

И2
Т0
Н2
Т2



СБОРНИК ТРУДОВ

**VIII Международной конференции
и молодежной школы «Информационные
технологии и нанотехнологии»**



**23-27 МАЯ
САМАРА, РОССИЯ**

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

ИНСТИТУТ СИСТЕМ ОБРАБОТКИ ИЗОБРАЖЕНИЙ РАН –
ФИЛИАЛ ФНИЦ "КРИСТАЛЛОГРАФИЯ И ФОТОНИКА" РАН

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
И НАНОТЕХНОЛОГИИ (ИТНТ-2022)
Том 3. Распознавание, обработка и анализ изображений

*Сборник трудов по материалам
VIII Международной конференции и молодежной школы
(г. Самара, 23-27 мая)*

Одобрено редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева»

© Самарский университет, 2022

ISBN 978-5-7883-1788-5

ISBN 978-5-7883-1791-5 (т. 3)

САМАРА
Издательство Самарского университета
2022

УДК 004.9
ББК 32.973
И741

И741 Информационные технологии и нанотехнологии (ИТНТ-2022): сборник трудов по материалам VIII Международной конференции и молодежной школы (г. Самара, 23-27 мая) / Министерство науки и высшего образования Российской Федерации, Самарский университет, Институт систем обработки изображений РАН – филиал ФНИЦ "Кристаллография и фотоника" РАН. – Самара: Издательство Самарского университета, 2022. – **Том 3: Распознавание, обработка и анализ изображений** / под ред. В. В. Мясникова. – 1 CD-ROM (7,3 Мб). – Загл. с титул. экрана. – Текст. Изображение: электронный.

ISBN 978-5-7883-1791-5 (т. 3)
ISBN 978-5-7883-1788-5

Тематика Конференции ИТНТ-2022 охватывает широкий круг областей применения информационных технологий в науке и высокотехнологичных отраслях промышленности. Одним из приоритетных направлений работы Конференции является образовательный аспект, заключающийся в предоставлении студентам и молодым ученым возможности ознакомиться с новейшими научными достижениями по тематике Конференции, а также с уникальным научным оборудованием и лабораторной базой Самарского университета, используемой для реализации современных научных проектов.

УДК 004.9
ББК 32.973

Минимальные системные требования:

PC, процессор Pentium, 160 МГц; оперативная память 32 Мб;
на винчестере 16 Мб; Microsoft Windows XP/Vista/7;
разрешение экрана 1024x768 с глубиной цвета 16 бит;
DVD-ROM2-х и выше, мышь; Adobe Acrobat Reader.

Редактор тома В.В. Мясников

Выпускающий редактор Д.В. Кудряшов

Подписано для тиражирования 13.07.2022.

Объем издания 7,3 Мб.

Количество носителей 1 диск.

Тираж 11 дисков.

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

443086, САМАРА, МОСКОВСКОЕ ШОССЕ, 34.

Издательство Самарского университета.
443086, Самара, Московское шоссе, 34.

Оглавление

Предисловие	7-10
1. Исследование точности восстановления сигнала на фоне шума методом максимума правдоподобия при двухпараметрическом анализе райсовских данных Т.В. Яковлева	030042
2. Modified Fizeau interferometer with the fringes polynomial smoothing algorithm I. Galaktionov, A. Nikitin, J. Sheldakova, V. Toporovsky, A. Kudryashov	030262
3. Метод дистанционного измерения скорости ветра в режиме реального времени В.В. Дудоров, С.О. Шестаков	030272
4. Application of NDVI and NDRE indices in the assessment of soybean productivity under controlled-release fertilizer В. Boiarskii	030292
5. Разработка общей структуры метода компрессии изображений, основанного на машинном обучении М.В. Гашников	030742
6. Выбор методов машинного обучения для сжатия изображений М.В. Гашников	030752
7. Разработка макета оптического триангуляционного сканера для контроля геометрии резьбового соединения бурильных труб Р.Р. Диязитдинов	030772
8. Проективное совмещение телевизионных изображений многокамерной системы технического зрения по данным натурной видеосъемки А.А. Диязитдинова	030792
9. Применение методов интеллектуального анализа изображений при определении типов лиц Ю.Ж. Пчелкина, Р.А. Парингер, П.Е. Савельева, М.В. Егорова	030802
10. Модификация алгоритмов машинного обучения для встраивания в методы компрессии изображений Р.Р. Юзькив, М.В. Гашников	030812
11. Correction of the interpolation effect in modeling the process of estimating image spatial deformations R. Kovalenko, A. Tashlinskii	030832
12. Обобщение основанного на машинном обучении метода компрессии на гиперспектральные изображения А.И. Максимов, М.В. Гашников	030842
13. Extraction of convex hulls of metal microstructure objects from metallographic images R. Magdeev, G. Safina	030872
14. Вероятность ошибки и вычислительная сложность классификации объектов в пространстве многоуровневых представлений М.М. Ланге, С.В. Парамонов	030882
15. Application of renyi mutual information in stochastic referencing of multispectral and multi-temporal images A. Tashlinskii, R. Ibragimov, G. Safina	030912
16. Оптимизация числа интервалов гистограммы, аппроксимирующей распределения вероятностей яркостей, при стохастическом совмещении изображений на основе взаимной информации Р.О. Коваленко, А.Г. Ташлинский	030982

17. Depth maps correction based on neighboring frames
A. Varlamova, V. Kitov 031052
18. Метод адаптивного управления дорожными сигналами на основе максимального взвешенного потока
А.А. Агафонов, А.С. Юмаганов, В.В.Мясников 031202
19. Восстановление пространственно неоднородных изображений на основе дважды стохастических фильтров
К.К. Васильев, В.Е. Дементьев, А.В. Белянчиков 031292
20. Использование алгоритмов обработки изображения для решения задачи навигации беспилотных аппаратов
К.В. Святков, В.Е. Дементьев, М.А. Гапонова 031302
21. Совместное применение нейронных сетей и псевдоградиентной привязки изображений при оценке дефектов мостовых конструкций
М.Н. Суетин, Р.Г. Магдеев, А.Г. Ташлинский 031312
22. Исследование стойкости компактных дескрипторов векторной карты к трансформациям её содержимого
А.С. Асанов, Ю.Д. Выборнова, В.А. Федосеев 031343
23. Сравнение алгоритмов управления сигналами светофоров в крупномасштабном сценарии моделирования движения транспортных средств
А.А. Агафонов, Е.Ю. Ефименко 031382
24. Сравнение методов реконструкции промежуточных кадров видеопоследовательностей с динамической сценой
Ю.Х. Ганеева 031472
25. Влияние реконструкции промежуточных кадров видеопоследовательности на результат 3D реконструкции объектов
Ю.Х. Ганеева, В.В. Мясников 031492
26. Влияние комплексирования спектральных каналов и сезонных космических снимков на различимость хвойного и лиственного леса
С.М. Зраенко 031592
27. Оптимизация траектории движения транспортных средств в задаче управления транспортными потоками на перекрестке
А.С. Юмаганов, А.А. Агафонов 031922
28. Аппаратно-программный комплекс для составления паспорта огнестрельного повреждения
И.Г. Пальчикова, Е.С. Смирнов, И.А. Будаева, И.В. Латышов, В.А. Васильев, А.В. Кондаков 031962
29. Исследование методов сегментации для выделения областей интереса на рентгеновских снимках лёгких
Д.В. Машков, Н.Ю. Ильясова, Н.С. Демин 032002
30. Повышение качества обработка видеоизображений на основе использования искусственной нейронной сети Yolo
В.Е. Дементьев, А.Г. Ташлинский, М. Н. Суетин 032052
31. Исследование эффективности метода стохастической модуляции для стеганографического встраивания в тепловые видеоданные
А.Ю. Баврина, Д.Д. Карнаухов, В.А. Федосеев 032112
32. Метод защиты видео при помощи фазовых цифровых водяных знаков
Д.А. Шапиро, В.В. Сергеев, В.А. Федосеев 032262
33. Распознавание оптических мод Эрмита Гаусса с использованием нейронной сети
В.В. Прудовская 032432

34. Распознавание оптических мод Лаггера Гаусса с использованием сверточной нейронной сети
А.В. Бехтерев 032442
35. Двухуровневый метод повышения качества сегментации изображений
Д.М. Мурашов 032452
36. Алгоритмы сокрытия и извлечения информации в аудио файлах
П.О. Волков 032503
37. Применение взаимной информации в задаче стохастической привязки расфокусированных
разноспектральных изображений
Р.М. Ибрагимов, О.В. Шидиков 033022
38. Оптимизация сверточных сетей с помощью квантизации и OpenVINO при распознавании
снимков багажа
Н.А. Андриянов, Дж. Папакостас 033052
39. Использование рецептивных полей в методах машинного обучения для обработки изображений,
представленных выборками случайных отсчетов
В.Е. Анциперов, В.А. Кершнер 033172
40. Обнаружение в илочкового маркера уреаплазмоза на изображении фации био логической жидкости
М.Л. Албутова, В.Р. Крашенинников, О.Е. Маленова, Л.И. Трубникова, Ю.Е. Кувайскова 033272
41. Новый алгоритм определения 3D координат центра роговицы глаза в системе бесконтактного
айтрекинга
Г.И. Громилин, В.П. Косых, Н.С. Яковенко 033372
42. Алгоритм разреженного представления в задаче устранения шумов на изображениях
Д. В. Архипова, Е.В. Гошин 033413
43. Исследование влияния шумов на результат сверхразрешения по цветным изображениям
Д.А. Аксенова, Е.В. Гошин 033432
44. Генеративная модель автокодировщиков изображений на основе рецептивных полей
В.Е. Агциперов 033662
45. Разработка сервиса сегментации спутниковых данных с помощью алгоритмов глубокого обучения
и библиотеки OpenVINO
Е.П. Васильев, А.А. Филатов, И.С. Микерин, Н.О. Ким, Д.А. Ермолаев 033792
46. Гиперспектральные и тепловые изображения в ранней диагностике засухи растений пшеницы
с помощью ХАІ
М.А. Лысов, В.Е. Турлапов 033892

ПРЕДИСЛОВИЕ

Конференция ИТНТ-2022 проводится с целью предоставления возможности научных дискуссий и обсуждения результатов фундаментальных и прикладных исследований в области информационных технологий и нанотехнологий, привлечения молодежи в сферу передовых научных исследований, обмена опытом научно-образовательной деятельности при подготовке ИТНТ-специалистов.

Тематика Конференции ИТНТ-2022 охватывает широкий круг областей применения информационных технологий в науке и высокотехнологичных отраслях промышленности.

Основными направлениями работы Конференции ИТНТ-2022 являются:

Компьютерная оптика и нанофотоника

- дифракционная оптика;
- планарные оптические структуры;
- гиперспектральные системы;
- нанофотоника;
- системы оптической сенсорики, передачи и обработки информации;
- сингулярная оптика.

Информационные технологии дистанционного зондирования Земли

- информационные технологии в проектировании космических аппаратов дистанционного зондирования Земли и полезных нагрузок для них;
- программные и математические решения для управления движением космических аппаратов наблюдения;
- программные и аппаратные средства для получения, обработки и анализа данных, получаемых с космических аппаратов дистанционного зондирования Земли;
- математическое моделирование процессов функционирования космических аппаратов дистанционного зондирования Земли;
- современные проектные решения для создания космических аппаратов мониторинга Земли и околоземного пространства и их группировок, в том числе на базе аппаратов типа CubeSat;
- системы дистанционного зондирования Земли на основе БПЛА.

Распознавание, обработка и анализ изображений

- математические методы цифровой обработки изображений и распознавания образов
- трёхмерное зрение
- биометрические системы на основе изображений
- геоинформационные системы и технологии
- защита и верификация мультимедиа

Искусственный интеллект

- новые подходы, тренды и фундаментальные результаты в сфере искусственного интеллекта и его приложениях к распознаванию образов и анализу изображений, обработке текстов, речевой информации;
- нейросетевые методы и глубокое обучение;
- прикладные технологии искусственного интеллекта в обработке изображений, беспилотном транспорте, производственных и сельскохозяйственных приложениях, медицинских приложениях, экологии, мониторинге окружающей среды и других;
- программные технологии для решения задач искусственного интеллекта – фреймворки, библиотеки, открытые инициативы и сообщества;
- мультидисциплинарные аспекты искусственного интеллекта и машинного обучения.

Науки о данных

- Компьютерные науки:
 - инженерия данных;
 - визуализация данных;
 - математические методы анализа данных;
 - программные платформы и библиотеки для работы с данными;
 - аппаратные средства хранения и обработки данных;
 - высокопроизводительные, параллельные и облачные вычисления, технологии обработки больших данных;
 - базы данных, инструменты и языки для работы с базами данных.
- Прикладные задачи интеллектуального анализа данных:
 - решение актуальных прикладных задач.

Одним из приоритетных направлений работы Конференции ИТНТ-2022 является образовательный аспект, заключающийся в предоставлении студентам и молодым ученым возможности ознакомиться с новейшими научными

достижениями по [тематике](#) Конференции, а также с уникальным научным оборудованием и лабораторной базой [Самарского университета](#), используемой для реализации современных научных проектов.

В рамках Конференции проводится Молодежная школа, где молодые ученые и студенты получают возможность повысить свой профессиональный уровень и [опубликовать](#) свои научные результаты.

В данный сборник трудов вошли материалы по 5 основным направлениям Конференции:

Том 1. Компьютерная оптика и нанофотоника (под редакцией к.ф.-м.н. Е.С. Козловой)

Том 2. Информационные технологии дистанционного зондирования Земли (под редакцией к.т.н. И.С. Ткаченко)

Том 3. Распознавание, обработка и анализ изображений (под редакцией д.ф.-м.н. В.В. Мясникова)

Том 4. Искусственный интеллект (под редакцией д.т.н. А.В. Никонорова)

Том 5. Науки о данных (под редакцией д.т.н. А.В. Куприянова).

Выпускающий редактор томов 1-5: Д.В. Кудряшов.

Официальный сайт Конференции ИТНТ-2022: <http://itnt-conf.org/>

ОРГАНИЗАТОРЫ

- Институт систем обработки изображений РАН (ИСОИ РАН) – филиал ФНИЦ «Кристаллография и фотоника» РАН, г. Самара, Россия;
- Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет), г. Самара, Россия.

Организационный комитет

Председатель

Богатырёв В.Д. – д.э.н., профессор, ректор Самарского национального исследовательского университета, г. Самара, Россия

Заместители председателя

Казанский Н.Л. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;

Сергеев В.В. – д.т.н., профессор, Самарский университет, г. Самара, Россия;

Куприянов А.В. – д.т.н., профессор, Самарский университет, г. Самара, Россия.

Ответственный секретарь

Хабибуллин Р.М. – к.т.н., Самарский университет, г. Самара, Россия.

Члены Организационного комитета

Бланк В.А. – к.ф.-м.н., Самарский университет, г. Самара, Россия;

Бояркин Ю.Н. – ИСОИ РАН, г. Самара, Россия;

Востокин С.В. – д.т.н., Самарский университет, г. Самара, Россия;

Гашников М.В. – к.т.н., Самарский университет, г. Самара, Россия;

Гошин Е.В. – к.т.н., Самарский университет, г. Самара, Россия;

Жердев Д.А. – к.т.н., ИСОИ РАН, г. Самара, Россия;

Кадомина Е.А. – Самарский университет, г. Самара, Россия;

Кириш Д.В. – к.т.н., Самарский университет, г. Самара, Россия;

Котов А.П. – к.т.н., ИСОИ РАН, г. Самара, Россия;

Кузнецов А.В. – к.т.н., Самарский университет, г. Самара, Россия;

Куприянов Д.Д. – Самарский университет, г. Самара, Россия;

Максимов А.И. – Самарский университет, г. Самара, Россия;

Мисиевич С.К. – Самарский университет, г. Самара, Россия;

Паренский Н.А. – Самарский университет, г. Самара, Россия;

Подлипов В.В. – ИСОИ РАН, г. Самара, Россия;

Попов С.Б. – д.т.н., проф., Самарский университет, г. Самара, Россия;

Пресняков К.Г. – Департамент информационных технологий и связи Самарской области, г. Самара, Россия;

Рыцарев И.А. – Самарский университет, г. Самара, Россия;

Савельева А.А. – Самарский университет, г. Самара, Россия;

Скиданова Е.Б. – Самарский университет, г. Самара, Россия;

Смагин С.В. – ИСОИ РАН, г. Самара, Россия;

Стафеев С.С. – к.ф.-м.н., ИСОИ РАН, г. Самара, Россия;

Тиц С.Н. – к.т.н., Самарский университет, г. Самара, Россия;

Ткаченко И.С. – к.т.н., доцент, Самарский университет, г. Самара, Россия;

Хнырева Е.С. – Самарский университет, г. Самара, Россия;

Фомченков С.А. – Самарский университет, г. Самара, Россия;

Якуненкова Д.М. – ИСОИ РАН, г. Самара, Россия.

Программный комитет

Председатель

Сойфер В.А. – академик РАН, д.т.н., профессор, президент Самарского университета, Самара, Россия.

Заместитель председателя

Казанский Н.Л. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия.

Члены Программного комитета

Korotkova O. – Prof., University of Miami, Coral Gables, USA;
Niemann H. – Prof., Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany;
O’Faolain L. – Prof., Tyndall National Institute, Cork, Ireland;
Sazhin S.- Prof., University of Brighton, Brighton, United Kingdom;
Sobolewski M. – Prof., Polish-Japanese Institute of IT, Warsaw, Poland;
Калачёв Л.В. – профессор, Университет Монтаны, г. Миссула, США;
Коробейников А. – профессор, Педагогический университет Шеньси, г. Сиань, Китай;
Бычков И.В. – академик РАН, д.т.н., профессор, Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения РАН, г. Иркутск, Россия;
Воеводин В.В. – член-корреспондент РАН, д.ф.-м.н., профессор, Московский государственный университет имени М.В. Ломоносова., г. Москва, Россия;
Головашкин Д.Л. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;
Гуляев Ю.В. – академик РАН, д.ф.-м.н., профессор, Институт радиотехники и электроники им. В.А. Котельникова РАН, г. Москва, Россия;
Желтов С.Ю. – академик РАН, д.т.н., профессор, ГосНИИ авиационных систем, г. Москва, Россия;
Журавлев Ю.И. – академик РАН, д.ф.-м.н., профессор, Вычислительный центр им. А.А. Дородницына РАН, г. Москва, Россия;
Калошин В.А. – д.ф.-м.н., профессор, Институт радиотехники и электроники им. В.А. Котельникова РАН, г. Москва, Россия;
Козлова Е.С. – к.ф.-м.н., ИСОИ РАН, г. Самара, Россия.
Конов В.И. – академик РАН, д.ф.-м.н., профессор, Институт общей физики им. А.М. Прохорова РАН, г. Москва, Россия;
Котляр В.В. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;
Кульчин Ю.Н. – академик РАН, д.ф.-м.н., профессор, Институт автоматизации и процессов управления Дальневосточного отделения РАН, г. Владивосток, Россия;
Куприянов А.В. – д.т.н., профессор, Самарский университет, г. Самара, Россия;
Лабунец В.Г. – д.т.н., профессор, Уральский государственный лесотехнический университет, г. Екатеринбург, Россия;
Мясников В.В. – д.ф.-м.н., профессор, Самарский университет, г. Самара, Россия;
Никитов С.А. – член-корреспондент РАН, д.ф.-м.н., профессор, Институт радиотехники и электроники им. В.А. Котельникова РАН, г. Москва, Россия;
Никоноров А.В. - д.т.н., ИСОИ РАН, г. Самара, Россия;
Новиков Д.А. – чл.-корр. РАН, д.т.н., профессор, Института проблем управления РАН, г. Москва, Россия;
Потатуркин О.И. – д.т.н., профессор, Институт автоматизации и электрометрии Сибирского отделения Российской академии наук, г. Новосибирск, Россия;
Ткаченко И.С. – к.т.н., доцент, Самарский университет, г. Самара, Россия;
Хонина С.Н. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;
Чочиа П.А. – д.т.н., Институт проблем передачи информации им. А.А. Харкевича РАН, г. Москва, Россия.

Исследование точности восстановления сигнала на фоне шума методом максимума правдоподобия при двухпараметрическом анализе райсовских данных

Т.В. Яковлева

Федеральный исследовательский центр «Информатика и управление» РАН
Москва, Россия
tan-ya@bk.ru

Аннотация—В работе исследуется точность расчета величины сигнала, формирующего магнитно-резонансное или ультразвуковое изображение и восстанавливаемого на фоне шума методом максимума правдоподобия при двухпараметрическом анализе райсовских данных, в сопоставлении с максимально достижимой точностью, определяемой величиной нижней границы для стандартного отклонения оценки искомого параметра, в зависимости от параметров задачи, таких как отношение сигнала к шуму и длина выборки измерений. Рассматриваются теоретические аспекты решаемой задачи и представлены результаты численного расчета стандартного отклонения искомого параметра в сопоставлении с теоретическим минимумом этой величины, определяемым неравенством Рао-Крамера. Результаты проведенных численных экспериментов показали, что стандартное отклонение сигнала, восстановленного методом максимума правдоподобия, лишь незначительно превосходит его теоретическую нижнюю границу, что подтверждает высокую эффективность двухпараметрического анализа данных в условиях распределения Райса.

Ключевые слова— *распределение Райса, метод максимума правдоподобия, стандартное отклонение.*

1. ВВЕДЕНИЕ

Эффективность обработки стохастических сигналов существенно зависит от особенностей статистического распределения анализируемых данных. Отмечаемый в последние годы интерес к изучению статистического распределения Райса объясняется тем, что данное распределение адекватно описывает широкий круг задач, в которых выходной сигнал формируется как сумма искомого исходного сигнала и гауссовского шума, а анализируемой величиной является амплитуда результирующего сигнала, [1-4]. В частности, распределению Райса подчиняются данные, формирующие магнитно-резонансное и ультразвуковое изображения. Для анализа изображений в рамках модели Райса были разработаны так-называемые методы двухпараметрического анализа, обеспечивающие совместный расчет информативной и шумовой составляющих сигнала [5, 6].

2. ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ И ЧИСЛЕННОЕ ТЕСТИРОВАНИЕ ТОЧНОСТИ МЕТОДА МАКСИМУМА ПРАВДОПОДОБИЯ ПРИ ДВУХПАРАМЕТРИЧЕСКОМ АНАЛИЗЕ РАЙСОВСКИХ ДАННЫХ

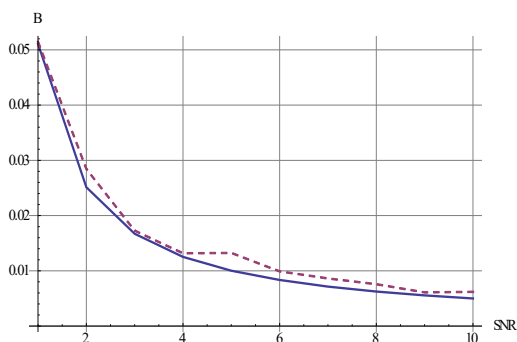
Теоретические исследования в области анализа стохастических данных [5-8], доказали возможность эффективного решения задач совместного расчета сигнала и шума в условиях применимости статистической модели Райса посредством применения методов двухпараметрического анализа, суть которых сводится к расчету параметров распределения Райса на основе данных выборочных измерений, чем обеспечивается математическое разделение полезной, информативной составляющей сигнала и его шумовой компоненты.

Настоящая работа направлена на изучение точности восстановления сигнала на фоне шума и сопоставление ее с максимально достижимой точностью, определяемой величиной нижней границы для дисперсии оценки неизвестного параметра на основе неравенства Рао-Крамера.

Метод максимума правдоподобия для расчета величины полезного сигнала состоит в решении следующей системы уравнений для параметров сигнала ν и шума σ^2 [5]:

$$\begin{cases} \nu = \frac{1}{n} \sum_{i=1}^n x_i \cdot I_1(x_i \cdot \nu / \sigma^2) / I_0(x_i \cdot \nu / \sigma^2) \\ \sigma^2 = (\langle x^2 \rangle - \nu^2) / 2. \end{cases} \quad (1)$$

где $I_\delta(z)$ - модифицированная функция Бесселя первого рода порядка δ , x_i - i -я выборка сигнала. Параметр ν соответствует исходной, не искаженной величине сигнала, а σ^2 - дисперсия шума. Как показали результаты численных экспериментов, при расчете сигнала методом максимума правдоподобия величина стандартного отклонения параметра $S = \nu^2$ лишь незначительно превосходит его нижнюю границу, что иллюстрируется Рисунком 1.



Полученная в результате численного эксперимента кривая стандартного отклонения оцениваемого параметра S (пунктирная линия) и нижняя граница стандартного отклонения этого параметра (сплошная линия) в зависимости от отношения сигнала к шуму

На Рис. 1 представлены кривые, отображающие численно рассчитанную величину стандартного отклонения B_{calc} оцениваемого параметра $P = \nu^2$ в сравнении с нижней теоретической границей данной величины B_{RC} , определяемой неравенством Рао-Крамера, в зависимости от отношения сигнала к шуму. При этом стандартное отклонение B_{calc} параметра P рассчитывается на основе решения системы (1) уравнений максимума правдоподобия для параметров сигнала ν и шума σ^2 .

Таким образом, результаты проведенного численного эксперимента демонстрируют следующее: метод максимума правдоподобия при двухпараметрическом анализе райсовских данных обеспечивает достаточно высокую точность оценивания полезной, информативной составляющей сигнала ν , такую, что величина стандартного отклонения B_{calc} параметра $P = \nu^2$ превосходит свой предельно достижимый теоретический уровень B_{RC} очень незначительно.

3. ЗАКЛЮЧЕНИЕ

В работе изучается точность восстановления сигнала на фоне шума методом максимума правдоподобия при двухпараметрическом анализе райсовских данных в сопоставлении с ее предельно достижимым значением,

определяемым на основе неравенства Рао-Крамера. Представлены теоретические аспекты решаемой задачи, а также результаты ее численного решения.

Полученные в ходе численных экспериментов результаты подтверждают возможность достижения высокой точности при восстановлении искомой информативной составляющей сигнала на фоне шума рассматриваемым методом максимума правдоподобия, что весьма важно для решения проблем обработки изображений в системах магнитно-резонансной и ультразвуковой визуализации.

Результаты демонстрируют, что метод максимума правдоподобия при двухпараметрическом анализе райсовских данных обеспечивает точность, практически совпадающую с ее максимально достижимым теоретическим пределом в широком диапазоне значений отношения сигнала к шуму.

ЛИТЕРАТУРА

- [1] Rice, S.O. Mathematical analysis of random noise / S.O. Rice // Bell Syst. Technological J. – 1944. – Vol. 23. – P. 282.
- [2] Benedict, T.R. The joint estimation of signal and noise from the sum envelope / T.R. Benedict, T.T. Soong // IEEE Trans. Inf. Theory. – 1967. – Vol. IT-13(3). – P. 447-454.
- [3] Talukdar, K.K. Estimation of the parameters of Rice distribution / K.K. Talukdar, W.D. Lawing // J. Acoust. Soc. Amer. – 1991. – Vol. 89(3). – P. 1193-1197.
- [4] Sijbers, J. Maximum-Likelihood Estimation of Rician Distribution Parameters / J. Sijbers, A.J. den Dekker, P. Scheunders, D. Van Dyck // IEEE Transactions on Medical Imaging. – 1998. – Vol. 17(3). – P. 357-361.
- [5] Yakovleva, T.V. Methods of Mathematical Statistics in Two-Parameter Analysis of Rician Signals / T.V. Yakovleva, N.S. Kulberg // Doklady Mathematics. – 2014. – Vol. 90(3). – P. 1-5. DOI: 10.1134/S1064562414070060.
- [6] Яковлева, Т.В. Условия применимости статистической модели Райса и расчет параметров Райсовского сигнала методом максимума правдоподобия / Т.В. Яковлева // Компьютерные исследования и моделирование. – 2014. – Т. 6, № 1. – С. 13-25. DOI: 10.20537/2076-7633-2014-6-1-13-25.
- [7] Яковлева, Т.В. Теоретическое обоснование методов математической статистики для оценивания параметров сигнала и шума при анализе райсовских данных / Т.В. Яковлева // Компьютерные исследования и моделирование. – 2016. – Т. 8, № 3. – С. 445-473. DOI: 10.20537/2076-7633-2016-8-3-445-473.
- [8] Yakovleva, T. Peculiarities of the Rice Statistical Distribution: Mathematical Substantiation/ T. Yakovleva // Applied and Computational Mathematics, Science Publishing Group. – 2018. – Vol. 7(4). – P. 188-196. DOI: 10.11648/j.acm.20180704.12.

Modified Fizeau interferometer with the fringes polynomial smoothing algorithm

I. Galaktionov
Institute of Geosphere Dynamics RAS
Moscow, Russia
galaktionov@activeoptics.ru

A. Nikitin
Institute of Geosphere Dynamics RAS
Moscow, Russia
nikitin@activeoptics.ru

J. Sheldakova
Institute of Geosphere Dynamics RAS
Moscow, Russia
sheldakova@activeoptics.ru

V. Toporovsky
Institute of Geosphere Dynamics RAS
Moscow, Russia
topor@activeoptics.ru

A. Kudryashov
Institute of Geosphere Dynamics RAS
Moscow, Russia
kud@activeoptics.ru

Abstract—The modified Fizeau interferometer that allows to diagnose plane and spherical optical elements with the diameter ranging from 10 to 100 mm is discussed. The method of interference patterns reconstruction based on reference lines is modified, the algorithm of 4th order polynomial smoothing is implemented. The modified method increases the reliability and accuracy of interferometric pattern reconstruction and suppress the influence of incoming noise. Accuracy of the measurements is about $\lambda/10$ ($\lambda=0.63$ μm).

Keywords— Fizeau interferometer, optical surface testing, polynomial smoothing algorithm.

1. INTRODUCTION

Interferometric techniques for optical surface quality control are well-known and well-established methods due to the high accuracy of the interferometer devices. In this paper we present the modified Fizeau interferometer that was assembled and tested in our laboratory as well as discuss the new algorithm of 4th order polynomial smoothing of reconstructed interference patterns.

Overall, the main goal of this work was to increase the efficiency of the interference pattern reconstruction algorithm of the Fizeau interferometer to measure the flatness of optical elements in case of noisy image conditions. The origin of the noise is not really matter for the algorithm – whether it is due to the presence of inhomogeneities along the optical path between the input aperture of the interferometer and the detail under test or it is due to digital noise of the video camera used.

2. OPTICAL SCHEME OF THE INTERFEROMETER

The interferometer described in this paper is based on Fizeau scheme [1]–[3]. The scheme of the interferometer assembled in our laboratory is presented on Fig. 1.

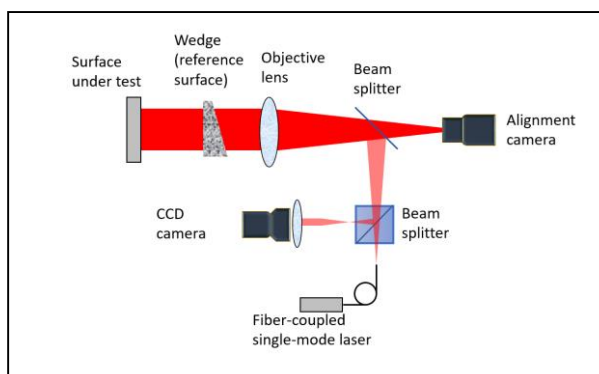


Fig. 1. Principal scheme of the Fizeau interferometer

3. PHASE SURFACE RECONSTRUCTION ALGORITHM

One of the main problems of analysis of interference pattern images is how to correctly determine extremuma. For this purpose we used the previously developed method that allowed to exclude false extremuma and to determine the position of the extremuma with high precision [3].

4. FOURTH ORDER POLYNOMIAL SMOOTHING ALGORITHM

Due to the presence of noise component on the interference pattern image it was necessary to reduce it. We used the algorithm of moving average to smooth the noise effects. If the noise impact is rather high, the interference pattern will be reconstructed with the error because of the error in determination of extremuma. But we found out that sometimes the strength of the moving average algorithm is not enough. There are still some reconstructed fringes with a false bends after applying the algorithm of moving average. This bends introduces the additional error to the calculated set of Zernike coefficients [4], [5] and to the resultant phase surface [6].

In order to decrease this error, we applied the 4th order polynomial smoothing algorithm. This algorithm takes the X and Y coordinates of each point of the determined interference patterns and approximates it using the polynomials of 4th order. By solving the least squares problem, the approximation coefficients C_i are obtained. The resultant Y_{new} coordinates of each point of the determined interference patterns are recalculated (based on the old Y_{old} values) using the formula:

$$Y_{new} = C_0 + C_1 \cdot Y_{old} + C_2 \cdot (Y_{old})^2 + C_3 \cdot (Y_{old})^3 + C_4 \cdot (Y_{old})^4$$

The application of this algorithm allows to eliminate false bends of the color lines in the interference patterns.

There are a few other algorithms that supposed to increase the efficiency of the fringe identification and reconstruction. In [7] the nonlinear regression analysis is described. But as is known, all nonlinear regression algorithms are iterative. As such they require significant amounts of computer time. Also, unlike linear regression analysis, they require initial guesses of the values of all coefficients. In [8] authors applied continuous wavelet transform and Hilbert transform algorithms to obtain phase map from the interference fringes pattern. However, Hilbert transform algorithm gives a noisy result, since its frequency spectrum has harmonic noise. In [9] the authors proposed a method for producing stabilized

interference patterns using CCD camera as the detector element. This method allowed to decrease the photoresist grating fill factor from ~60% to less than 50%.

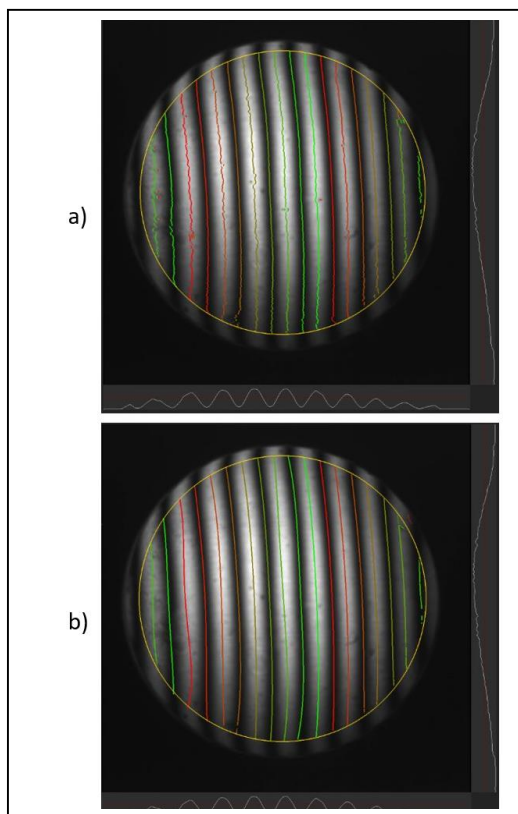


Fig. 2. a) interference fringe pattern and reconstructed fringes (color lines) after applying classical moving average algorithm; b) reconstructed fringes after applying the 4th order polynomial smoothing algorithm

The advantages of the proposed fourth order polynomial smoothing algorithm are independence of noise and fast calculation speed which allows to reconstruct and smooth the interference fringes in real-time (figure 2).

Table 1 contains the resultant PV, RMS and Zernike coefficients obtained after interference fringes pattern reconstruction before (row #1) and after (row #2) applying 4th order polynomial smoothing algorithm.

TABLE 1. ZERNIKE COEFFICIENTS (IN WAVELENGTH) FOR THE RECONSTRUCTED FRINGES

	PV	RMS	Z3	Z4	Z5	Z6	Z7	Z8
1.	0.3	0.046	0.069	0.032	-0.014	0.005	-0.01	0.01
2.	0.13	0.038	0.061	0.025	-0.004	-0.002	-0.001	0.003

The proposed algorithm of polynomial smoothing can possibly be effective in solving the problems of phase surface reconstruction in Shack-Hartmann wavefront sensors [10], [11] with the B-spline approximation technique [12]. Moreover, such a technique can be applied in case of approximation of average wavefronts [13] of the radiation passed through the scattering medium [14], because the wavefront in that case is almost flat in the central part of the beam and dramatically increases to the edge of the beam [14]. It can cause an error in surface approximation which can be decreased by means of the proposed algorithm.

5. CONCLUSION

The method of interference patterns reconstruction based on reference lines is modified, the algorithm of 4th order polynomial smoothing is implemented and successfully tested. The modified method increases the reliability and accuracy of interferometric pattern reconstruction and suppress the influence of incoming noise. The proposed algorithm can also be applied to the phase reconstruction in Shack-Hartmann sensor.

ACKNOWLEDGMENT

The research was carried out within the Russian Science Foundation project # 20-69-46064.

REFERENCES

- [1] Goodwin, E. Interferometric Optical Testing / E.P. Goodwin, L.C. Wyant // SPIE Field Guides. – 2006. – FG10.
- [2] Romanov, P. Modified interferometer fizeau for diagnostics of wide aperture optical elements / P. Romanov, J. Sheldakova, V. Zavalova, A. Alexandrov, V. Dubikovskiy, A. Kudryashov // Proc. CAOL. – 2008. – P. 393-394.
- [3] Sheldakova, J. Shack-Hartmann wavefront sensor versus Fizeau interferometer for laser beam measurements / J. Sheldakova, A. Kudryashov, V. Zavalova, P. Romanov // Proc. SPIE. – 2009. – Vol. 7194. – P. 71940B.
- [4] Wyant, J. Basic wavefront aberration theory for optical metrology / J.C. Wyant, K. Creath // Proc. of Applied optics and optical engineering. – 1992. – P. 27-39.
- [5] Galaktionov, I. Laser beam propagation and wavefront correction in turbid media / I. Galaktionov, A.V. Kudryashov, J. Sheldakova, A. Byalko, G. Borsoni // Proc. of SPIE. – 2015. – Vol. 9617. – P. 96170D.
- [6] Galaktionov, I. Phase correction of laser beam passed through turbid medium / I. Galaktionov, J. Sheldakova, A. Kudryashov // International Conference Laser Optics. – 2014. – P. 1-1. DOI: 10.1109/LO.2014.6886307.
- [7] Schemm, J. Fringe pattern recognition and interpolation using nonlinear regression analysis / J. Schemm, C. Vest // Applied Optics. – 1983. – Vol. 22(18). – P. 2850-2853.
- [8] Saraç, Z. Phase recovery from interference fringes by Hilbert transform / Z. Saraç, H. Birkök, A. Emir, A. Dursun // Proceedings of SPIE. – 2009. – Vol. 7443.
- [9] Young, P. Simple interferometric fringe stabilization by charge-coupled-device-based feedback control / P. Young, P. Priambodo, T. Maldonado, R. Magnusson // Appl. Opt. – 2006. – Vol. 45. – P. 4563-4566.
- [10] Southwell, W. Wavefront estimation from wavefront slope measurements / W.H. Southwell // J. Opt. Soc. Am. – 1980. – Vol. 70(8). – P. 998-1005.
- [11] Nikitin, A. Absolute calibration of a Shack-Hartmann wavefront sensor for measurements of wavefronts / A. Nikitin, I. Galaktionov, D. Denisov, V. Karasik, A. Sakharov, N. Baryshnikov, J. Sheldakova, A. Kudryashov // Proc. of SPIE Photonic Instrumentation Engineering VI. – 2019. – Vol. 10925. – P. 109250K.
- [12] Galaktionov, I. B-spline approximation of a wavefront measured by Shack-Hartmann sensor / I. Galaktionov, A. Nikitin, J. Sheldakova, A. Kudryashov // Proc. SPIE Laser Beam Shaping XXI. – 2021. – Vol. 11818. – P. 118180N.
- [13] Galaktionov, I. Laser beam focusing through the scattering medium-low order aberration correction approach / I. Galaktionov, A. Nikitin, V. Samarkin, J. Sheldakova, A. Kudryashov // Proc. SPIE, Unconventional and Indirect Imaging, Image Reconstruction, and Wavefront Sensing. – 2018. – Vol. 10772. – P. 107720W.
- [14] Galaktionov, I. Laser beam focusing through the scattering medium by means of adaptive optics / I. Galaktionov, A. Kudryashov, J. Sheldakova, A. Nikitin // Proc. SPIE Adaptive Optics and Wavefront Control for Biological Systems III. – 2017. – Vol. 10073. – P. 100731L.
- [15] Galaktionov, I. Laser beam focusing through the scattering medium using 14-, 32- and 48-channel bimorph mirrors / I. Galaktionov, J. Sheldakova, A. Kudryashov, A. Nikitin // International Conference Laser Optics (ICLO). – 2018. – P. 223-223.

Метод дистанционного измерения скорости ветра в режиме реального времени

В.В. Дудоров

Институт оптики атмосферы имени В. Е. Зуева СО РАН
Томск, Россия
dvv@iao.ru

С.О. Шестаков

Институт оптики атмосферы имени В. Е. Зуева СО РАН
Томск, Россия
sso@iao.ru

Аннотация—Мы предлагаем метод дистанционного измерения поперечной составляющей скорости ветра в режиме реального времени на основе обработки искажённых атмосферой изображений. Метод реализован в программном обеспечении и работает с использованием высокочастотной видеокамеры высокого разрешения. Представлены результаты лабораторного эксперимента с использованием разработанной программы.

Ключевые слова— обработка изображений, анализ изображений, режим реального времени, измерение скорости ветра.

1. ВВЕДЕНИЕ

Известно, что степень искажения оптических полей в атмосфере определяется силой турбулентности [1-3]. На основании этой закономерности разрабатываются методы дистанционного зондирования таких характеристик атмосферы как структурная характеристика флуктуаций показателя преломления, масштабы турбулентности, скорость ветра и другие [4-12]. В задачах измерения скорости ветра при анализе турбулентных искажений оптических полей можно определять компоненту вектора скорости ветра, перпендикулярную направлению распространения излучения [8, 9], в отличие от доплеровских лидаров, зондирующих продольную скорость ветра на основе рассеянного аэрозолем излучения [11].

Существующие пассивные методы измерения поперечного ветра основываются на определении сдвига функции корреляции дрожания изображений некоторого контрастного элемента объекта, сформированных двумя рядом расположенными приемными телескопами, либо корреляции дрожания изображений двух однотипных элементов объекта, сформированных единственным телескопом [9]. Кроме этого, также анализируется корреляция флуктуаций интенсивности излучения в плоскости изображения [10]. Общей особенностью данных подходов является необходимость накопления статистики флуктуаций анализируемых величин, что соответствует измерению скорости ветра, усредненной за некоторый интервал времени, который обычно составляет не менее нескольких секунд [12]. Это является недостатком использования данного подхода в задачах, когда необходимо знать мгновенное значение скорости ветра, как, например, при корректировке линии прицеливания или определении благоприятных условий взлета и посадки летательных аппаратов.

Предлагаемый нами метод позволяет измерять мгновенную скорость поперечного ветра на трассе между объектом и приёмной системой. Это является главным преимуществом метода перед аналогами, которые проводят измерения с усреднением по некоторому временному промежутку. Кроме того, при правильной

реализации этого метода и использовании подходящей аппаратной части возможно удовлетворение требований к скорости регистрации кадров и скорости их обработки, возникающих из-за изменчивости атмосферных неоднородностей и необходимости работы в режиме реального времени.

Фактически задача заключается в детектировании атмосферных неоднородностей на трассе наблюдения и отслеживании их перемещения с течением времени. Скорость перемещения неоднородностей на трассе является искомой скоростью ветра. То есть для измерения скорости ветра необходимо вначале выделить атмосферные абберрации из снятых кадров и затем проследить за скоростью их перемещения.

2. ОПИСАНИЕ МЕТОДА

Предлагаемый алгоритм начинается с выделения атмосферных неоднородностей в кадрах видеопотока. Это выделение производится с помощью характеристики изображения, которая называется картой качества:

$$Q(\mathbf{r}, a_k, t) = |\nabla_{\mathbf{r}} I(\mathbf{r}, t)|^2 \otimes K(\mathbf{r}, a_k), \quad (1)$$

где $I(\mathbf{r}, t)$ – изображение, $\mathbf{r} = \{x, y\}$ – вектор в плоскости, перпендикулярной оси наблюдения OZ, $\nabla_{\mathbf{r}} = \mathbf{i} \partial/\partial x + \mathbf{j} \partial/\partial y$ – вектор-градиент, $|\nabla_{\mathbf{r}} I(\mathbf{r}, t)|^2$ – функционал резкости изображения, \otimes – операция свёртки, a_k – коэффициент сглаживания, $K(\mathbf{r}, a_k)$ – функция сглаживания в виде гауссоиды.

Используя карты качества соседних кадров (зарегистрированных в моменты времени t и $t - \Delta t$) рассчитывается функция взаимной корреляции (2) с целью детектирования смещения выделенных абберраций на трассе. Аргумент максимального значения функции взаимной корреляции (3) будет указывать на величину смещения неоднородностей.

$$R(\mathbf{r}, a_k, t) = \int d\mathbf{r}' Q(\mathbf{r}', a_k, t - \Delta t) Q(\mathbf{r} + \mathbf{r}', a_k, t), \quad (2)$$

$$\Delta \mathbf{r}(t) = \arg \max [R(\mathbf{r}, a_k, t)]. \quad (3)$$

Далее скорость ветра рассчитывается с использованием следующего выражения:

$$\mathbf{v}(t) = \Delta \mathbf{r}(t) \frac{z_{is}}{\Delta t} \frac{1}{z_{im}}, \quad (4)$$

где z_{is} – расстояние до анализируемого слоя атмосферы, z_{im} – расстояние от плоскости апертуры до плоскости изображения.

3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

На основе вышеизложенного метода было разработано программное обеспечение, позволяющее

рассчитывать поперечную составляющую скорости ветра на горизонтальной трассе в режиме реального времени. Программа использует высокоскоростную цифровую видеокамеру Мако U-130В с интерфейсом подключения USB 3.0 и разрешением 1280×1024 точек. Как было сказано выше, скорость регистрации и обработки кадров имеют решающее значение при использовании предлагаемого метода. Используемая нами камера при своем максимальном разрешении позволяет регистрировать около 160-ти кадров в секунду, что является достаточным для метода. Разработанная программа обрабатывает 100-160 кадров в секунду в зависимости от размеров областей интереса в кадре. Все вычисления происходят на центральном процессоре ПК. Такая скорость работы была достигнута на 4-х ядерном процессоре с 8-ю потоками и базовой тактовой частотой работы 4.00 ГГц. Таким образом, выполняются требования скорости съемки и обработки кадров для успешного применения метода.

4. ЛАБОРАТОРНЫЙ ЭКСПЕРИМЕНТ

С использованием разработанного программного обеспечения были проведены лабораторные эксперименты. Целью экспериментов являлись проверка и демонстрация работы метода на практике в контролируемых условиях. Таким образом можно более точно проследить за реакцией метода на запланированные возникновения и затухания ветра, в отличие от экспериментов в натуральных уличных условиях, когда ветер ведёт себя неконтролируемо и из-за этого проблематично определить распределение ветра на всей трассе.

Имитацию ветра в эксперименте выполнял комнатный вентилятор с нагревателем. В качестве тестовой трассы выступал коридор в главном корпусе ИОА СО РАН. Длина трассы от видеокамеры до тестового объекта составляла 75 метров. Результат расчетов метода показан на рис. 1.

На графике можно увидеть динамику изменений скорости смещения неоднородностей в течение 30-ти секунд продолжительности эксперимента. График колеблется около нуля в периоды, когда вентилятор был выключен, и явные смещения неоднородностей отсутствовали. А в периоды, когда вентилятор был включен вначале наблюдается равномерный быстрый рост кривой, затем колебания в пределах смещения 40-50 пикселей, что соответствует скорости примерно 1 м/с, и в конце такое же равномерное быстрое падение кривой графика.

Таким образом, в эксперименте было показано, что предлагаемый метод измерения поперечной скорости

ветра на трассе работоспособен и адекватно детектирует динамику изменения скорости смещения атмосферных неоднородностей.

5. ЗАКЛЮЧЕНИЕ

В настоящей работе представлен метод дистанционного измерения поперечной составляющей скорости ветра, а также его экспериментальная апробация в режиме «реального времени». Показана работоспособность метода в контролируемых лабораторных условиях. Дальнейшее исследование будет направлено на апробацию предложенного метода в реальных атмосферных условиях на трассах различной протяженности.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке РФФИ (проект № 20-32-90214).

ЛИТЕРАТУРА

- [1] Лукин, В.П. Адаптивное формирование пучков и изображений в атмосфере / В.П. Лукин, Б.В. Фортес. – Новосибирск: Издательство Сибирского отделения РАН, 1999. – 214 с.
- [2] Воронцов, М.А. Принципы адаптивной оптики / М.А. Воронцов, В.И. Шмальгаузен. – Москва: Наука, 1985. – 336 с.
- [3] Roggemann, M.C. Imaging Through Turbulence / M.C. Roggemann, B.M. Welsh, B.R. Hunt. – CRC Press, 1996. – 320 p.
- [4] Гурвич, А.С. К методике определения структурной характеристики показателя преломления в атмосфере по изображениям радиальной миры / А.С. Гурвич, И.А. Старобинец, А.М. Черемухин // Изв. АН СССР. Физ. атмосф. и океана. – 1974. – Т. 10, № 4. – Р. 413-416.
- [5] Ochs, G.R. Refractive-turbulence profiles measured by one-dimensional spatial filtering of scintillations // Appl. Opt. – 1976. – Vol. 15(10). – P. 2504.
- [6] Fisher, M.J. The crossed-beam correlation technique / M.J. Fisher, F.R. Krause // J. Fluid Mech. – 1967. – Vol. 28(4). – P. 705.
- [7] Wang, T. Wind measurements by the temporal cross-correlation of the optical scintillations / T. Wang, G.R. Ochs, R.S. Lawrence // Appl. Opt. – 1981. – Vol. 20(23). – P. 4073.
- [8] Holmes, R.B. Passive optical wind profilometer / R.B. Holmes // US patent N 5.469.250 US005469250A USA. – 1995.
- [9] Belenkii, M. Passive crosswind profiler / M. Belenkii // US patent N 2010/0128136 A1. – 2010.
- [10] Porat, O. Crosswind sensing from optical-turbulence-induced fluctuations measured by a video camera / O. Porat, J. Shapira // Appl. Opt. – 2010. – Vol. 49(28). – P. 5236.
- [11] Hall, F.F. Wind measurement accuracy of the NOAA pulsed infrared Doppler lidar / F.F. Hall, R.M. Huffaker, R.M. Hardesty // Appl. Opt. – 1984. – Vol. 23(15). – P. 2503-2506.
- [12] Маракасов, Д.А. Оценка средней скорости ветра из корреляции смещений центров тяжести изображений некогерентных источников в турбулентной атмосфере / Д.А. Маракасов // Оптика атмосферы и океана. – 2016. – Т. 29, № 04. – С. 294-299.

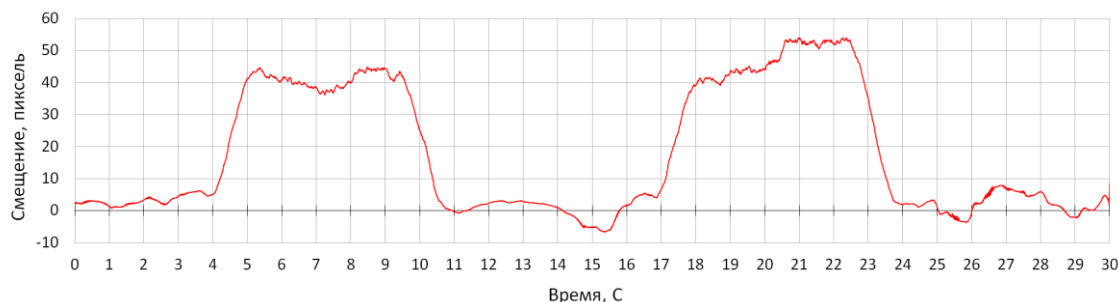


Рис. 1. График скорости смещения атмосферных неоднородностей на трассе в ходе выполнения лабораторного эксперимента

Application of NDVI and NDRE indices in the assessment of soybean productivity under controlled-release fertilizer

B. Boiarskii

All-Russia Scientific Research Institute of Soybean
Blagoveschensk, Russia
haomoris@gmail.com

Abstract—UAVs are becoming widespread in modern agriculture for data collection and measuring instruments for solving agriculture tasks such as field surveys and crop monitoring. Multispectral sensors and obtained spectral images make it possible to conduct an in-depth and high-precision analysis of the crops and fields. This study used vegetation indices to determine vegetation status and chlorophyll activity in soybean plants. Thereby finding out the difference in soybean vegetation between standard fertilizer and controlled-release fertilizer (CRF) applications. This study aimed to make a correlation between soybean productivity (yield, grain mass, and plant height) of two soybean varieties and two indices normalized difference vegetation index (NDVI) and normalized difference red edge (NDRE), depending on the applied fertilizers ratio. The research was carried out in an experimental field of the All-Russia Scientific Research Institute of Soybean, Amur Region, Russia. The analysis showed a significant increase in yield and a positive relation with vegetation indices on plots using CRF.

Keywords— control-released fertilizer, NDRE, NDVI, Russia, soybean, UAV.

1. INTRODUCTION

The first UAV (unmanned aerial vehicles) was created in 1916 by Americans [1]. UAVs for a long time were mainly used for military and civilian purposes. At the beginning of the 90s, drones were used for surface optical surveys and research [2]. Nowadays, in the world, UAVs and multispectral sensors are widely used in agriculture due to the continuous development of food quality and safety [3].

VIs (vegetation indices) applications are the most reliable instruments in agriculture across the world. VIs can provide data for use in field surveys, crop monitoring, yield predictions, crop status mapping, and detecting weeds, diseases, and nutrient deficiency (for example, nitrogen shortage in soybean plants) [3], [4]. Change in leaf mesophyll content of the healthy crop shows increasing the reflectance of wavelengths within the infrared channel and decreasing the red channel's reflectance. Thus, the ratio of different channels can be used in crop health assessment [5].

Precision agriculture in Russia was started to develop rapidly in recent years. The government, scientists, and farmers are increasingly aware of the importance of precision agriculture and the need for development in this direction as agriculture development expands significantly. In 2017, the government program "Digitalizing of Agriculture" of the Russian Federation Program was approved, which aimed to transform agriculture by introducing intelligent and precious agriculture and platform solutions to ensure a technological breakthrough in the country's agricultural sector [6].

2. MATERIALS AND METHODS

The tests were carried out in the southern zone of the Amur region, on the experimental field of the All-Russia Scientific Research Institute of Soybean. The zone belongs to a warm, relatively humid agro-climatic region of the Region. The crop Heat Units (CHU) indicator, the sum of temperatures accumulated above a base of 10 °C mean air temperature, is 2471 °C.

This study used the UAV model DJI Mavic Air 2, manufactured by the DJI company located in China. This UAV is a rotary-wing drone used for vertical take-off, hovering, and closer crop inspection. The survey was carried out at the height of 30 meters above the ground. The study used a multispectral camera, RedEdge-3, manufactured by Micasense, USA. The camera captured five spectral bands (red 668 nm, green 560 nm, blue 475 nm, NIR (infrared) 840 nm, and red-edge 717 nm). It provided NDVI, which is defined as a ratio of the difference between infrared and red channel bands to the sum of two bands. The camera provided NDRE, which was defined as a ratio of the difference between an infrared band and red-edge band to the sum of two bands instead of red.

3. RESULTS

This study used NDVI and NDRE to analyze vegetation's greenness, with the greenness of plant leaves determined by chlorophyll's concentration and plants' green pigment. Analysis of NDVI and NDRE values showed the highest index in the plots where the pigment concentration was high [4].

The plant density of the Zakat variety was more developed than that of the Nega-1 variety. However, the maximum greenness values were higher in the Nega-1 soybean variety. This remark indicated the different characteristics of each of the varieties. Therefore, in this experiment, the author calculated and compared the state of vegetation separately for each variety.

In this experiment, the author made correlations between soybean productivity (yield, grain mass, and plant height) of two soybean varieties and two indices, NDVI and NDRE. Fig. 1 shows the soybean variety of Zakat's productivity structure, which significantly increased plants' height by 10 cm and crop yield by 0.6 t/ha under CRF10. However, the soybean variety Zakat showed an adverse coefficient determination in terms of the mass of 1000 grains. The seed weight varied from 140 grams to 160 grams of 1000 grains. This gap is relatively small in this experiment and very sensitive to calculation. Therefore, in this case, the author pointed out the error in the calculations within the normal range. As a result, the correlation coefficient for plant height to indices was 0.74 for NDRE, 0.72 for NDVI,

seed weight was 0.53 for NDRE, 0.59 for NDVI, and yield 0.75 for NDRE, 0.77 for NDVI.

Fig. 2 represents the productivity structure for soybean variety Nega-1. The height of soybean plants increased slightly, by 2 cm and 5 cm for CRF5 and CRF10, respectively, compared to the control. The grain mass showed an increase of 10 grams for CRF5 and 20 grams for CRF10 compared to the control. The increase in yield was insignificant, with a rate of 0.48 t/ha and 0.16 t/ha for CRF5 and CRF10, respectively.

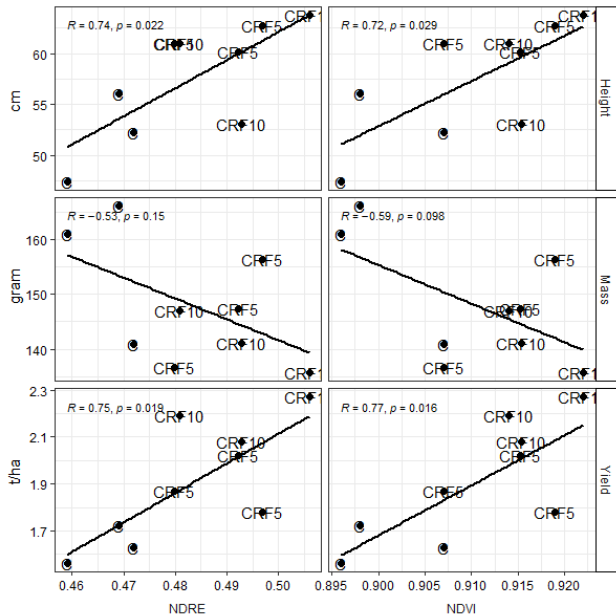


Fig. 1. Correlation analysis of vegetation indices, NDVI, NDRE vs. crop height, grain mass (1000 seeds) and crop yield for soybean variety Zakat

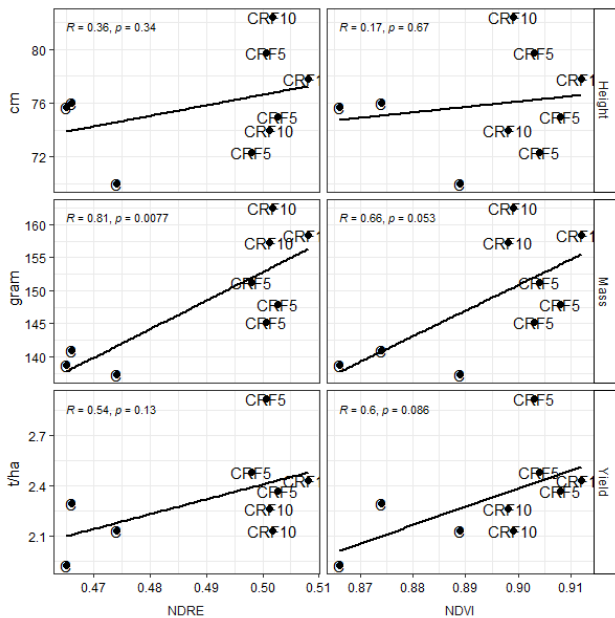


Fig. 2. Correlation analysis of vegetation indices, NDVI, NDRE vs. crop height, grain mass (1000 seeds) and crop yield for soybean variety Nega-1

4. CONCLUSION

This article described and demonstrated the application of two vegetation indices, NDVI and NDRE, on the experiment of soybean productivity assessment under different fertilizers, as a part of remote sensing tests. The result showed the distribution of index values depending on the applied fertilizers on the map obtained from the UAV based on multispectral images. The study observed that CRF did not significantly affect the soybean grain's quality during the yield analysis. However, there was a tendency to increase the fat content and decrease protein in soybean seeds of Zakat variety with CRF dose increasing. Conversely, there was an increase in protein and decrease fat in Nega-1 variety.

The study showed that CRF has a beneficial effect on the productivity of soybeans. Moreover, the author showed that the indices reflected chlorophyll activity plants since the release of nitrogen in the soil occurred at its active uptake by soybeans.

ACKNOWLEDGMENT

The author would like to thank the Academician of the Russian Academy of Sciences, Sinogovskaia Valentina Timofeevna, for the provided materials for this paper. The author declare that they have no conflict of interest.

REFERENCES

- [1] Nonami, K. Prospect and Recent Research & Development for Civil Use Autonomous Unmanned Aircraft as UAV and MAV / K. Nonami // J. Syst. Des. Dyn. – 2007. – Vol. 1(2). – P. 120-128. DOI: 10.1299/jsdd.1.120.
- [2] Baek, H. Design of Future UAV-Relay Tactical Data Link for Reliable UAV Control and Situational Awareness / H. Baek, J. Lim // IEEE Commun. Mag. – 2018. – Vol. 56(10). – P. 144-150. DOI: 10.1109/MCOM.2018.1700259.
- [3] Whelan, B.M. Downscaling for site-specific crop management needs? / B.M. Whelan, A.B. McBratney // Digital Soil Assessments and Beyond - Proceedings of the Fifth Global Workshop on Digital Soil Mapping. – 2012. – P. 353-356. DOI: 10.1201/b12728-69.
- [4] Berni, J.A.J. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle / J.A.J. Berni, P.J. Zarco-Tejada, L. Suárez, E. Fereres // IEEE Trans. Geosci. Remote Sens. – 2009 – Vol. 47(3). P. 722-738. DOI: 10.1109/TGRS.2008.2010457.
- [5] Carter, G.A. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration / G.A. Carter, A.K. Knapp // Am. J. Bot. – 2001. – Vol. 88(4). – P. 677-684. DOI: 10.2307/2657068.
- [6] Drobot, E. Prospects for digitalization of agriculture as a priority direction of import substitution / E. Drobot, M. Vartanova // J. Int. Econ. Aff. – 2018. – Vol. 8(1). – P. 1-18 (in Russian).

Разработка общей структуры метода компрессии изображений, основанного на машинном обучении

М.В. Гашников

Самарский национальный исследовательский университет им. академика С.П. Королева

Самара, Россия

mih-fastt@yandex.ru

Аннотация—Эта статья предлагает архитектуру основанного на машинном обучении (МО) метода для компрессии изображений и двумерных сигналов. Метод использует основанное на квадродереве представление двумерного сигнала. Это представление позволяет использовать характерные особенности различных этажей этого квадродерева для настройки МО-методов. Это представление также позволяет нам предсказывать отсчёты двумерного сигнала на основе менее детальных этажей квадродерева и кодировать отклонения предсказания. Мы используем МО-методы на всех этапах метода компрессии. В частности, мы используем основанные на МО алгоритмы сверхразрешения и встраивания цифровых двумерных сигналов, для предсказания отсчётов сигнала: свёрточные и генеративно-состязательные нейросети, а также автоэнкодеры. Мы проводим вычислительные эксперименты по исследованию эффективности основанного на МО предсказания на наиболее детальном этапе основанного на квадродереве представления двумерного сигнала при компрессии. Мы экспериментально доказываем, что исследуемый основанный на МО предсказатель отсчётов имеет существенное преимущество на реальных цифровых двумерных сигналах.

Ключевые слова— сверхразрешение, предсказание, двумерный сигнал, нейросети, архитектура метода.

1. ВВЕДЕНИЕ

Научная значимость и актуальность исследований в области сжатия цифровых изображений [1] с использованием методов машинного обучения (в частности, методов глубокого обучения) [2] обусловлена целым рядом факторов: ростом пространственного разрешения, качества цветопередачи и в целом качества восприятия визуальных данных, ростом объема хранимых и передаваемых визуальных данных в сетях общего пользования, ростом ожиданий пользователей относительно качества получаемой ими видеoinформации и т.д.

В данной работе разрабатывается структура основанного на машинном обучении (МО) метода компрессии изображений. Метод использует избыточный вариант основанного на квадродереве представления изображения [3], позволяющий учитывать при настройке МО структурные особенности изображения, проявляющие на различных масштабных уровнях. Такое представление позволяет интерполировать отсчёты изображения по более прореженным версиям того же самого изображения с последующим энтропийным кодированием постинтерполяционных остатков. Для интерполяции

используются алгоритмы [2, 4] сверхразрешения и доопределения цифровых изображений, основанные на машинном обучении: свёрточные (CNN) и генеративно-состязательные нейросети (GAN), а также автоэнкодеры. Использование МО позволяет существенно повысить эффективность отдельных этапов метода компрессии, что, в конечном итоге, приводит к существенному повышению эффективности всей технологии в целом.

2. МЕТОД КОМПРЕССИИ ИЗОБРАЖЕНИЙ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Предложенная структура основанного на МО метода компрессии изображений показана на рис. 1.

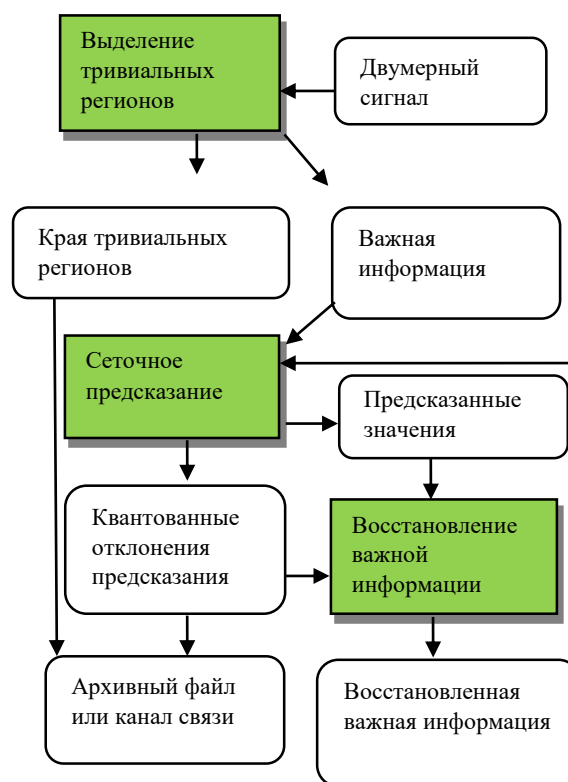


Рис. 1. Архитектура основанного на машинном обучении метода компрессии двумерных сигналов

Мы выделяем тривиальные регионы в двумерном сигнале. Мы используем основанную на МО сегментацию для этого. Тривиальные регионы (например, внутренняя часть озёр, морей, рек, лесов в снимках Земной поверхности и т.п.) не содержат важной информации. Мы помещаем в архивный файл или канал связи только края тривиальных регионов, так как мы можем использовать алгоритмы встраивания

изображений для заполнения этих регионов при декомпрессии. Мы архивируем только «важную информацию», которая расположена вне этих регионов.

Мы используем основанный на квадродере алгоритм [3] для компрессии регионов, содержащих «важную информацию». Мы представляем двумерный сигнал внутри «важных» регионов как квадродерево. Этажи квадродерева представляют из себя непересекающиеся части прореженного двумерного сигнала внутри «важных» регионов.

Мы компрессируем и сразу декомпрессируем этажи квадродерева последовательно, начиная с наименее детального этажа. Мы получаем искаженные (декомпрессированные) значения отсчетов в результате обработки каждого этажа квадродерева внутри регионов двумерного сигнала, содержащих важную информацию.

Мы предсказываем отсчеты каждого этажа квадродерева на основе отсчетов менее детальных этажей посредством нейросетей. Мы используем декомпрессированные значения этих отсчетов в качестве базы для этого предсказания.

Используемые свёрточные нейросети модифицируются посредством ортогонализации и параметризации настраиваемых свёрточных ядер в целях повышения устойчивости и сокращения времени обучения. Ортогонализация выполняется через внедрение специализированных штрафов на линейную зависимость свёрточных ядер в функцию потерь при обучении нейросети. Для параметризации свёрточных ядер применяется подход, заимствованный из методологии построения мультипараметрических вейвлетов.

3. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

Экспериментальный раздел настоящей работы посвящён исследованию эффективности нейросетевой интерполяции в рамках предлагаемого метода компрессии на натуральных изображениях (см. рис. 2).

В качестве нейросетевого интерполятора использован метод [5]. На рис. 3 показаны результаты интерполяции на наиболее детальном уровне основанного на квадродере представления изображения при компрессии. Видно, что исследуемый нейросетевой интерполятор имеет существенное преимущество над билинейным интерполятором.

4. ЗАКЛЮЧЕНИЕ

Предложена архитектура основанного на МО метода компрессии изображений. Метод использует алгоритмы сверхразрешения, встраивания и кластеризации, основанные на CNN и GAN. Вычислительные эксперименты подтвердили существенное преимущество исследованных алгоритмов на реальных изображениях.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда (проект 22-21-00662).



Рис. 2. Пример исходного натурального тестового изображения

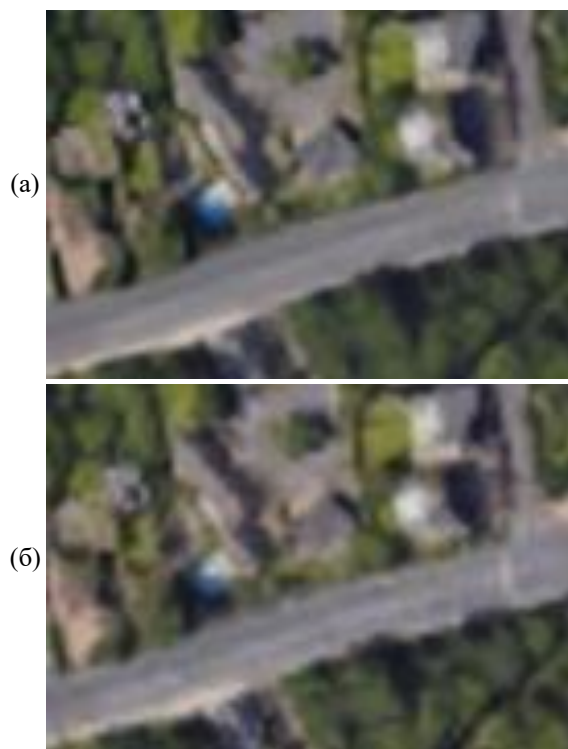


Рис. 3. Нейросетевая (а) и билинейная (б) интерполяция на наиболее детальном этапе основанного на квадродере представления изображения при компрессии

ЛИТЕРАТУРА

- [1] Зевалова, Л.А. Краткий обзор методов и алгоритмов компрессии зашумлённых изображений / Л.А. Зевалова, Ю.С. Бехтин // Современные научные исследования: теория, методология, практика. – 2018. – Т. 1. – С. 26-35. DOI: 10.26425/2658-3445-2018-1-26-35.
- [2] Коротеев, М.В. Обзор некоторых современных тенденций в технологии машинного обучения / М.В. Коротеев // E-Management. – 2018. – Т. 1, № 1. – С. 26-35. DOI: 10.26425/2658-3445-2018-1-26-35
- [3] Сергеев, В.В. Информационная технология компрессии изображений в системах оперативного дистанционного зондирования / М.В. Гашников, Н.И. Глузов, В.В. Сергеев // Известия Самарского научного центра РАН. – 1999. – № 1. – С. 99-107.
- [4] Гашников, М.В. Параметризованная интерполяция при комплексировании многомерных сигналов различного разрешения / М.В. Гашников // Компьютерная оптика. – 2020. – Т. 44, № 3. – С. 436-440. DOI: 10.18287/2412-6179-CO-696.
- [5] Wang, X. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data / X. Wang, L. Xie, C. Dong, Y. Shan // IEEE/CVF International Conference on Computer Vision. – 2021. – P. 1905-1914. DOI: 10.48550/arXiv.2107.10833.

Выбор методов машинного обучения для сжатия изображений

М.В. Гашников
Самарский национальный
исследовательский университет
им. академика С.П. Королева
Самара, Россия
mih-fastt@yandex.ru

Абстракт—Статья выполняет сравнение и выбор нейросетевых алгоритмов в рамках метода сжатия изображений, основанного на машинном обучении. Этот метод сжатия изображений использует свёрточные нейросети, автоэнкодеры и генеративно-состязательные нейросети в качестве базовых технологий. Мы сравниваем свёрточные нейросети и генеративно-состязательные нейросети при заполнении тривиальных областей изображений в рамках этого метода компрессии. Эти тривиальные области не содержат важных данных, поэтому мы не сохраняем эти области в архивный файл. Вместо этого мы с высоким качеством заполняем эти тривиальные области фейковыми данными при декомпрессии. Мы проводим вычислительные эксперименты по сравнению эффективности различных нейросетевых алгоритмов при заполнении тривиальных областей в натуральных изображениях в рамках задачи сжатия изображений. Мы оцениваем качество фейковых данных внутри тривиальных регионов как визуально, так и численно. Мы выбираем генеративную нейросеть, которая имеет преимущество и является перспективной для использования в рамках метода сжатия изображений, основанного на машинном обучении.

Ключевые слова— цифровое изображение, заполнение областей, тривиальные области, нейросети, компрессия изображений.

1. ВВЕДЕНИЕ

Научная значимость изысканий в области сжатия изображений [1] на основе алгоритмов глубокого машинного обучения [2] обусловлена актуальностью этих исследований в таких сферах, как телекоммуникации (широковещательная передача, организация удаленной работы, удаленное обучение и т.п.), игровая индустрия (интерактивные развлечения), здравоохранение (телемедицина) и др.

В данной работе выполняется сравнение и выбор конкретных нейросетевых алгоритмов в рамках метода сжатия изображений, основанного на машинном обучении (МО). Метод основан на использовании свёрточных нейросетей (CNN), автоэнкодеров и генеративно-состязательных нейросетей (GAN) [2, 3] в качестве базовых технологий.

Свёрточные нейросети, также как и генеративно-состязательные нейросети, в рамках этого метода используются для доопределения малоинформативных областей (являются обширные участки растительности или водных поверхностей и т.п.), которые довольно часто присутствуют на реальных изображениях. Такие области можно вообще не помещать в архивный файл, так как они

не несут важной информации, и могут быть с высоким качеством сгенерированы при декомпрессии.

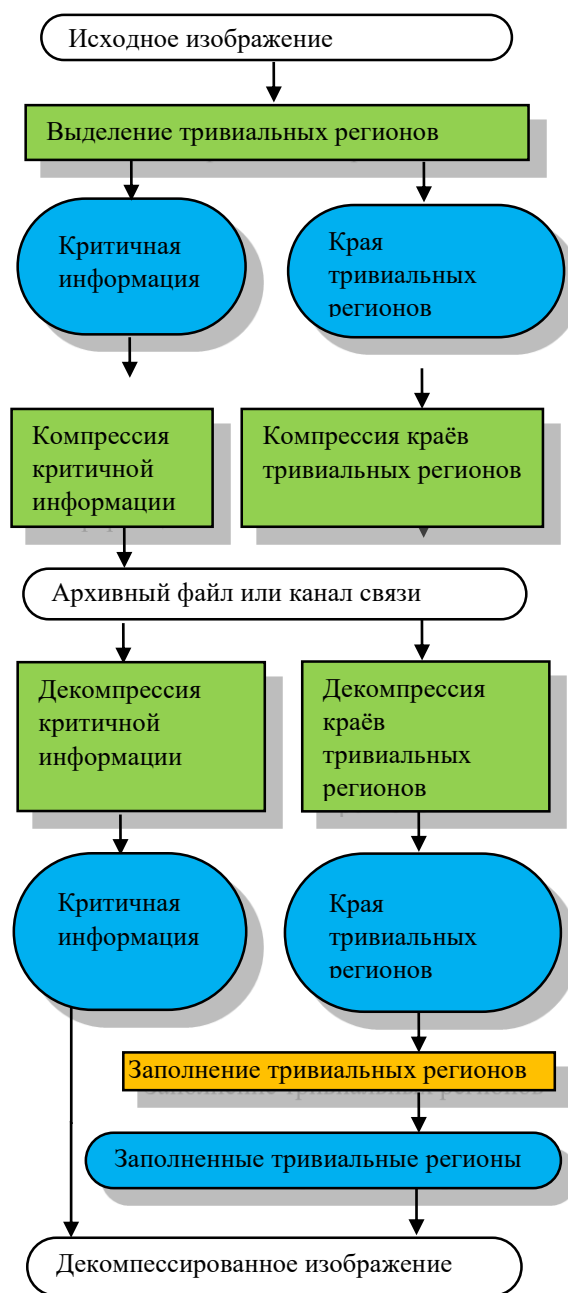


Рис. 1. Место основанного на машинном обучении алгоритма заполнения тривиальных регионов в технологии компрессии и декомпрессии цифровых изображений

2. СРАВНЕНИЕ И ВЫБОР МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В РАМКАХ МЕТОДА СЖАТИЯ ИЗОБРАЖЕНИЙ

Используемая архитектура основанного на машинном обучении метода сжатия показана на рис. 1.

В данной работе в рамках задачи сжатия сравниваются следующие доопределяющие изображения нейросети: свёрточная нейросеть [4], основанная на модификации UNet [2], в которой свёрточные слои заменены на слои частичной свёртки с частичной пакетной нормализацией, а также генеративно-состязательная нейросеть [5], основанная на тандемной композиции двух GAN, включающей специализированный механизм доопределения вытянутых структурных элементов.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Некоторые результаты показана на рис. 2, пример исходного изображения показан на рис. 3. Исследуемые в данной работе нейросети, доопределяющие изображения, сравнивались на изображениях, содержащих малоинформативные области.



Рис. 2. Сравнение нейросетевых алгоритмов доопределения малоинформативных областей изображений для сжатия: результаты доопределения посредством CNN и GAN

Задавалась некоторая, относительно крупная область на изображении, и информация внутри этой области заменялась подделкой с помощью двух сравниваемых нейросетей. Типичные результаты показаны на рис. 2, из которого видно, что исследуемая GAN более эффективна при доопределении малоинформативных областей.



Рис. 3. Пример исходного тестового изображения с контурами доопределяемых областей

4. ЗАКЛЮЧЕНИЕ

Выполнен выбор нейросетевых алгоритмов в рамках метода сжатия изображений, основанного на МО. Метод использует CNN, автоэнкодеры и GAN в качестве базовых технологий. Вычислительные эксперименты позволили выявить нейросеть, которая имеет преимущество и является перспективной для использования в рамках метода сжатия изображений.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда (проект 22-21-00662).

ЛИТЕРАТУРА

- [1] Володькин, М.Д. Методы сжатия данных в информационноизмерительных системах / М.Д. Володькин, А.В. Левенец // Информационные технологии XXI века. – 2019. – С. 126-130.
- [2] Аксютин, Е. М. Обзор архитектур и методов машинного обучения для анализа больших данных / Е.М. Аксютин, Ю.С. Белов // Электронный журнал: наука, техника и образование. – 2016. – Т. 1, № 5. – С. 134-141.
- [3] Кузнецов, А.В. Ретуширование данных дистанционного зондирования с использованием алгоритмов доопределения изображений в задаче генерирования подделок / А.В. Кузнецов, М.В. Гашников // Компьютерная оптика. – 2020. – Т. 44, № 5. – С. 763-771. DOI: 10.18287/2412-6179-CO-721.
- [4] Zhao, C. Inpainting to hide structures in satellite images [Electronical Resource] / C. Zhao. — Access mode: <https://github.com/ChenchaoZhao/NeuralCamouflage> (21.01.2021).
- [5] Nazeri, K. EdgeConnect: Generative image inpainting with adversarial edge learning / K. Nazeri, E. Ng, T. Joseph, F.Z. Qureshi, M. Ebrahimi // arXiv preprint. – 2019. –1901.00212.

Разработка макета оптического триангуляционного сканера для контроля геометрии резьбового соединения бурильных труб

Р.Р. Диязитдинов

Поволжский государственный университет телекоммуникаций и информатики

Самара, Россия

rinat.diyazitdinov@gmail.com

Аннотация—Контроль геометрии резьбового соединения бурильных труб представляет собой актуальную задачу для предприятий, специализирующихся на их выпуске и эксплуатации. Механический способ контроля с использованием калибров является точным, но не является универсальным. Каждый год проводится изменение номенклатуры резьбовых соединений, и существующие механические устройства контроля могут становиться непригодными для использования. Механические измерительные устройства имеют высокую стоимость, и их замена является сложной технической и организационной задачей. По этой причине в данном сегменте промышленного производства стоит проблема разработки универсального устройства контроля геометрии резьбового соединения бурильных труб. В работе представлено описание макета оптического триангуляционного сканера для решения этой задачи.

Ключевые слова— триангуляционный, сканер, геометрия, резьба, оптический, бурильная труба, макет.

1. ВВЕДЕНИЕ

Оценка геометрии резьбового соединения бурильной трубы является специфической задачей, в решении которой заинтересованы промышленные предприятия, выпускающие и эксплуатирующие их. Контроль геометрии интересен в первую очередь с точки зрения оценки прочностных характеристик (прогнозирование времени эксплуатации), а также с точки зрения оценки качества изготовления. В частности, наличие несоответствий в геометрии трубы может свидетельствовать о неверной работе или неполадках в резьбонарезных станках. Данная область производства в течение длительного времени являлась достаточно закрытой в силу своей специфики. Контроль резьбовых соединений производился с помощью соответствующих калибров, которые являются частью соответствующих механических измерительных установок.

Фактически, с помощью калибра проверяется, нет ли заедания резьбового соединения. Измерительные устройства такого типа являются стационарными, характеризуются большими габаритами и высокой ценой. Возрастающий номенклатурный ряд типов резьбовых соединений привел к тому, что промышленность, специализирующаяся на контроле геометрии, «не успевает» разрабатывать и поставлять измерительные устройства, способные проверять «новые» типы резьбовых соединений.

Следует отметить две проблемы, связанные с использованием подобных измерительных устройств:

- высокая стоимость (замена или закупка устройства является нетривиальной задачей);
- стационарность (невозможность использования установки в полевых условиях, вне цеха предприятия).

В связи с этим в последние несколько лет появилась задача разработки универсального устройства оценки геометрии резьбового соединения, которое не будет зависеть от формы резьбового соединения.

2. СУЩЕСТВУЮЩИЕ ПОДХОДЫ НА БАЗЕ ОПТИЧЕСКИХ ДАТЧИКОВ

В настоящее время параллельно ведется два направления:

- использование точечного датчика совместно с устройством шарико-винтовой пары;
- разработка специализированного триангуляционного сканера.

На Рис. 1 показана схема устройства на основе точечного датчика.

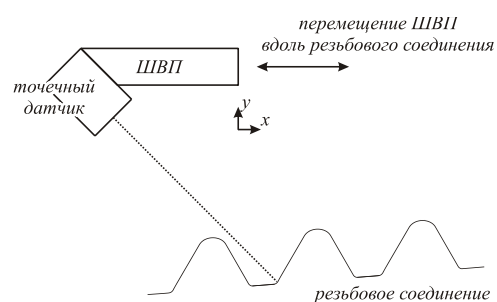


Рис. 1. Схема с точечным датчиком

Вдоль оси перемещения ШВП измеряется координата X. Координата Y определяется расстоянием, измеренным точечным датчиком и углом между системами координат точечного датчика и измерительного устройства.

В настоящий момент такие устройства являются внутренними разработками компаний, и результаты точностных характеристик подобных приборов еще не опубликованы в открытой печати. Однако оценить порядок погрешностей подобных приборов можно косвенным способом.

Точечный оптический датчик имеет разброс измерений в одной точке порядка $\Delta_{ОД} = 2 \dots 5$ мкм [1]. Катаные ШВП имеют класс точности с 7 по ISO DIN, и накопительная погрешность перемещения 52 мкм на длине 300 мм [2]. Средняя длина резьбового соединения составляет 100 мм, т.е. вклад в погрешность от ШВП составляет $\Delta_{ШВП} = 52 \cdot 100 / 300 \approx 17$ мкм. А общая погрешность: $\delta = (\Delta_{ОД}^2 + \Delta_{ШВП}^2)^{1/2} = 17..18$ мкм.

На Рис. 2 показана схема измерения с помощью триангуляционного сканера.

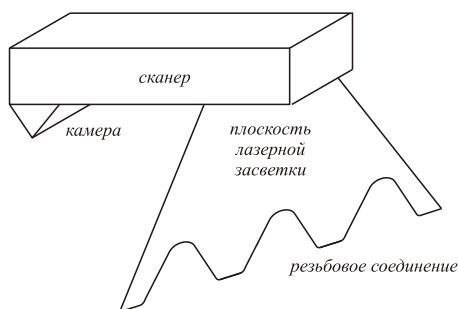


Рис. 2. Схема триангуляционного сканера

В настоящее время не существует готовых промышленных решений для измерения резьбовых соединений на базе триангуляционных сканеров.

Существующие решения [3] ориентированы на измерения достаточно крупных объектов (от нескольких сантиметров до метра) преимущественно с выпуклой формой поверхности. Такие поверхности отражают луч лазера в светочувствительную матрицу, не формируя переотражения.

Резьбовые соединения бурильных труб имеют шаг резьбы 5...10 мм и форма поверхности не является выпуклой. Это приводит к тому, что из-за переотражения помехи искажают профиль резьбы настолько сильно, что полученные данные нельзя использовать для оценки качества изготовления резьбы.

Для решения этой проблемы был разработан макет триангуляционного сканера, предназначенный именно для решения этой задачи.

3. МАКЕТ ТРИАНГУЛЯЦИОННОГО СКАНЕРА

При разработке макета триангуляционного сканера:

- рассматривались различные схемы расположения светочувствительной матрицы относительно лазера;
- выбиралась экспозиция;
- проверялись различные типы лазеров.

Основным результатом проведенных исследований является макет триангуляционного сканера. На Рис. 3 показано резьбовое соединение и лазерный луч на поверхности резьбы.

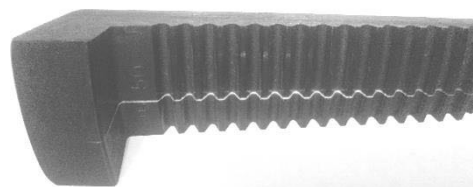


Рис. 3. Резьбовое соединение

На Рис. 4 показан результат совмещения измеренного профиля неизношенного резьбового соединения и эталонного профиля (примечание: для совмещения использовалась методика обработки [4]).

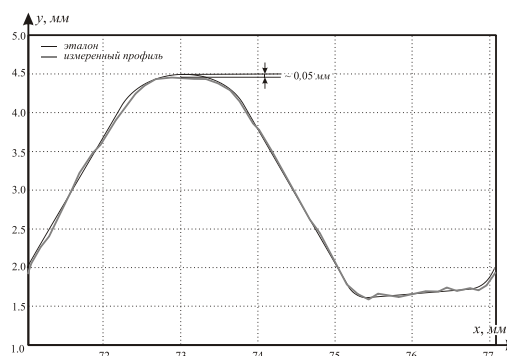


Рис. 4. Результат совмещения эталонного и измеренного профиля

Совмещение измеренного профиля с эталонным профилем показывает, что точность измерения составляет порядка 0,05 мм. Такая точность измерения позволяет проводить оценку качества изготовления резьбы, а макет может быть использован в дальнейшем для разработки устройства автоматического контроля геометрии резьбового соединения бурильных труб.

4. ЗАКЛЮЧЕНИЕ

В работе представлены материалы, посвященные разработке триангуляционного сканера для контроля геометрии резьбового соединения бурильной трубы. Разработанный макет позволяет проводить измерение с погрешностью порядка 0,05 мм. Исследования и разработанный макет могут быть интересны производителям измерительного оборудования, специализирующимся на данной области.

ЛИТЕРАТУРА

- [1] Sensors and Measurement Systems with More Precision [Электронный ресурс]. – Режим доступа: <https://www.micro-epsilon.ru/> (28.01.2022).
- [2] International Standard. ISO 3408-3. Second edition 2006-06-15. Ball screws – Part 3: Acceptance conditions and acceptance tests [Электронный ресурс]. – Режим доступа: https://infostore.saiglobal.com/preview/iso/updates2014/wk42/iso_3408-3-2006.pdf?sku=291826 (28.01.2022).
- [3] Photoelectronic Sensors [Электронный ресурс]. – Режим доступа: <https://www.wenglor.com/en/> (28.01.2022).
- [4] Васин, Н.Н. Обработка данных оптических триангуляционных сканеров для измерения профилей рельсов / Н.Н. Васин, Р.Р. Дязитдинов // Компьютерная оптика. – 2018. – Т. 42, № 6. – С. 1054-1061. DOI: 10.18287/2412-6179-2018-42-6-1054-1061.

Проективное совмещение телевизионных изображений многокамерной системы технического зрения по данным натурной видеосъемки

А.А. Диязитдинова

Поволжский государственный университет телекоммуникаций и информатики

Самара, Россия

alphiya.diyazitdinova@gmail.com

Аннотация—В работе представлен алгоритм проективного совмещения телевизионных изображений многокамерной системы технического зрения. Особенностью задачи является расчет параметров по данным натурной видеосъемки. Сложность задачи заключается в сопоставлении точек на изображениях. Применение метода полного перебора определяет огромное количество вариантов, время проверки которых не удовлетворяет требованиям на настройку систем видеонаблюдения. Для уменьшения количества вариантов были разработаны две процедуры, позволяющие уменьшить количество возможных вариантов. Эти процедуры легли в основу алгоритма, который позволяет оценивать параметры проективного совмещения за время, приемлемое для промышленных многокамерных систем видеонаблюдения.

Ключевые слова— проективный, совмещение, телевизионный, изображение, многокамерный, особая точка, корреляция, логарифмически-полярный.

1. ВВЕДЕНИЕ

Для решения специфических промышленных задач используются видеосистемы, которые называются многокамерными системами технического зрения. Такие системы нашли применение для исключения «слепых зон» на охраняемой территории и для повышения разрешающей способности областей изображения без потери информации обо всем поле зрения [1]. Неотъемлемой частью подобных многокамерных систем является процедура настройки. Если для настройки нельзя использовать тест-объект (часто используется «шахматная доска»), то возникает проблема сопоставления точек для оценки параметров совмещения. Для решения этой проблемы необходимо разрабатывать методы настройки по данным, снятым в ходе натурной видеосъемки, без использования тест-объектов. В представленной работе представлено исследование, посвященное решению данной задачи.

2. ОСНОВНЫЕ ЭЛЕМЕНТЫ РАЗРАБОТАННОГО АЛГОРИТМА СОВМЕЩЕНИЯ

Для совмещения была выбрана проективная модель. Она описывается выражениями:

$$x' = (h_{11}x + h_{12}y + h_{13}) / (h_{31}x + h_{32}y + 1) \quad (1)$$

$$y' = (h_{21}x + h_{22}y + h_{23}) / (h_{31}x + h_{32}y + 1) \quad (2)$$

Идея проективного совмещения телевизионных изображений базируется на том, чтобы вычислить особые

точки на изображениях, затем выбрать 4 произвольные точки на изображениях, и по ним оценить совмещение. В качестве решения использовать тот вариант, который обеспечит максимальное совпадение общего фрагмента по коэффициенту корреляции. Количество вариантов согласно методу полного перебора будет равно:

$$V = C_M^4 \cdot A_N^4, \quad (3)$$

где C – количество сочетаний, A – количество размещений, M – количество точек на первом изображении, N – количество точек на втором изображении.

При количестве особых точек порядка 1000, количество вариантов V будет порядка 10^{22} . Если проверка даже одного варианта будет составлять 0,01 с, то время обработки составит $3 \cdot 10^{13}$ лет.

Поэтому необходимо разрабатывать процедуры, позволяющие уменьшить количество вариантов путем предварительной отбраковки, чтобы время совмещения удовлетворяло требованиям по настройке систем видеонаблюдения [2].

Первая процедура отбраковки основана на возможном геометрическом расположении особых точек относительно друг друга.

Сформированные ограничения представлены ниже:

$$y_2 > y_1; \quad x_3 > x_1; \quad y_3 > y_1; \quad x_4 > x_1; \quad r_{13} \cdot m_1 < r_{12} < r_{13} \cdot m_2;$$

$$r_{13} \cdot m_1 < r_{14} < r_{13} \cdot m_2; \quad R_1 < r_{13} < R_2; \quad k_{14} < k_{13} < k_{12},$$

$$\text{где } r_{12} = ((x_1 - x_2)^2 + (y_1 - y_2)^2)^{1/2}, \quad r_{13} = ((x_1 - x_3)^2 + (y_1 - y_3)^2)^{1/2}, \quad r_{14} = ((x_1 - x_4)^2 + (y_1 - y_4)^2)^{1/2}, \quad k_{12} = (y_1 - y_2) / (x_1 - x_2), \quad k_{13} = (y_1 - y_3) / (x_1 - x_3), \quad k_{14} = (y_1 - y_4) / (x_1 - x_4).$$

Примечание: для изображения с размером 1280x720 пикселей использовались следующие параметры: $m_1=0,333$, $m_2=0,666$; $R_1 = 200$ пикселей, $R_2=500$ пикселей.

На Рис. 1 показано графическое представление ограничений.

Вторая процедура позволяет провести предварительное сопоставление особых точек. Если особые точки соответствуют одному и тому же фрагменту изображения, то и окрестность изображения этих точек также будут похожи. Однако они могут отличаться поворотом и масштабом. Для сравнения использовался прием представления изображений в логарифмически-

полярной системе координат [3], который обеспечивает инвариантность к повороту и масштабу.

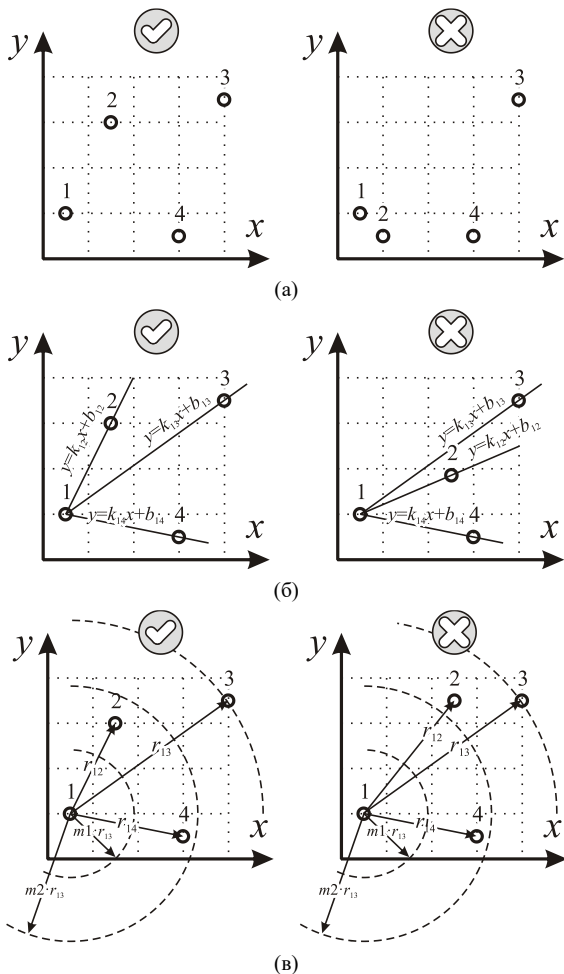


Рис. 1. Ограничения: (а) на взаимное расположение точек, (б) по углу, (в) по расстоянию

3. ЭКСПЕРИМЕНТ

Разработанный алгоритм был реализован в виде программы C++, тестирование проводилось на изображениях с разрешением 1280x720 пикселей, количество особых точек на изображениях по Harris detector составляло порядка 2000. Обработка на персональном компьютере с процессором Intel Core i7 занимала порядка 40-50 минут. На Рис. 2 показан пример совмещения.

Как можно видеть из Рис. 2б, разработанный алгоритм позволяет корректно совмещать изображения.

4. ЗАКЛЮЧЕНИЕ

Разработанный алгоритм проективного совмещения позволяет корректно обрабатывать изображения в автоматическом режиме без участия оператора. Алгоритм может быть внедрен на предприятиях, занимающихся

разработкой и внедрением промышленных систем для охраны территорий гражданского и стратегического назначения.



(а)



(б)

Рис. 2. Съемка периметра наблюдения системой, установленной на высотном здании: (а) исходные изображения, (б) результат совмещения

ЛИТЕРАТУРА

- [1] Дязитдинова, А.А. Повышение помехоустойчивости при оценке параметров проективного совмещения телевизионных сигналов / А.А. Дязитдинова // Физика волновых процессов и радиотехнические системы. – 2021. – Т. 24, № 1. – С. 58-66. DOI: 10.18469/1810-3189.2021.24.1.58-66.
- [2] EN 50132-7:2012 – Alarm systems – CCTV surveillance systems for use in security applications – Part 7: Application guidelines, 2012 [Электронный ресурс]. – Режим доступа: <https://cdn.standards.iteh.ai/samples/21916/0337918c36d24d739fb5c37af1035ed4/SIST-EN-50132-7-2012.pdf> (02.02.2022).
- [3] Мясников, Е.В. Определение параметров геометрических трансформаций для совмещения портретных изображений / Е.В. Мясников // Компьютерная оптика. – 2007. – Т. 31, № 3. – С. 77-82.

Применение методов интеллектуального анализа изображений при определении типов лиц

Ю.Ж. Пчелкина
Самарский национальный
исследовательский
университет им.
академика С.П. Королева
Самара, Россия
musina@yandex.ru

Р.А. Парингер
Самарский национальный
исследовательский
университет им.
академика С.П. Королева
Самара, Россия
rusparinger@gmail.com

П.Е. Савельева
Московский областной
научно-
исследовательский
клинический институт
им. М.Ф. Владимирского
Москва, Россия
gezulya76@yandex.ru

М.В. Егорова
Московский областной
научно-
исследовательский
клинический институт
им. М.Ф. Владимирского
Москва, Россия

Аннотация—Разработана и обучена модель нейронной сети для автоматизации процесса распознавания цефалометрических признаков лица по фотоснимкам. Проведен анализ точности обучения полученной модели. Подтверждена зависимость точности от объема, сбалансированности и качества данных.

Ключевые слова— распознавание изображений, нейронные сети, ортодонтия, цефалография.

1. ВВЕДЕНИЕ

Цефалометрический анализ — геометрическая абстракция, созданная ортодонтами и челюстно-лицевыми хирургами для дифференциации нормы от патологии посредством тригонометрических функций. Одним из пунктов обязательного осмотра врачом ортодонтом является антропометрическое и фотометрическое исследование головы.

Автоматизирование процесса распознавания типов лиц по фотоснимкам для заполнения обязательной описательной части фотопротокола [1] позволит снизить нагрузку на врача, сэкономить время осмотра и облегчить процесс заполнения обязательной медицинской документации.

2. ОПИСАНИЕ АНТРОПОМЕТРИЧЕСКОГО МЕТОДА В ОРТОДОНТИИ

При антропометрическом анализе выделяют ключевые цефалометрические измерительные точки головы.

Таблица I. НЕКОТОРЫЕ ИЗМЕРИТЕЛЬНЫЕ ТОЧКИ ГОЛОВЫ

Цефалометрические точки		
условное обозначение	наименование	описание
eu	эурион	латерально выступающая точка на голове (слева и справа)
go	гонион	наиболее выступающая кнаружи и кзади точка угла нижней челюсти (слева и справа)
n	назион	наиболее выраженное углубление между лбом и носом
zy	зигион	наиболее выступающая кнаружи точка скуловой дуги (слева и справа)
gn	гнатион	наиболее нижняя точка подбородка

Антропометрические параметры определяются рядом вертикальных, горизонтальных и угловых величин.

Таблица II. НЕКОТОРЫЕ ПАРАМЕТРЫ ИЗМЕРЕНИЯ ГОЛОВЫ

Заголовок столбца таблицы		
условное обозначение	наименование	описание
n-gn	назион-гнатион	морфологическая высота
eu-eu	эурион- эурион	расстояние между двумя латерально выступающими точками на голове
eu-go	эурион- гонион	касательные к боковому контуру головы
zy-zy	зигион- зигион	скуловая ширина - расстояние между наиболее выступающими кнаружи точками скуловых дуг

Форма лица может быть определена по морфологическому фациальному индексу Izard (по соотношению вертикальных и горизонтальных параметров).

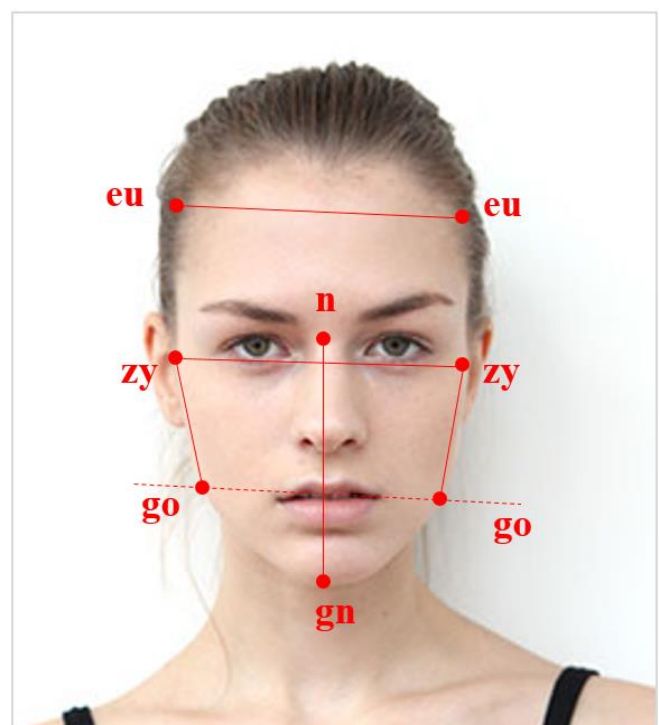


Рис. 1. Параметры для определения типа лица по признаку Izard

Рассчитав все необходимые характеристики, можем определить к какому типу по индексу Izard относится лицо на изображении:

- Прямоугольная форма – морфологическая высота лица ($n-gn$) превосходит его скуловую ширину (zuz), а касательные к боковому контуру головы ($eu-go$) параллельны между собой или конвергируют незначительно до 30 %.
- Квадратная форма – при равном или меньшем значении морфологической высоты по отношению к скуловой ширине, при параллельных касательных $eu-go$.
- Треугольная форма – при которой касательные $eu-go$ резко конвергируют книзу.

3. ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ

В качестве входных данных использовались фотографии из различных открытых источников в интернете, содержащие изображения людей в анфас. На первом этапе предобработки данных при реализации программного модуля для распознавания лиц на изображении использовалась предварительно обученная модель Хаара [2]. По полученным координатам прямоугольных областей путём обрезки исходных изображений были получены новые изображения. Все новые (вырезанные по контуру лиц) изображения, были преобразованы в оттенки серого, и масштабированы до одного размера. Расстановка контрольных точек на изображении лица являлось лишь вспомогательным инструментом, а не целью работы. Поэтому для локализации контрольных точек была использована готовая обученная модель [3], осуществляющая расстановку 68 контрольных точек.

После получения координат точек, были заданы уравнения соединяющих их прямых, определено расстояние между ключевыми точками ($n-gn$, zuz , $eu-go$), найдены углы наклона прямых zuz относительно прямой $go-go$ и, как следствие было определено к какому типу относится лицо на изображении.

Созданный набор данных был использован при обучении нейронной сети. Данные, на которых проводилось обучение, представляют собой набор пикселей в изображении со значениями от 0 до 255. Слишком широкий диапазон данных может отрицательно влиять на работу сети, поэтому предварительно была проведена нормировка данных.

Архитектура построенной модели нейронной сети представляет собой простую свёрточную нейронную сеть с тремя свёрточными слоями (включая входной слой). В качестве функции активации на всех слоях (и свёрточных и полносвязных) кроме выходного использовалась функция Relu (Rectified Linear Unit), как наименее затратная в плане времени вычислений. На выходном слое использовалась функция активации Softmax. Исключающий слой (Dropout), необходимый для предотвращения переобучения, удаляет случайным образом определенное количество (в данной модели на каждом шаге – 20 процентов) соединений между слоями.

Размер обучающей выборки составлял 75% от общего количества данных.

При наборе данных 1703 изображений на 250 эпохах обучения максимальная точность достигала 64,84%. После увеличения объема данных до 3318 изображений (из них 1584 – первого типа, 1278 – второго типа, 456 – третьего типа) на 250 эпохах обучения достоверность определения типа лица была уже выше 76,03%. В исходном наборе лиц третьего типа в выборке было значительно меньше, чем лиц двух других типов. Для того чтобы избежать влияния несбалансированности классов на ошибку обучения, набор данных был увеличен и включал в себя 4110 изображений лиц (из них 1584 – первого типа, 1278 – второго типа, 1248 – третьего типа). При обучении модели на новых данных при 250 эпохах обучения точность на тестовых данных достигала 92,71 процента, при 300 эпохах – 95,43%.

При оценке точности обучения модели стоит учитывать и качество исходных данных. Так при анализе ошибочно распознанных типов лиц было замечено, что неверно были отнесены к тому или иному классу изображения плохого качества; изображения, где лицо не расположено прямо (сильный наклон или поворот головы); изображения, на которых нечетко различима или неявно определяется контур лица.

4. ЗАКЛЮЧЕНИЕ

Применение интеллектуального анализа данных в медицине возможно для различных классов биомедицинских изображений [4].

Имея в качестве исходных данных достаточно большой и к тому же сбалансированный по типам лица набор изображений нужного качества (высокое разрешение, светлый однотонный фон, освещение, расположение лица точно в анфас и т.д.), возможно автоматизировать процесс распознавания типов лиц по фотоснимкам. Что в свою очередь поможет решить следующие задачи:

- помощь врачам в диагностике, профилактике и лечении зубочелюстных аномалий,
- сбор диагностических данных, получение необходимых цефалометрических признаков и анатомических ориентиров без участия врача (при помощи специалистов среднего звена);
- облегчение процесса заполнения обязательной медицинской документации, снижение нагрузки на врача и вспомогательный медперсонал.

ЛИТЕРАТУРА

- [1] Токаревич, И.В. Общая ортодонтия / И.В. Токаревич, Н.В. Корхова, И.В. Москалева, Л.В. Кипкаева, Т.В. Терехова, Д.В. Хандогий, Ю.Я. Наумович, С.С. Денисов. – Минск: БГМУ, 2015. – 219 с.
- [2] Jones, M. Robust Real-Time Face Detection / M. Jones, P. Viola // International Journal of Computer Vision. – 2004. – Vol. 57(2). – P. 137-154.
- [3] King, D.E. Dlib-ml: A Machine Learning Toolkit // Journal of Machine Learning Research. – 2009. – Vol. 10. – P. 1755-1758.
- [4] Ilyasova, N. Particular Use of BIG DATA in Medical Diagnostic Tasks / N. Ilyasova, A. Kupriyanov, R. Paringer, D. Kirsh // Pattern Recognition and Image Analysis. – 2018. – Vol 28(1). – P. 114-121.

Модификация алгоритмов машинного обучения для встраивания в методы компрессии изображений

Р.Р. Юзькив

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
yuzkiv@geosamara.ru

М.В. Гашников

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
mih-fastt@yandex.ru

Аннотация—Статья имеет дело с модификацией методов машинного обучения для встраивания в фреймворк компрессии изображений. Эта модификация приводит к необходимости изменения как структуры методов машинного обучения, так и алгоритмов настройки этих методов. При настройке методов машинного обучения мы используем показатель качества, релевантный задаче компрессии изображений. Мы разрабатываем алгоритм настройки бинарных классификаторов, а затем обобщаем этот алгоритм на ситуацию настройки произвольных алгоритмов машинного обучения при компрессии цифровых изображений. Мы описываем структуры данных, числовые массивы, и расчётные формулы, используемые в алгоритме настройки модифицированного метода машинного обучения. Мы встраиваем модифицированный основанный на машинном обучении метод в фреймворк компрессии цифровых изображений. Мы проводим вычислительные эксперименты по исследованию эффективности модифицированных методов машинного обучения при компрессии натуральных изображений. Вычислительные эксперименты подтверждают высокую эффективность модифицированных методов машинного обучения при компрессии цифровых изображений.

Ключевые слова— цифровое изображение, машинное обучение, компрессия изображений, показатель качества, классификатор.

1. ВВЕДЕНИЕ

Исследования алгоритмов машинного обучения (МО) в области компрессии изображений, безусловно, в настоящее время являются актуальными, так как потенциал алгоритмов машинного обучения в настоящее время раскрыт далеко не полностью, что в перспективе должно позволитькратно улучшить результаты распространённых в настоящее время технологий компрессии изображений.

В данной работе выполняется модификация алгоритмов машинного обучения для встраивания в методы компрессии изображений. При таком встраивании следует учитывать, прежде всего, специфику задачи компрессии [1], связанную с показателем качества. При компрессии наиболее естественным показателем качества является коэффициент сжатия, в то время как большинство алгоритмов машинного обучения [2] используют иные меры качества. Например, алгоритмы машинного обучения, основанные на деревьях решений (и деревьях регрессии, которые в рассматриваемом контексте не имеют отдельной специфики и поэтому далее рассматриваются полностью аналогично деревьям

решений), могут быть встроены в методы компрессии, основанные на интерполяции и экстраполяции. Однако деревья решений обычно используют такие показатели качества [3], как показатель Джинни, прирост информации и СКО.

В данной работе модификация алгоритмов машинного обучения для встраивания в методы компрессии изображений выполняется на примере деревьев решений, которые применяются для интерполяции изображений [4]. При этом используется показатель качества на основе энтропии кодируемых данных, использование которого, в свою очередь, влечёт за собой необходимость модификации алгоритма обучения [5], включающей специализированный алгоритм поиска разделяющей поверхности в каждом нетерминальном узле дерева.

Кроме того, мы встраиваем модифицированный основанный на машинном обучении метод в фреймворк компрессии [4], основанный на представлении изображения как квадродерева. Затем мы проводим вычислительные эксперименты по исследованию эффективности модифицированных методов машинного обучения при компрессии натуральных цифровых изображений. Вычислительные эксперименты подтверждают высокую эффективность модифицированных методов машинного обучения при компрессии цифровых изображений.

2. МОДИФИКАЦИЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КОМПРЕССИИ

Рассмотрим достаточно широкий класс методов компрессии цифровых изображений, основанных на предсказании пикселей изображения по соседним пикселям того же самого изображения. Этот класс включает методы компрессии изображений, основанные на дифференциальной импульсно-кодовой модуляции, иерархической сеточной интерполяции, и т.д.

Рассмотрим задачу встраивания методов машинного обучения в основанный на предсказании фреймворк компрессии изображений. Мы используем для предсказания некоторый классификатор C , который выбирает номер $c_{i,j}$ предсказанного значения $\hat{b}_{i,j}^c$ в каждом пикселе изображения $b_{i,j}$ с координатами (i, j) :

$$\hat{b}_{i,j}(\vec{w}) = \hat{b}_{i,j}^c, \quad c_{i,j} = C(\vec{u}_{i,j}, \vec{w}), \quad (1)$$

где \bar{w} — настраиваемый параметр классификатора, $\bar{u}_{i,j}$ — функция классификатора, $\hat{b}_{i,j}(\bar{w})$ — прогнозируемое значение.

Мы вычисляем набор предсказанных значений $\{\hat{b}_{i,j}^k\}$ посредством некоторых предсказателей $\{\hat{B}^k(\bar{b}'_{i,j})\}$:

$$\{\hat{b}_{i,j}^k, 1 \leq k \leq K\} = \{\hat{B}^k(\bar{b}'_{i,j}) : 1 \leq k \leq K\}, \quad (2)$$

где K количество предикторов, $\bar{b}'_{i,j}$ массив базовых пикселей для предсказания в координатах (i, j) .

Мы вычисляем функцию классификатора $\bar{u}_{i,j}$ аналогично посредством векторнозначной функции $\hat{U}(\bar{b}'_{i,j})$:

$$\bar{u}_{i,j} = \hat{U}(\bar{b}'_{i,j}). \quad (3)$$

После предсказания пикселя фреймворк компрессии изображений вычисляет предсказательное отклонение $(b_{i,j} - \hat{b}_{i,j}(\bar{w}))$ и переквантовывает его:

$$\delta_{i,j}(\bar{w}) = \lfloor b_{i,j} - \hat{b}_{i,j}(\bar{w}) \rfloor, \quad (4)$$

где $\lfloor \dots \rfloor$ это переквантование, а $\delta \in [1, \Delta]$ это переквантованное предсказательное отклонение.

Затем фреймворк компрессии изображений кодирует это переквантованное предсказательное отклонение и записывает его в архивный или канал связи. Мы настраиваем классификатор C при компрессии каждого изображения $b_{i,j}$ для вычисления настраиваемого параметра \bar{w} . Мы настраиваем этот классификатор исходя из минимизации размера сжатых данных при компрессии изображений.

Мы используем энтропию переквантованного предсказательного отклонения как оценку размера сжатых данных. Вместо минимизации энтропии мы максимизируем «энтропию со знаком минус» g . Критерий при настройке методов машинного обучения при компрессии приобретает вид:

$$g(\bar{w}) = \sum_{\tilde{\delta}} n(\tilde{\delta}, \bar{w}) \ln n(\tilde{\delta}, \bar{w}) \rightarrow \max_{\bar{w}}, \quad (5)$$

где $n(\tilde{\delta}, \bar{w})$ - количество значений переквантованного прогностического отклонения $\delta_{i,j}(\bar{w})$, равное $\tilde{\delta}$:

$$n(\tilde{\delta}, \bar{w}) = \text{card}\{(i, j) : \delta_{i,j}(\bar{w}) = \tilde{\delta}\}. \quad (6)$$

Описанная модификация структуры и критерия методов машинного обучения позволяет нам встроить эти методы в фреймворк компрессии изображений.

3. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ МОДИФИЦИРОВАННЫХ АЛГОРИТМОВ

В данной работе выполнено экспериментальное исследование эффективности модифицированных алгоритмов МО, основанных применении деревьев решений для интерполяции, в рамках задачи компрессии изображений. Усреднённый по набору реальных изображений процентный выигрыш по энтропии кодируемых данных показан на рис. 1. Нетрудно видеть, что модифицированный алгоритм МО показывает существенный выигрыш.

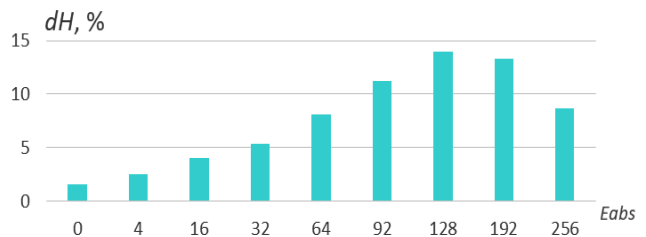


Рис. 1. Процентный выигрыш по энтропии модифицированного алгоритма машинного обучения на основе дерева решений в рамках задачи компрессии изображений

4. ЗАКЛЮЧЕНИЕ

Выполнена модификация методов МО для встраивания в фреймворк компрессии изображений. Модификация затронула как структуру методов МО, так и алгоритмы их настройки. Вычислительные эксперименты показали высокую эффективность модифицированных методов МО при компрессии.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда (проект 22-21-00662).

ЛИТЕРАТУРА

- [1] Тюльков, В.С. Обзор некоторых методов сжатия изображений / В.С. Тюльков // Молодежный научный форум. – 2019. – С. 27.
- [2] Вьюгин, В.В. Математические основы машинного обучения и прогнозирования / В.В. Вьюгин // Литрес, 2017. – 305 с.
- [3] Сулейманова, А.Н. Обзор развития алгоритмов деревьев решений / А.Н. Сулейманова // Социология: методология, методы, математическое моделирование. – 2020. – № 50-51. – С. 63-96.
- [4] Сергеев, В.В. Программная система компрессии и быстрого мультимасштабного просмотра изображений / В.В. Сергеев, М.В. Гашников, Н.И. Глузов, В.В. Мясников // VI Всероссийская с участием стран СНГ конференция “Методы и средства обработки сложной графической информации”. – 2001. – С. 66-67.
- [5] Гашников, М.В. Адаптивная интерполяция на основе оптимизации решающего правила в многомерном признаковом пространстве / М.В. Гашников // Компьютерная оптика. – 2020. – Т. 44, № 1. – С. 101-108. DOI: 10.18287/2412-6179-CO-661.

Correction of the interpolation effect in modeling the process of estimating image spatial deformations

R. Kovalenko
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
r.kovalenko.o@gmail.com

A. Tashlinskii
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
tag@ulstu.ru

Abstract—The paper proposed a technique for removing the influence of image interpolation artifact in the estimation algorithm of image spatial deformations. The technique is considered on examples of several similarity measures using bilinear and bicubic interpolation.

Keywords—simulation, interpolation, target function, similarity measures, spatial deformations.

1. INTRODUCTION

When developing algorithms to estimate the parameters of image spatial deformations [1], the study of the accuracy of developed algorithms and their performance is carried out, as a rule, on simulated images. Similar algorithms are used in determining the dynamic deformation of a scene from an image sequence [2], which is usually reduced to a sequential inter-frame estimation of deformations. In the study, one image $\mathbf{Z}=[z_{i,j}]$ (reference) can also be real, and the second $\tilde{\mathbf{Z}}=[z_{\tilde{i},\tilde{j}}]$ (deformed) is formed from it according to the given parameters of deformations using some kind of interpolation, where $z_{i,j}$ and $z_{\tilde{i},\tilde{j}}$ pixels of reference and deformed images, and (i,j) , (\tilde{i},\tilde{j}) coordinates of the regular grid on which they are given, in a single coordinate system. Then these given parameters are estimated under different conditions, in particular, the influence of noise. However, when forming a deformed image, interpolation makes changes in the correlation properties of the interpolated image, which affects the adequacy and correctness of the study.

2. PROBLEM STATEMENT

Let us assume that the origin of the coordinate coincides with the $(0,0)$. Then the coordinates (\tilde{i},\tilde{j}) of the nodes of the deformed image reference grid are determined by the deformation model adopted. In particular, when using the similarity model [3], which includes the parameters of shift $\vec{h}=(h_x,h_y)^T$, rotation angle φ and scale factor κ :

$$\tilde{i}=i\cos\varphi+j\sin\varphi+h_x,\tilde{j}=i\sin\varphi+j\cos\varphi+h_y. \quad (1)$$

The development of an algorithm for estimating the deformation parameters implies specifying some target function (TF) that characterizes the quality of estimation. Different similarity measures of images can be used as TF [4]. In this work, three most common similarity measures were used: mean square of inter-frame difference (MSID), inter-frame correlation coefficient (ICC), and mutual

information (MI) (in the given examples Shannon's MI was used [5,6]).

The pixel intensity $z_{\tilde{i},\tilde{j}}$ of the interpolated image can be predicted, as already noted, through some interpolation:

$$\tilde{z}_{\tilde{i},\tilde{j}}=\sum_{n=a}^c\sum_{m=b}^c a_{nm}z_{nm}, \quad (2)$$

where $a_{n,m}$ are coefficients determined by the type of interpolation [7]; (n,m) are coordinates of reference image pixels in the local coordinate system, the center of which (point $(n=0,m=0)$) corresponds to integer values of coordinates (\tilde{i},\tilde{j}) . The fractional (subpixel) component $(\Delta x,\Delta y)$ of these coordinates determines the $a_{n,m}$ coefficients. For $b=0,c=1$, we obtain bilinear interpolation, for $b=-1,c=2$, we obtain bicubic one. Thus, in bilinear interpolation: $a_{0,0}=(1-\Delta x)(1-\Delta y)$, $a_{0,1}=(1-\Delta x)\Delta y$, $a_{1,0}=\Delta x(1-\Delta y)$, $a_{1,1}=\Delta x\Delta y$.



Fig. 1. Satellite image of the Ulyanovsk state technical university

Due to the correlation of the reference image pixels, the inter-pixel correlation coefficient of the deformed image does not change at the points corresponding to the reference image nodes and increases at the points between the nodes. In this case, the maximum increase is achieved by a half shift of the counts grid along both coordinates. This leads to a distortion of the TF shape as a function of the deformation parameters. An example of such distortions when simulating a parallel shift of the image Fig. 1 along one of the coordinate axes is shown in Fig. 2. Fig. 2a corresponds to the use of MSID, Fig. 2b – ICC, Fig. 2c – MI. The red curve corresponds to the undistorted TF, the blue curve to the TF under bilinear interpolation, and the green curve to the bicubic interpolation. We see that interpolation affects different TFs variously. The largest distortions are noticeable at MSID, the smallest – at MI. This is explained by the fact that different TFs reflect

different characteristics of images. Namely, MSID shows brightness characteristics, ICC shows covariance one, MI shows entropic one. Note also that different deformation parameters also have various effects on TF. For example, at the rotation angle, the distortions of TF depend on the location of the rotation center. The distortions will be minimal when the center of rotation coincides with the center of the image.

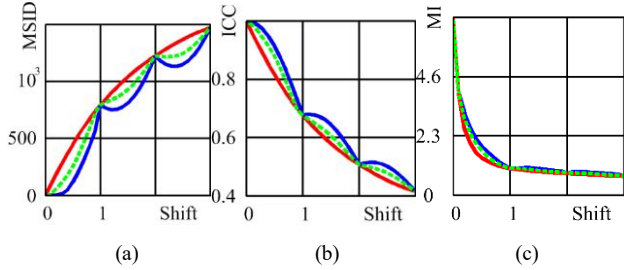


Fig. 2. TF when shifting the interpolated image

3. CORRECTION OF INTERPOLATION EFFECT

To compensate for the interpolation effect, we used the addition of independent Gaussian noise to the interpolated image pixels, the intensity of which is a function of the deformation parameters. This functional dependence can be found from the following assumptions. Denote the variance of the reference image by σ_z^2 . The pixel brightness at the node of the interpolated image is determined by the chosen type of interpolation, for example, (2), or other. To determine the variance of interpolated pixels it is necessary to use the same dependencies but take into account the correlation of the summands. For this purpose, the correlation function $R(l)$ of the reference image is used, where l is the distance between the corresponding pixels of the reference image. The value $\sigma_{\theta_{ij}}^2$ of the noise variance to be added to the pixels of the interpolated image can be found from the condition that the variances of all pixels $\{z_{\tilde{i},\tilde{j}}\}$ of the interpolated and reference images are equal.

$$\sigma_{\theta}^2(\tilde{i}, \tilde{j}) = \sigma_z^2(\tilde{i}, \tilde{j}) - \sigma_z^2. \quad (3)$$

Let us give an example for a simple shift deformation h_x of the image along the basic axis. Then the intensity at node (\tilde{i}, \tilde{j}) of the image under bilinear interpolation is determined by the expression:

$$z_{\tilde{i},\tilde{j}} = z_{i,j}(1 - \Delta h_x) + z_{i+1,j}\Delta h_x, \quad (4)$$

and at the bicubic:

$$z_{\tilde{i},\tilde{j}} = (2 - \Delta h_x)(1 + \Delta h_x)(z_{i,j}(1 - \Delta h_x) + z_{i+1,j}\Delta h_x)/2 - (\Delta h_x(1 - \Delta h_x))(z_{i-1,j}(2 - \Delta h_x) + z_{i-2,j}(1 + \Delta h_x))/6, \quad (5)$$

Accordingly, for the variance $\sigma_{\theta}^2(\tilde{i}, \tilde{j})$ with bilinear interpolation we obtain

$$\sigma_{\theta}^2(\tilde{i}, \tilde{j}) = \sigma_z^2(\tilde{i}, \tilde{j}) - \sigma_z^2 = -2(\Delta h_x^2 - \Delta h_x)(1 - \rho(1))\sigma_z^2, \quad (6)$$

In particular, for the image of Fig. 1 the interpixel correlation coefficient is $R(l) = 0.678$, and the variance is 1212.83. Substituting these values into (6), we obtain the value for finding the variance of added Gaussian noise. The plots of the target functions similar to Fig. 2, obtained with correction, are shown in Fig. 3. The results show that the interpolation effect is corrected and practically unnoticeable.

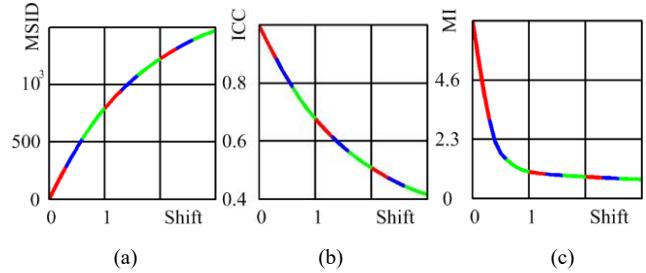


Fig. 3. TF after correction of the interpolation effect

4. CONCLUSION

The proposed method to correct the interpolation effect on TF quality for the problem of estimating the geometric deformation parameters of images has shown high efficiency. The method was considered by using bilinear and bicubic interpolation, for three similarity measures of images: MSID, ICC, and Shannon MI. However, it is not limited to these measures and is also applicable to others. Also, the method implies the possibility of using any interpolations: spline, using Lagrange polynomials, Newton, power functions and others.

ACKNOWLEDGMENT

The work was supported by the Russian Science Foundation according to research projects No. 22-21-00513, <https://rscf.ru/en/project/22-21-00513/>.

REFERENCES

- [1] Fida, A.D. Automated combination of optical coherence tomography images and fundus images / A.D. Fida, A.V. Gaidel, N.S. Demin, N.Y. Ilyasov, E.A. Zamytskiy // Computer Optics. – 2021. – Vol. 45(5). – P. 721-727. DOI: 10.18287/2412-6179-CO-892.
- [2] Tashlinskii, A.G. The convergence rate optimization of geometrical image inter-frame transformations parameters at recurrent estimation / A.G. Tashlinskii, G.L. Safina // J. Phys.: Conf. Ser. – 2018. – Vol. 1096. – P. 012032. DOI: 10.1088/1742-6596/1096/1/012032.
- [3] Gonzalez, R.C. Digital image processing / R.C. Gonzalez, E. Woods. – London: Pearson, 2018. – 1022 p.
- [4] Tashlinskii, A.G. Similarity and dissimilarity measures as objective function in image registration / A.G. Tashlinskii, S.V. Voronov // SWorld Journal. – 2015. – Vol. J11510. – P. 32-47.
- [5] Shanon, C.E. The mathematical theory of communication / C.E. Shanon, W. Weaver. – Urbana: The university of illinois press, 1994. – 132 p.
- [6] Tashlinskii, A.G. Usage of mutual information as similarity measures for stochastic binding images / A.G. Tashlinskii, G.L. Safina, R.O. Kovalenko, R.M. Ibragimov // IEEE Conference Proceedings. – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649386.
- [7] Keys, R. Cubic convolution interpolation for digital image processing / R. Keys // IEEE transactions on acoustics, speech, and signal processing. – 1981. – Vol. 29(6). – P. 1153-1160.

Обобщение основанного на машинном обучении метода компрессии на гиперспектральные изображения

А.И. Максимов

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
aleksei.maksimov.ssau@gmail.com

М.В. Гашников

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
mih-fastt@yandex.ru

Аннотация—Статья имеет дело с задачей обобщения основанных на машинном обучении двумерных фреймворков компрессии на гиперспектральные данные. Количество компонент в этих данных обычно настолько велико, что мы можем рассматривать эти данные как трёхмерные массивы с полноценным третьим измерением. Однако архитектура основанных на машинном обучении фреймворков компрессии обычно существенным образом опирается на двумерность исходных данных. Поэтому задача обобщения основанных на машинном обучении фреймворков компрессии на гиперспектральные данные является нетривиальной и актуальной. Большинство компонент в таких данных обычно очень похожи друг на друга. Мы используем сильную похожесть этих компонент для обобщения основанных на машинном обучении фреймворков компрессии на гиперспектральные данные. Мы проводим вычислительные эксперименты по исследованию эффективности предложенного подхода к обобщению основанных на машинном обучении фреймворков компрессии на гиперспектральные данные. Экспериментальные результаты доказывают, что использование предложенного подхода позволяет существенно увеличить эффективность компрессии гиперспектральных данных.

Ключевые слова— гиперспектральные данные, машинное обучение, компрессия, обобщение, отклонение.

1. ВВЕДЕНИЕ

Сфера применения гиперспектральных изображений, в том числе данных дистанционного зондирования, постоянно расширяется. В этих условиях актуальной является задача обобщения основанных на машинном обучении (МО) методов компрессии на гиперспектральные изображения.

Основной спецификой гиперспектральных изображений [1, 2], безусловно, является высокая спектральная размерность, так как количество спектральных компонент может достигать нескольких сотен. Кроме того, динамические характеристики спектральных компонент значительно различаются, значительная часть этих компонент обычно сильно коррелирована, а остальные компоненты при этом зачастую зашумлены. Все эти факторы затрудняют как сжатие самих спектральных компонент [3-5], так и использование взаимосвязей между ними.

Перспективным решением (см. рис. 1) обозначенной проблемы, как на этапе компрессии спектральных компонент, так и на этапе использования межкомпонентных корреляций для повышения коэффициента сжатия, являются алгоритмы машинного

обучения, потенциал использования которых в области компрессии изображения ещё не полностью раскрыт в настоящее время.

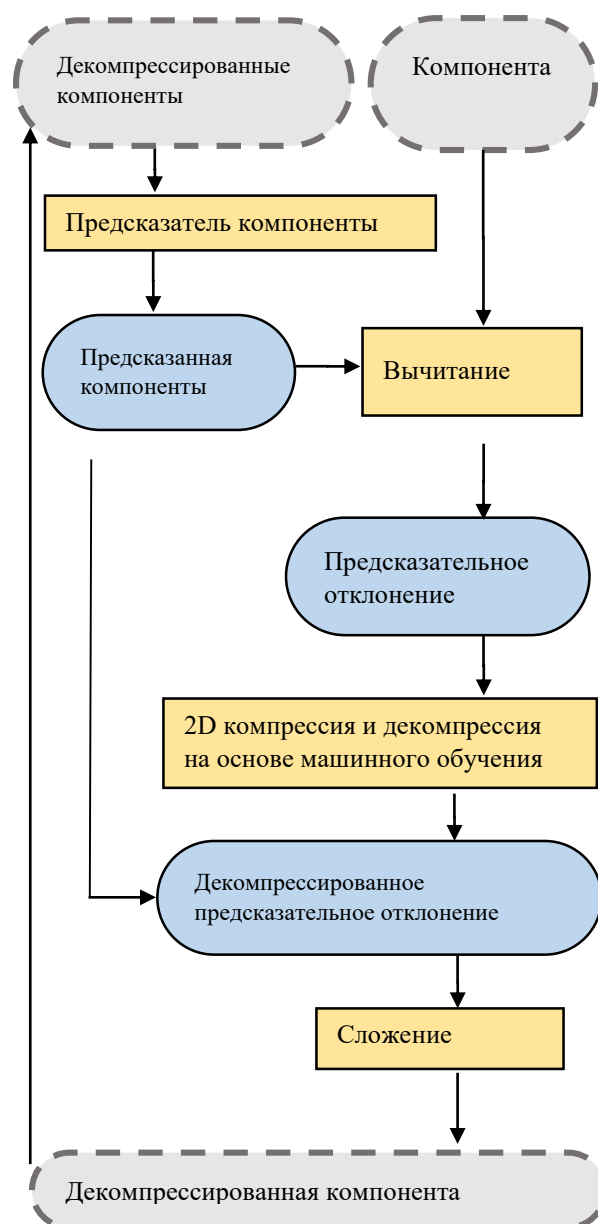


Рис. 1. Предсказатель компонент для компрессии гиперспектральных данных

В данной работе выполняется адаптация нейросетевых методов компрессии на случай гиперспектральных

изображений за счёт использования высокой корреляции между спектральными компонентами.

2. ОБОБЩЕНИЕ МЕТОДА КОМПРЕССИИ НА ГИПЕРСПЕКТРАЛЬНЫЕ ИЗОБРАЖЕНИЯ

Предлагаемый подход к обобщению основанных на МО методов компрессии на гиперспектральные изображения проиллюстрирован на рис. 1.

В этой статье мы исследуем задачу обобщения основанных на машинном обучении двумерных фреймворков компрессии на гиперспектральные данные. Количество компонент в этих данных обычно настолько велико, что мы можем рассматривать гиперспектральные данные как трёхмерные массивы с полноценным третьим измерением.

Однако архитектура основанных на машинном обучении фреймворков компрессии обычно существенным образом опирается на двумерность исходных данных. Поэтому задача обобщения основанных на машинном обучении фреймворков компрессии на гиперспектральные данные является нетривиальной и актуальной.

Большинство компонент в гиперспектральных данных обычно очень похожи друг на друга. Мы используем сильную похожесть компонент для обобщения основанных на машинном обучении фреймворков компрессии на гиперспектральные данные. Мы разрабатываем эффективный предсказатель гиперспектральных этажей, который позволяет нам компрессировать предсказательное отклонение вместо исходной компоненты.

Разброс значений этого предсказательного отклонения гораздо меньше, чем разброс значений исходной компоненты, поэтому степень компрессии предсказательного отклонения гораздо больше, чем степень компрессии исходной компоненты. Такой подход позволяет нам обобщить основанные на машинном обучении фреймворки компрессии на гиперспектральные данные.

3. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Проведены вычислительные эксперименты по использованию высокой межкомпонентной корреляции гиперспектральных изображений для повышения эффективности методов компрессии (см. рис. 2).

Использование межкомпонентной корреляции выполнялось за счёт использования специализированного аппроксимационного алгоритма, осуществляющего переход к разностному представлению компонент, основанному на вычислении постаппроксимационных остатков, возникающих при аппроксимации спектральных компонент.

Вычислительные эксперименты на реальных гиперспектральных данных дистанционного зондирования показали, что указанный аппроксимационный алгоритм позволяет существенно улучшить коэффициент компрессии, что позволяет сделать вывод о его эффективности при решении задачи обобщения нейросетевых методов компрессии на гиперспектральные изображения.

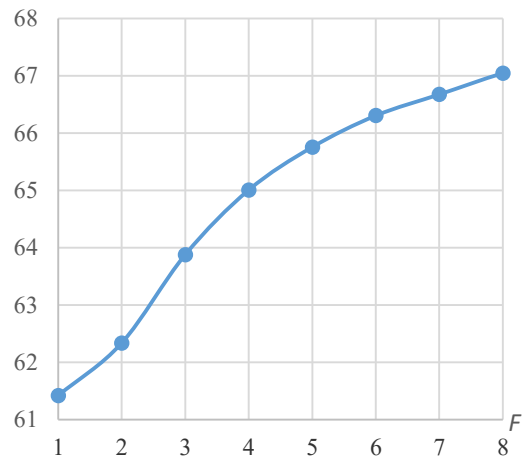


Рис. 2. Усреднённый процентный выигрыш при компрессии натуральных гиперспектральных данных, полученный благодаря использованию предсказателя гиперспектральных компонент (F – количество опорных компонент предсказателя)

4. ЗАКЛЮЧЕНИЕ

Основанные на МО двумерные фреймворки компрессии обобщены на гиперспектральные данные. Для обобщения использована сильная похожесть спектральных компонент. Экспериментальные результаты доказали, что использование предложенного подхода позволяет существенно увеличить эффективность компрессии гиперспектральных данных. Следовательно, применение предсказателя компонент является перспективным подходом к обобщению основанных на машинном обучении фреймворков компрессии на гиперспектральные данные.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда (проект 22-21-00662).

ЛИТЕРАТУРА

- [1] Бурдуковский, С.О. Обзор литературы по методам машинного обучения при обработке изображений / С.О. Бурдуковский // Приоритетные направления инновационной деятельности в промышленности. – 2021. – С. 154-155.
- [2] Фирсов, Н.А. Нейросетевая классификация гиперспектральных изображений растительности с формированием обучающей выборки на основе адаптивного вегетационного индекса / Н.А. Фирсов, В.В. Подлипов, Н.А. Ивлиев, П.П. Николаев, С.В. Машков, П.А. Ишкин, Р.В. Скиданов, А.В. Никоноров // Компьютерная оптика. – 2021. – Т. 45, № 6. – С. 887-896. DOI: 10.18287/2412-6179-CO-1038.
- [3] Кузьмина, М.Г. Многослойные сети-автоэнкодеры в задачах анализа и обработки гиперспектральных изображений / М.Г. Кузьмина // Препринты Института прикладной математики им. М.В. Келдыша РАН. – 2021. – С. 28-21.
- [4] Васин, Д.Ю. Регулярные методы кодирования растровых изображений дистанционного зондирования Земли / Д.Ю. Васин // Графикон-конференция по компьютерной графике и зрению. – 2019. – Т. 29. – С. 152-158.
- [5] Минкин, А.С. Сжатие гиперспектральных данных методом главных компонент / А.С. Минкин, О.В. Николаева, А.А. Руссков // Компьютерная оптика. – 2021. – Т. 45, № 2. – С. 235-244. DOI: 10.18287/2412-6179-CO-806.

Extraction of convex hulls of metal microstructure objects from metallographic images

R. Magdeev
Ulyanovsk State Technical University
Ulyanovsk, Russia
radiktkd2@yandex.ru

G. Safina
National Research Moscow State University of Civil
Engineering
Moscow, Russia
minkinag@mail.ru

Abstract—A comparative analysis of the effectiveness of algorithms for convex hull identification of perlite spots on binary metallographic images is carried out. The adequacy and computational complexity of the Graham, Jarvis algorithms and the “fast convex hull” algorithm on images of micro-grinds of metal pipelines with various lifespan are investigated.

Keywords— metallographic image, convex hull, algorithm, perlite spot, microstructural characteristics, steel.

1. INTRODUCTION

One of the most important problem in the production and operation of low-carbon steel products is the control of compliance of these products with the required characteristics (mechanical properties, residual life, the possibility of usage in certain conditions, etc.), which is primarily provided by the characteristics of the steel itself [1]. One of the effective approaches to estimate the steel microstructural characteristics is a technique based on the analysis of metallographic images of material microplate [2, 3]. In turn, an important component of this technique is the construction of convex hulls (CH) of detected perlite grains, according to the parameters of which the microstructural characteristics of perlite steels, which are the main products of ferrous metallurgy, are determined. The effectiveness of identification of objects (for the considered problem – perlite spots) on binary images is considered, for example, in the paper [4].

When constructing CH, only the points lying on the boundary of the object whose hull is being detected are used to increase the algorithm performance. Many convex hull identification algorithms are known, in particular, the algorithms of Chan, Kirkpatrick, Melkman [5], but in practice, the algorithms based on the procedures of Graham [6], Jarvis [7] and the so-called “fast convex hull” algorithm (FCH) [8] have received the greatest prevalence. The paper investigates their comparative effectiveness in the construction of CH of perlite grains (also called perlite spots) from binary metallographic images of steel micro-grinds.

2. A BRIEF DESCRIPTION OF THE ALGORITHMS

A. The Graham algorithm

The main algorithm operators are:

1°. Determining the point c_{\min} on the object with the minimum coordinate on the ordinate axis (if there are several, then the one with the smallest value on the abscissa axis is selected).

2°. Ranking of points from the object boundaries in ascending order of the polar angle counterclockwise relative

to the point c_{\min} (if the polar angles for several points coincide, the furthest from c_{\min} is selected).

3°. Graham bypass, which is based on the concepts of «left» and «right» corners. As a result, the points that match with vertices of CH are highlighted. At the same time, vertices that have not passed the «right» corner test are not vertices of CH.

4°. Connecting the found vertices with a hull.

B. The Jarvis algorithm

The Jarvis algorithm, also known as the «gift wrapping» algorithm, is slightly simpler than the Graham algorithm and consists of the following basic operators:

1°. Determining the minimum point of the object (as in the Graham algorithm).

2°. Jarvis bypass, highlighting the points of the convex hull.

3°. Connecting the found points with a hull.

B. The “fast convex hull” algorithm

The “fast convex hull” algorithm consists of the following main steps:

1°. The choice of two extreme points of the spot having the largest and smallest values along the abscissa axis: the left c_l and right c_r , which are the vertices of the CH (if there are several points with the same values, any of them is selected).

2°. Constructing a straight line passing through the points c_l and c_r , and dividing the set of all points into two subsets: located above and below the line $c_l c_r$, respectively.

3°. Consideration of a subset of points located above the straight line $c_l c_r$. Selection of the point c_1^{ch} that is the furthest from the straight line (if there are several, then the one with the largest angle $\angle c_l^{\text{ch}} c_l c_r$ is selected). Such a point is recognized as a vertex of CH.

4°. Construction of vectors $\overrightarrow{c_l c_1^{\text{ch}}}$ and $\overrightarrow{c_r c_1^{\text{ch}}}$ and exclusion from further consideration of points located to the right of them (internal points of the triangle $c_l^{\text{ch}} c_l c_r$).

5°. Consideration of a subset of points located to the left of the line $c_l c_1^{\text{ch}}$, for which there is a point c_2^{ch} furthest from the line $c_l c_1^{\text{ch}}$ (similar to paragraph 3), which is recognized as a vertex of CH.

6°. For all subsequent formed subsets operations similar to paragraphs 4 and 5 are performed until there is not a single non-empty subset left.

7°. Similarly to operators 3-6, a subset of points located below the straight line c_1c_r is considered.

3. COMPUTATIONAL COMPLEXITY OF ALGORITHMS

The computational complexity of Graham algorithm does not depend on the number of detected vertices and is proportional to L_p :

$$W = L_p \log(L_p),$$

where L_p is the number of external points of the object perimeter.

The computer operating capacity for the Jarvis algorithm, unlike the Graham algorithm, depends on the number of the polygon vertices (on the shape of the pearlite grain) and is proportional to $L_p L_{CH}$, where L_{CH} is the number of common points of the grain and its convex hull, which in the worst case is equal to $(L_p)^2$.

The computational complexity of the “fast convex hull” algorithm is determined by the complexity of constructing all subsets. In the best case, the problem is divided into two equally powerful sub-problems, then the complexity of the algorithm is from $2L_p$ to $(L_p)^2$. The advantage of the “fast convex hull” algorithm is also the possibility of parallel calculations for all subsets of points.

4. RESEARCH RESULTS

The algorithms of Graham, Jarvis and FCH algorithm for CH identification were investigated on metallographic images of pearlite spots and binary images of simple shapes (from the collection of test binary images, from the Internet portal [9]). On test binary images of simple shapes, all algorithms showed adequate results, differing mainly in speed. On binary images of real objects – pearlite spots, obtained from images of microstructures of metal pipelines with different lifespan, the Jarvis algorithm and the FCH algorithm adequately detected pearlite spots. An example of the results is shown in Fig. 1a and Fig. 1b respectively. At the same time, errors of CH identification are characterized for the Graham algorithm. It is explained by the complex, sometimes chaotic structure of pearlite spots. A typical example of such an erroneous identification is shown in Fig. 1c. As for computational costs, the average operating time of the Graham algorithm is about 1.1 times less than that of the FCH algorithm. In turn, the Jarvis algorithm loses to the FCH algorithm by about 1.9 times in terms of speed. Thus, it is advisable to use the FCH algorithm to solve the problem.

After the convex hulls are detected, the geometric characteristics associated with them are calculated: the pearlite spot area, the CH area, the length of the spot perimeter, the length of the CH perimeter, and others, which are then used

to evaluate the microstructural parameters of pearlite-grade steels.

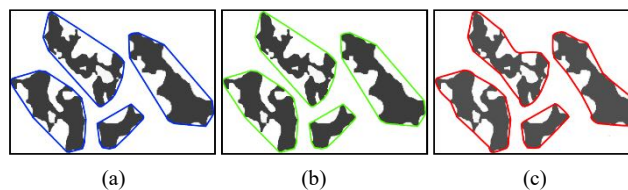


Fig. 1. An example of identification of convex hulls of pearlitic spots

5. CONCLUSION

When finding the microstructural characteristics of pearlite-grade steels from metallographic images of micro-grinds, the initial information are the parameters of the CH pearlite grains, which determine the important role of the quality of CH detection. A comparative study of the most used algorithms of Graham, Jarvis and the FCH algorithm on micro-grinds of steels with various lifespan has shown that for this problem CH are detected most adequately by the Jarvis and FCH algorithms. The Graham algorithm requires the least computational costs, however, with a complex configuration of spots, it makes errors in CH identification. The Jarvis algorithm is about 1.7 times inferior to the FCH algorithm, which makes the latter preferable when solving the problem of detecting pearlite grains from metallographic images of pearlite-grade steels.

ACKNOWLEDGMENTS

The work was supported by RFBR and Government of Ulyanovsk Region according to the research projects № 19-29-09048 and 19-47-730004.

REFERENCES

- [1] GOST R 54384-2011 Steel. Definition and classification by chemical composition and quality classes. – Moscow: Standartinform” Publisher, 2011. – 7 p. (in Russian).
- [2] Magdeev, R.G. Method for identification of pearlite-class steel microstructure parameters using metallographic images / R.G. Magdeev, A.G. Tashlinskiy // CEUR Workshop Proceedings. – 2017. – Vol. 1902. – P. 169-175. DOI 10.18287/1613-0073-2017-1901-169-175.
- [3] Magdeev, R.G. Identification of pearlite grains on metallographic images of mild steel / R.G. Magdeev, A.G. Tashlinskiy // Radiotekhnika. – 2018. – Vol. 6. – P. 33-37. (in Russian).
- [4] Magdeev, R.G. Efficiency of object identification for binary images / R.G. Magdeev, A.G. Tashlinskiy // Computer Optics. – 2019. – Vol. 43(2). – P. 277-281. DOI 10.18287/2412-6179-2019-43-2-277-281.
- [5] Kormen, T.H. Introduction to Algorithms / T.H. Kormen, C.E. Leiserson, R.L. Rivest, C. Stein. – Boston: MIT Press, 2009. – 1312 p.
- [6] Graham, R.L. An efficient algorithm for determining the convex hull of a finite planar set / R. . Graham // Information Processing Letters. – 1972. – Vol. 1. – P. 132-133. DOI: 10.1016/0020.
- [7] Jarvis, R.A. On the identification of the convex hull of a finite set of points in the plane / R.A. Jarvis // Information Processing Letters. – 1973. – Vol. 2. – P. 18-21.
- [8] Barber, C.B. The quickhull algorithm for convex hulls / C.B. Barber, D.P. Dobkin, H. Huhdanpaa // ACM Transactions on Mathematical Software. – 1996. – Vol. 22(4). – P. 469-483. DOI: 10.1145/235815.
- [9] Image Databases [Electronic resource]. – Access mode: http://www.imageprocessingplace.com/root_files_V3/image_database_s.htm (02.01.2021).

Вероятность ошибки и вычислительная сложность классификации объектов в пространстве многоуровневых представлений

М.М. Ланге
Федеральный исследовательский центр
«Информатика и управление» РАН
Москва, Россия
lange_mm@ccas.ru

С.В. Парамонов
Федеральный исследовательский центр
«Информатика и управление» РАН
Москва, Россия
psvpobox@gmail.com

Аннотация—Исследуется схема классификации объектов, заданных многоуровневыми древовидными представлениями, в терминах зависимости вероятности ошибки от количества обрабатываемой информации. Используя разделяющие функции на последовательных уровнях представления объектов, предлагается стратегия направленного поиска решения. Вводится избыточность вероятности ошибки относительно нижней границы, зависящая от параметра стратегии поиска решения. Экспериментальные оценки, полученные на множествах изображений лиц и подписей, демонстрируют разрыв между вычислительной сложностью и качеством классификации при различных значениях параметра стратегии поиска решения.

Ключевые слова— классификация, вероятность ошибки, взаимная информация, разделяющие функции, многоуровневое представление, вычислительная сложность.

1. ВВЕДЕНИЕ

В работе [1] получена аналитическая нижняя граница вероятности ошибки классификации на множестве объектов с заданной метрикой, и для любого набора разделяющих функций введена избыточность вероятности ошибки относительно нижней границы. Найденная граница является обобщением границы Шеннона для скорости кодирования символов дискретного источника с допустимой погрешностью в метрике Хемминга [2], когда символы источника передаются по каналу без памяти с искажениями. Приведенные в [1] численные оценки вероятности ошибки и избыточности получены для разделяющих функций экспоненциального типа на множествах древовидных представлений изображений лиц и подписей при переборном поиске решения.

В настоящей работе исследуются характеристики качества и вычислительной сложности классификации в пространстве древовидных представлений информативных объектов, заданных изображениями, с использованием разделяющих функций на всех уровнях представления. Для параметрической стратегии направленного поиска решения приводится оценка вычислительного выигрыша по сравнению с полным перебором. Предлагаемый подход может быть полезен для получения соотношения характеристик качества и быстродействия в схемах поиска приближенного ближайшего соседа на множестве объектов, заданных изображениями с высоким разрешением [3].

Исследуется схема классификации $\Omega \rightarrow \mathbf{X} \rightarrow \hat{\Omega}$, в которой Ω и $\hat{\Omega}$ – множества меток и их оценок для $c \geq 2$ классов по объектам из множества \mathbf{X} . При заданных априорных вероятностях на множестве Ω и условных по классам вероятностях на множестве \mathbf{X} , безусловные вероятности объектов образуют распределение $P = \{P(\mathbf{x}), \mathbf{x} \in \mathbf{X}\}$. Классификация выполняется в пространстве многоуровневых представлений, которое образовано набором $\mathbf{X}^L = \{\mathbf{X}_l\}_{l=1}^L$ представлений множества \mathbf{X} с нарастающим разрешением. В пространстве \mathbf{X}^L каждый объект $\mathbf{x} \in \mathbf{X}$ задан последовательностью представлений $\mathbf{x}^L = (\mathbf{x}_1, \dots, \mathbf{x}_L)$, в которой $\mathbf{x}_l \in \mathbf{X}_l, l = 1, \dots, L$ – поддерево глубины l в бинарном дереве \mathbf{x}_L глубины L [1].

2. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Используя квадратичную метрику на множествах представлений $\mathbf{X}_l, l = 1, \dots, L$, введены наборы разделяющих функций

$$G_l = \{g_{jl}(\mathbf{x}), \mathbf{x} \in \mathbf{X}\}_{j=1}^c, l = 1, \dots, L, \quad (1)$$

которые являются «мерами правдоподобия» оценок классов по предъявляемым объектам. Безусловное распределение P и условные по объектам распределения

$$Q_l = \{Q_l(j|\mathbf{x}) = g_{jl}(\mathbf{x}) / \sum_{i=1}^c g_{il}(\mathbf{x})\}_{j=1}^c, l = 1, \dots, L$$

вероятностей оценок меток классов позволяют ввести среднюю взаимную информацию $I_{G_l}(\mathbf{X}; \hat{\Omega})$ и среднюю вероятность ошибки $E_{G_l}(\mathbf{X}; \hat{\Omega})$ для разделяющих функций (1) на уровнях представления $l = 1, \dots, L$.

Для направленного поиска решения по объекту \mathbf{x} с последовательностью представлений \mathbf{x}^L предложена параметрическая стратегия направленного отбора меток классов с наибольшими значениями разделяющих функций на уровнях $l = 1, \dots, L$. Согласно такой стратегии, количество меток l -го уровня, среди которых отбираются наиболее правдоподобные по функциям (1)

метки $l+1$ -го уровня, определяется экспоненциально убывающей функцией

$$c_l = \left\lfloor c 2^{-\alpha(l-1)} \right\rfloor, l=1, \dots, L, \quad (2)$$

где $\alpha = (L-1)^{-1} \log(c/c^*) < 1$ и c^* – параметр, принимающий целые значения на отрезке $[1, c]$. На L -м уровне принимается решение j^* по наибольшему значению $g_{j^*L}(\mathbf{x})$ среди c^* меток, отобранных на $(L-1)$ -м уровне. В случае $c^* = c$ предлагаемая стратегия эквивалентна полному перебору.

С учетом структуры бинарных представляющих деревьев и функции (2), вычислительная сложность поиска решения по объекту в терминах количества обрабатываемых вершин в представляющем дереве удовлетворяет оценке

$$C_\alpha = \sum_{l=1}^L c_l 2^l \leq 2c \frac{2^{(1-\alpha)L} - 1}{2^{(1-\alpha)} - 1}. \quad (3)$$

Оценка (3) обеспечивает вычислительный выигрыш

$$\frac{C_{\alpha=0}}{C_{\alpha>0}} \geq \frac{2^L - 1}{2^L} \left(2 - (c/c^*)^{1/(L-1)} \right) c/c^* \quad (4)$$

направленного поиска решения по сравнению с полным перебором. При достаточно больших значениях $L \gg 1 + \log(c/c^*)$ правая часть в (4) составляет величину порядка c/c^* .

Семейство наборов разделяющих функций $G^L = \{G_l\}_{l=1}^L$ с весами $w_l = (\log c_l - \log c_{l+1}) / \log c$, $l=1, \dots, L$, где $c_1 = c$ и $c_{L+1} = 1$, порождает усредненные по уровням характеристики

$$I_{G^L} = \sum_{l=1}^L w_l I_{G_l}(\mathbf{X}; \hat{\Omega}) \quad \text{и} \quad E_{G^L} = \sum_{l=1}^L w_l E_{G_l}(\mathbf{X}; \hat{\Omega}),$$

которые зависят от параметра c^* . Указанным характеристикам соответствует избыточность

$$r_{G^L} = E_{G^L} - E_{\min}(I_{G^L})$$

средней вероятности ошибки E_{G^L} относительно значения нижней границы E_{\min} при среднем количестве обрабатываемой информации I_{G^L} .

Используя на множествах представлений изображений лиц [4] и подписей [5] слабые одномодовые разделяющие функции экспоненциального типа и композиции таких функций для ансамбля указанных источников, получены численные оценки характеристик I_{G^L} , E_{G^L} и $E_{\min}(I_{G^L})$ при различных значениях c^* . Данные источников содержали по 1000 объектов от $c = 25$ персон (классов), по 40 объектов в каждом классе. Для представлений объектов использованы бинарные деревья глубины $L=8$. Параметры разделяющих функций вычислены в режиме скользящего контроля по схеме «leave-one-out».

Графики указанных характеристик, вычисленных для ансамбля лиц и подписей в диапазоне значений $10 \leq c^* \leq 25$, даны на рисунке 1. Полученные

зависимости демонстрируют слабое увеличение вероятности ошибки и избыточности с уменьшением параметра c^* и, следовательно, с ростом вычислительного выигрыша, определенного в (4).

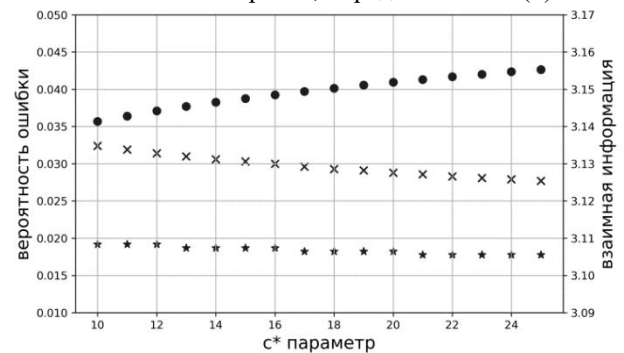


Рис.1. Численные оценки средней вероятности ошибки (x), средней взаимной информации (•) и нижней границы вероятности ошибки (*)

Приближение к нижней границе может быть достигнуто за счет использования многомодовых разделяющих функций для отдельных источников, а также путем увеличения количества источников в ансамбле.

3. ЗАКЛЮЧЕНИЕ

В рамках теоретико-информационной модели исследованы характеристики качества и вычислительной сложности классификации на множестве древовидных представлений объектов с многоуровневым разрешением. Показана возможность уменьшения времени принятия решения по предъявляемому объекту за счет некоторой потери качества классификации. Универсальность предложенного подхода позволяет применить его для анализа эффективности многоклассовых SVM классификаторов [6] в пространстве многоуровневых представлений с разделяющими функциями сигмоидального типа.

ЛИТЕРАТУРА

- [1] Lange, M.M. On a Lower Bound to Classification Error Probability in an Ensemble of Data Sources / M.M. Lange, S.V. Paramonov // IEEE Proceedings. – 2021. – P. 1-6. DOI: 10.1109/ITNT52450.2021.9649088
- [2] Gallager, R.G. Information Theory and Reliable Communication – Wiley & Sons, 1968. – 588 с.
- [3] Andoni, A. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions / A. Andoni, P. Indyk // Communications of the ACM. – 2008. – Vol. 51(1). – P. 117-122. DOI: 10.1145/1327452.1327494.
- [4] Distance matrices for face dataset [Electronic resource]. – Access mode: <http://sourceforge.net/projects/distance-matrices-face> (2020, June).
- [5] Distance matrices for signature dataset [Electronic resource]. – Access mode: <http://sourceforge.net/projects/distance-matrices-signature> (2020, June).
- [6] Sueno, H.T. Multi-class document classification using support vector machine (SVM) based on improved naïve Bayes vectorization technique / H.T. Sueno, B.D. Gerardo, R.P. Medina // International Journal of Advanced Trends in Computer Science and Engineering. – 2020. – Vol. 9(3). – P. 3937-3944. DOI: 10.30534/ijatcse/2020/216932020.

Application of renyi mutual information in stochastic referencing of multispectral and multi-temporal images

A. Tashlinskii
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
tag@ulstu.ru

R. Ibragimov
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
ibragimow.it@gmail.com

G. Safina
National Research Moscow State
University of Civil Engineering
Moscow, Russia
minkinag@mail.ru

Abstract—The efficiency of relay stochastic algorithms for referencing multispectral and multi-temporal images is considered when using Renyi mutual information as a measure of similarity of the referenced images. A comparison is made with the situation of using correlation similarity measures.

Keywords— image, referencing, image combination, mutual information, Renyi, stochastic algorithm, gradient, objective function.

1. INTRODUCTION

The necessity for image referencing arises in various applications: in the processing of medical and satellite images, the operation of autopilot systems, trajectory construction and many others [1,2]. One of the approaches to solve this problem is reduced to the problem of estimating the parameters of mutual geometric deformations of these images. The solution of the latter problem, in turn, is usually implemented iteratively as a search for the extremum of some multidimensional target function of evaluation quality in the space of referenced parameters [3]. The problem becomes more complicated when referenced multi-temporal and multispectral images, since in addition to geometric deformations caused by changes in the scene and the position of photodetectors, there are nonlinear brightness distortions in the images, the dependencies of which are not a priori known. This imposes additional strict requirements on the similarity measures of images, on the basis of which the target functions are constructed. Recently, information-theoretic [4, 5] has been increasingly used as such similarity measures. The paper investigates the effectiveness of using one of such measures – Renyi mutual information [6] – with stochastic non-identification image referencing.

2. DESCRIPTION OF ESTIMATION ALGORITHM

To synthesize an algorithm for estimating mutual geometric deformations of images $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ let us choose a recurrent relay stochastic procedure in the form [7]:

$$\hat{\mathbf{a}}_t = \hat{\mathbf{a}}_{t-1} \pm \Lambda_t \text{sign} \left(\boldsymbol{\beta} \left(J \left(\hat{\mathbf{a}}_{t-1}, \mathbf{Z}^{(1)}, \mathbf{Z}^{(2)} \right) \right) \right), \quad (1)$$

where $\hat{\mathbf{a}}$ is a vector of estimates of deformation parameters \mathbf{a} ; $\boldsymbol{\beta}(J(\bullet))$ is the stochastic gradient of target function $J(\bullet)$ of estimation quality; Λ is a learning matrix that determines the rate of change in grades during estimation on iterations; $t = \overline{1, T}$ is the number of iterations. A two-dimensional local sample Z_t of pixels $z_{j_t}^{(1)} \in \mathbf{Z}^{(1)}$, $z_{j_t}^{(2)} \in \mathbf{Z}^{(2)}$ with small size μ is used at each iteration. On the base of it sample the stochastic gradient of target function is calculated (in this problem it is

the gradient estimate), $\mathbf{j}_t \in \Omega_t$ is the coordinate vector of the element Z_t .

To find Renyi mutual information at the next iteration of the algorithm, it is necessary to estimate the probability density (PD) of images $\tilde{\mathbf{Z}}_t^{(1)}$ and $\mathbf{Z}^{(2)}$ on the base of a local sample Z_t , where $\tilde{\mathbf{Z}}_t^{(1)}$ is the oversampled image $\mathbf{Z}^{(1)}$ according to estimates $\hat{\mathbf{a}}_{t-1}$. At the same time, in order to reduce computational costs, the Parsen window method was used [8]. The PD estimate is found as a superposition of approximating functions $f(z - z_i)$ (elementary PD) of the same shape, centered on the brightness of all pixels z_i , $i = \overline{1, \mu}$ that are in the local sample:

$$p(z) = \mu^{-1} \sum_{z_i \in Z_t} f(z - z_i). \quad (2)$$

In the paper, a Gaussian function was used as an approximation function.

Renyi mutual information is found from Renyi entropy:

$$S = \left(H(\tilde{\mathbf{Z}}_t^{(1)}) + H(\mathbf{Z}^{(2)}) \right) / H(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) \quad (3)$$

where $H(\mathbf{Z}) = (1 - \alpha)^{-1} \lg \sum_i p_{z_i}^\alpha$ is the image entropy, $H(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) = (1 - \alpha)^{-1} \sum_i \sum_k p_{z_i, z_k}^\alpha$ is the joint image entropy, p_z and p_{z_1, z_2} are estimates of PD and joint brightness PD, $i, k = \overline{1, \mu}$ [9].

When usage of Renyi mutual information as the target function to calculate the stochastic gradient of MI, it is necessary to find its partial derivatives with respect to the estimated parameters. In this paper, analytical expressions for derivatives were found. So for the derivatives of the single and joint Shannon entropy according to the estimated parameters, we obtain expressions (4) and (5), respectively

$$\frac{\partial \hat{H}(\mathbf{Z}^{(k)})}{\partial \bar{\alpha}} = \sum_{z_i^{(k)} \in Z_b} \frac{\left(\sum_{z_j^{(k)} \in Z_a} \nabla_{ij}^k \right)^{(\alpha-1)} \left(\sum_{z_j^{(k)} \in Z_a} \nabla_{ij}^k \Delta_{ij}^k \frac{\partial \Delta_{ij}^k}{\partial \bar{\alpha}} \right)}{(1 - \alpha) \sigma^2 \sum_{z_i^{(k)} \in Z_b} \left(\sum_{z_j^{(k)} \in Z_a} \nabla_{ij}^k \right)^\alpha} \quad (4)$$

$$\frac{\partial \hat{H}(\mathbf{Z}^1, \mathbf{Z}^2)}{\partial \bar{\alpha}} = \sum_{z_i^{(k)} \in Z_b} \frac{\left(\sum_{z_j^{(k)} \in Z_a} \nabla_{ij}^1 \nabla_{ij}^2 \right)^{(\alpha-1)} \left(\sum_{z_j^{(k)} \in Z_a} \nabla_{ij}^1 \nabla_{ij}^2 \Delta_{ij}^1 \frac{\partial \Delta_{ij}^1}{\partial \bar{\alpha}} \right)}{(1-\alpha)\sigma^2 \sum_{z_i^{(k)} \in Z_b} \left(\sum_{z_j^{(k)} \in Z_a} \nabla_{ij}^1 \nabla_{ij}^2 \right)^\alpha} \quad (5)$$

where $\nabla_{ij}^{(1,2)} = f(z_i^{(1,2)} - z_k^{(1,2)})$; $\Delta_{ij}^{(1,2)} = z_i^{(1,2)} - z_k^{(1,2)}$; σ^2 is variance of $f(\bullet)$; $Z_{a,b} \in Z_t$ (in the experiments the local sample is divided equally); $k = 1, 2$.

In this case, the gradient of Renyi MI is

$$\boldsymbol{\beta} = \left(H^{-2}(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) H(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) \partial H(\tilde{\mathbf{Z}}_t^{(1)}) / \partial \bar{\alpha} - \left(H(\tilde{\mathbf{Z}}_t^{(1)}) + H(\mathbf{Z}^{(2)}) \right) \partial H(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) / \partial \bar{\alpha} \right). \quad (6)$$

3. THE RESULTS OF THE ALGORITHM

The approbation of the developed algorithm was carried out on multi-temporal and multi-spectral images. The results were compared with the results of the algorithm based on the meringue cross-correlation similarity measure [10] (hereinafter the correlation algorithm). The parameters of the similarity model, including shift, rotation angle and scale factor, were used as estimated. So, in Fig. 1a, for example, fragments of two different-time satellite images of the same domain are shown, and in Fig. 1b of two images of the same object in the visible and infrared ranges of the spectrum, having mutual geometric deformations, in particular a rotation of 9^0 . The results of estimating this parameter, averaged over 30 implementations, are presented in Fig. 2, where, for example, the dependences on the local sample size (LSS) of the variance of the estimated parameter (Fig. 2a) and the number of iterations required to achieve a steady-state evaluation mode are given. The steady-state mode is characterized by the fact that the change in the average value of estimates in a sliding window of some size does not exceed a specified threshold. In the same window, the variance of estimates of deformation parameters is also calculated.

4. CONCLUSION

It can be seen from the graphs that with a small μ the correlation algorithm shows the estimation failure and only, starting from $\mu=160$ the convergence of parameter estimates is observed. The proposed algorithm has stable convergence in the entire reduced range of LSS. In addition, it provides approximately 1.4 times less error of estimate parameters referencing, in the domain where the convergence of the correlation algorithm is achieved. It should also be noted that in the convergence domain, the correlation algorithm provides a slightly higher rate of convergence of deformation parameter estimates, which is explained by the fact that finding the numerical value of MI itself requires a larger sample size.

Thus, the usage of Renyi MI as the target function is advisable when image referencing of general morphology under conditions of uncertainty of large nonlinear brightness distortions.

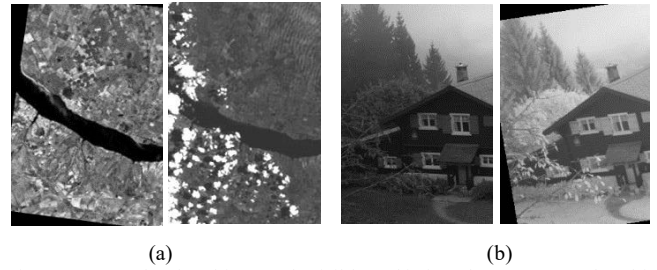


Fig. 1. Example of multispectral (visible and infrared spectrum) and multi-temporal images

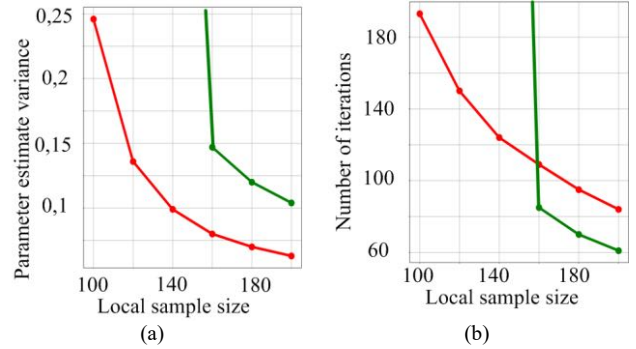


Fig. 2. The dependence of the number of iterations for convergence and variance on LSS

ACKNOWLEDGMENT

The work was supported by the Russian Science Foundation according to research projects No. 22-21-00513, <https://rscf.ru/en/project/22-21-00513/>.

REFERENCES

- [1] Kamaev, A.N. Visual navigation of an autonomous underwater vehicle based on the global search of image correspondences / A.N. Kamaev, D.A. Karmanov // *Computer Optics*. – 2018. – Vol. 42(3). – P. 457-467. DOI: 10.18287/2412-6179-2018-42-3-457-467.
- [2] Magdeev, R.G. Efficiency of object identification for binary images / R.G. Magdeev, A.G. Tashlinskii // *Computer Optics*. – 2019. – Vol. 43(2). – P. 277-281. DOI: 10.18287/2412-6179-2019-43-2-277-281.
- [3] Nag, S. Image Registration Techniques: A Survey / S. Nag // *Computer Vision and Pattern Recognition* [Electronic resource]. – Access mode: <https://arxiv.org/abs/1712.07540> (15.12.2021).
- [4] Chehade, W.E.H. Comparison of mutual information and its point similarity implementation for image registration / W.E.H. Chehade, P. Rogelj // *International Journal of Electrical and Computer Engineering (IJECE)*. – 2021. – Vol. 11(3). – P. 2613-2620.
- [5] Nan, A. DRMIME: Differentiable Mutual Information and Matrix Exponential for Multi-Resolution Image Registration / A. Nan, T. Matthew, R. Uriel, R. Nilanjan // *Software Jan 2020* [Electronic resource]. – Access mode: <http://arxiv.org/abs/2001.09865> (10.12.2021).
- [6] Pan, M. Medical Image Registration Based on Renyi's Quadratic Mutual Information / M. Pan, Z. Fen // *IETE Journal of Research*. – 2020. – P. 1-9.
- [7] Tashlinskii, A.G. Pseudogradient Estimation of Digital Images Interframe Geometrical Deformations / A.G. Tashlinskii // *Vision Systems: Segmentation & Pattern Recognition*. – 2007. – P. 465-494.
- [8] Mussa, H.Y. The Parzen Window method: In terms of two vectors and one matrix / H.Y. Mussa, J.B.O. Mitchell, A. M. Afzal // *Pattern Recognition Letters*. – 2015. – Vol. 63. – P. 30-35.
- [9] Rényi, A. On measures of entropy and information / A. Rényi // *Proc. Fourth Berkeley Symposium on Mathematical Statistics Probability*. – 2006. – Vol. 1. – P. 547-561.
- [10] Tashlinskii, A.G. Analysis of methods of estimating objective function gradient during recurrent measurements of image parameters / A.G. Tashlinskii, P.V. Smirmov, S.S. Zhukov // *Pattern Recognition and Image Analysis*. – 2012. – Vol. 22(3). – P. 399-405.

Оптимизация числа интервалов гистограммы, аппроксимирующей распределения вероятностей яркостей, при стохастическом совмещении изображений на основе взаимной информации

Р.О. Коваленко

Ульяновский государственный технический университет

Ульяновск, Россия

r.kovalenko.o@gmail.com

А.Г. Ташлинский

Ульяновский государственный технический университет

Ульяновск, Россия

tag@ulstu.ru

Аннотация—Проведен анализ влияния числа интервалов группирования гистограммы, аппроксимирующей плотность распределения вероятностей яркости, при совмещении изображений с использованием адаптивных стохастических процедур на основе взаимной информации Шеннона, Реньи и Тсаллиса.

Ключевые слова— совмещение изображений, стохастический градиент, взаимная информация, плотность распределения вероятности, гистограмма.

1. ВВЕДЕНИЕ

Совмещение изображений одна из востребованных процедур при их анализе. Она используется при решении прикладных задач обработки изображений и видеопотоков в различных областях [1, 2]. Так, медицинские снимки компьютерной томографии содержат нужную информацию не только в системе координат, но и в смысле морфологических и/или функциональных структур на изображении. Для повышения информативности изображения от различных видов визуализации совмещают, что позволяет получить информацию для лучшей клинической диагностики. Важную роль совмещение играет при обработке спутниковых изображений, когда нужно обеспечить точную географическую привязку или метрическую корректность.

Широкое применение при оценке параметров совмещения изображений нашли рекуррентные безыдентификационные процедуры стохастической аппроксимации [3], идея которой была предложена французским математиком Огюстеном Луи Коши ещё в 1847 году. Синтез таких процедур основан на разнообразных мерах сходства изображений [4], которые используются как целевые функции (ЦФ) качества оценивания параметров совмещения. Как правило, задача сводится к рекуррентному поиску экстремума этой ЦФ в пространстве параметров. Одним из видов таких мер подобия являются корреляционно-экстремальные, в частности, взаимная информация (ВИ), меры F-информации, энтропия совместной плотности распределения вероятностей и другие. Их использование при обработке изображений длительное время сдерживалось вычислительной трудоёмкостью. Однако в последние время с развитием вычислительных средств и новых подходов к развитию плотностей распределения вероятностей (ПРВ) яркостей изображений, они находят все большее распространение [5].

Ключевой составляющей при нахождении ВИ и её стохастического градиента по оцениваемым параметрам совмещения является оценка ПРВ яркостей совмещаемых изображений и их взаимной ПРВ. Одним из подходов к решению этой задачи является использование гистограмм яркостей изображений. Но при этом возникает проблема выбора числа интервалов группирования гистограммы. Если, например, яркость кодируется байтом, то максимально возможное число интервалов группирования 256. Гистограмма с таким числом разбиений, построенная по всему изображению, адекватно характеризует ПРВ. Однако в стохастических процедурах объем локальной выборки, которая используется для очередной итерации оценивания параметров совмещения, как правило, относительно небольшой – десятки отсчетов. В этом случае при большом количестве интервалов в часть из них попадет малое число отсчетов яркостей или не попадет вовсе, что негативно сказывается на качестве аппроксимации ПРВ. С другой стороны, чем меньше число интервалов, тем с меньшей адекватностью гистограмма отражает ПРВ. Таким образом, существует оптимальное число интервалов, которое, в свою очередь является предметно ориентированным, т.е. зависит от решаемой прикладной задачи.

Отметим, что вопрос оптимизации числа интервалов группирования, например, для задачи согласованности χ^2 статистик с χ^2 -распределением Пирсона еще несколько десятилетий назад исследовался в различных работах. Предложены различные правила нахождения оптимального числа интервалов, которые связаны и с объемом исходной выборки: Старджеса [6], Хейнхолда [7], Брукса-Каррузера, Манна-Вальда, Смирнова, Скотта, Таушанова, Тонева, Фридмана-Диакониса и другие. Для статистической обработки экспериментальных результатов рекомендации по выбору числа интервалов разработаны ВНИИ Метрологии имени Д.И. Менделеева [8]. Однако для задачи стохастического оценивания параметров совмещения изображения при использовании в качестве ЦФ теоретико-информационных мер подобия изображений сведения о подобных исследованиях не известны. Поэтому целью настоящей работы было исследование влияния числа интервалов группирования гистограмм яркостей изображений на эффективность стохастических процедур совмещения изображений при использовании в качестве мер подобия изображений ВИ.

2. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В исследованиях в качестве ЦФ использовались ВИ Шеннона [9], Реньи [10] и Тсаллиса [11], на базе которых синтезировались безыдентификационные процедуры релейного типа [12, 13]. В качестве оцениваемых параметров совмещаемых спутниковых изображений задавались вектор сдвига (h_x, h_y) , коэффициент масштаба k и угол поворота ϕ . Деформированное изображение моделировалось. Оптимальное число интервалов группирования гистограмм определялось по серии экспериментов с последующим сравнением с известными аналитическими правилами нахождения оптимального числа интервалов. Для характеристики точности совмещения использовалось евклидово расстояние рассогласования (интегральная мера точности вектора параметров в совокупности).

Для примера приведем результаты исследования спутникового изображения территории Ульяновского государственного технического университета (фрагмент представлен на рис. 1а) при параметрах деформации: $h_x = -2.2, h_y = 2.7, \phi = -1.5^\circ, k = 1.1$. Объем локальной выборки составлял 50 отчётов. Результаты усреднены по серии из 60 экспериментов. На рис. 2б приведены графики для ВИ Шеннона, рис. 2в - Реньи, рис. 2г - Тсаллиса. По оси ординат отложено число итераций, которое потребовалось для сходимости оценок параметров, по оси абсцисс - используемое число разбиений гистограммы для оценки ПРВ. Звёздочкой обозначено медианное число итераций, точкой - среднее, которое является центром прямоугольника, половина длины которого соответствует среднеквадратическому отклонению. Крестики снизу и сверху от прямоугольника соответствуют минимальному и максимальному значениям числа итерации сходимости.

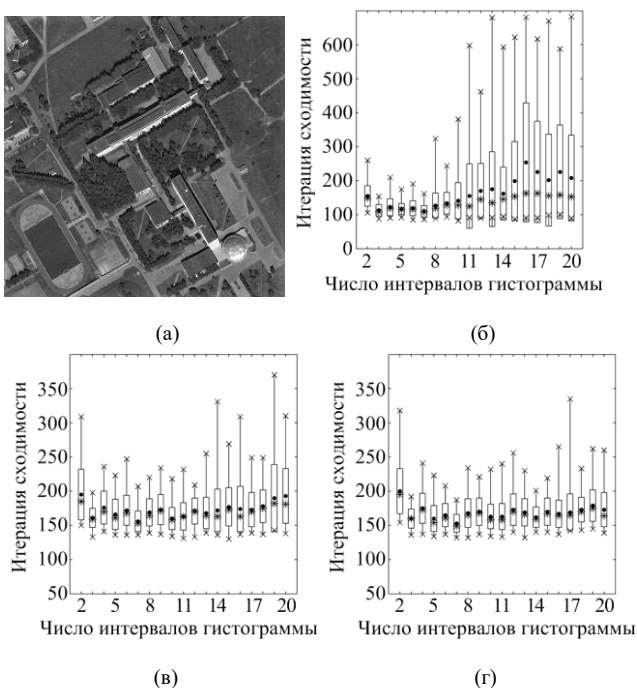


Рис. 1. Фрагмент исследуемого изображения и характеристики эффективности стохастической процедуры при использовании в качестве ЦФ различных видов ВИ

3. ЗАКЛЮЧЕНИЕ

Анализ результатов показал, что оптимальное число разбиения гистограмм яркостей, обеспечивающее наилучшую скорость сходимости параметров привязки существует. Так, для приведенного примера, при ВИ Шеннона и 7 интервалах разбиениями среднее время сходимости составило 110 итераций, медианное 108, а СКО составляет 17 итераций. Результаты исследований показали, что для рассматриваемой задачи могут быть рекомендованы правила нахождения оптимального числа интервалов Старджеса и Хейнхолда. Отметим также, что при аддитивном зашумлении изображений оптимальное число интервалов несколько уменьшается, а число итераций до сходимости параметров увеличивается. Изменение же в несколько раз объема локальной выборки слабо влияет на оптимальное число интервалов. По сравнению с ВИ Шеннона, ВИ Реньи и Тсаллиса более устойчивы к отклонению числа разбиений гистограммы от оптимального.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда № 22-21-00513, <https://rscf.ru/project/22-21-00513/>.

ЛИТЕРАТУРА

- [1] Fida, A.D. Automated combination of optical coherence tomography images and fundus images / A.D. Fida, A.V. Gaidel, N.S. Demin, N.Y. Ilyasov, E.A. Zamyatskiy // *Computer Optics*. – 2021. – Vol. 45(5). – P. 721-727. DOI: 10.18287/2412-6179-CO-892.
- [2] Magdeev, R.G. Efficiency of object identification for binary images / R.G. Magdeev, A.G. Tashlinskii // *Computer Optics*. – 2019. – Vol. 43(2). – P. 277-281. DOI: 10.18287/2412-6179-2019-43-2.
- [3] Ташлинский, А.Г. Оценивание параметров пространственных деформаций последовательностей / А.Г. Ташлинский. – Ульяновск: Издательство УлГТУ, 2000. – 132 с.
- [4] Tashlinskii, A.G. Similarity and dissimilarity measures as objective function in image registration / A.G. Tashlinskii, S.V. Voronov // *SWorld Journal*. – 2015. – Vol. J11510. – P. 32-48.
- [5] Chehade, W.E.H. Comparison of mutual information and its point similarity implementation for image registration / W.E.H. Chehade, P. Rogelj // *International Journal of Electrical and Computer Engineering (IJECE)*. – 2021. – Vol. 11(3). – P. 2613-2620.
- [6] Sturges, H.A. The choice of classic intervals / H.A. Sturges // *Journal of the American Statistical Association*. – 1926. – Vol. 21(153). – P. 65-66.
- [7] Heinhold, I. *Ingenieur statistic* / I. Heinhold, K.W. Gaede. – München: Springer Verlag, 1964. – 352 p.
- [8] Бурдун, Г.Д. Основы метрологии: учебное пособие / Г.Д. Бурдун, Б.Н. Марков. – М.: Изд-во стандартов, 1985. – 120 с.
- [9] Shannon, C.E. *The mathematical theory of communication* / C.E. Shannon, W. Weaver. – Urbana: The university of illinois press, 1994. – 132 p.
- [10] Wachowiak, M.P. Similarity metrics based on nonadditive entropies for 2D-3D multimodal biomedical image registration / M.P. Wachowiak, R. Smolikova, G.D. Tourassi, A.S. Elmaghaby // *Medical Imaging: Image Processing*. – 2003. – Vol. 5032. – P. 1090-1100.
- [11] Cvejic, N. Image fusion metric based on mutual information and Tsallis entropy / N. Cvejic, C.N. Canagarajah, D.R. Bull // *Electronics Letters*. – 2006. – Vol. 42(11). – P. 626-627.
- [12] Tashlinskii, A.G. Usage of mutual information as similarity measures for stochastic binding images / A.G. Tashlinskii, G.L. Safina, R.O. Kovalenko, R.M. Ibragimov // *IEEE Conference Proceedings*. – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649386.
- [13] Voronov, S.V. Efficiency analysis of information theoretic measures in image registration / S.V. Voronov, A.G. Tashlinskii // *Pattern recognition and image analysis*. – 2016. – Vol. 26(3). – P. 502-505. DOI: 10.1134/S1054661816030226.

Depth maps correction based on neighboring frames

A. Varlamova
Lomonosov Moscow State University
Moscow, Russia
avarlamova1996@gmail.com

V. Kitov
Lomonosov Moscow State University
Moscow, Russia
v.v.kitov@yandex.ru

Abstract—The use of convolutional neural networks in image processing tasks often allows achieving significantly better results in comparison with traditional methods. For the problem of calculating a real-time depth map, the state-of-the-art method is the MiDaS method - a convolutional neural network trained on a large dataset that is not publicly available. However, this approach does not include the ability to use information from neighboring frames which can improve the prediction. The method proposed in this paper uses the depth maps generated by MiDaS for several frames of a video sequence and their further refinement, which makes it possible to achieve an improvement in quality without a significant decrease in the algorithm performance.

Keywords— *depth maps, convolutional neural networks, supervised learning, structure-from-motion.*

1. INTRODUCTION

Depth information obtained from one input image or a set of them is an important part of solving a wide range of problems. For example, an algorithm that allows you to get depth maps in real-time would make it possible to replace LIDAR systems in self-driving vehicles, which could lead to a reduction in their cost. Depth maps can be useful in segmentation and object detection problems [1, 2]. They even can be applied to the style transfer problem, significantly improving the quality of the generated image [3, 4].

The best results are achieved by the neural networks-based approach. The current state-of-the-art model – MiDaS [5] is capable of generating accurate depth maps. However, in the case of a video sequence, one can extract additional spatial information from neighboring frames, which can be used for further refinement.

Papers focusing on depth prediction using video sequence [6, 7] typically use it to organize learning in an unsupervised manner with the neural network taking one frame at a time as its input. The paper [8] considers an approach based on two frames, however, the method loses [5] both in terms of performance and quality.

This article focuses on the possibility of using information from consequent frames of a video sequence simultaneously.

Various input data configurations are explored and a link between the quality of prediction and frames distance from each other in a sequence is studied. This paper also proposes a modified loss function capable of improving the prediction. The resulting method gives an increase in quality while maintaining comparable performance.

2. RELATED WORK

Depth maps calculation is traditionally associated with structure-from-motion algorithms, however, there are some serious disadvantages such as significant processing time, incomplete depth maps, and inconsistent results in the case of moving objects. Another group of methods is based on training on video data in an unsupervised manner, however,

they use several frames only to calculate the loss function, leaving the prediction stage with one image [6, 7].

3. METHOD

The main goal of this paper is to improve the inverse depth maps obtained using the state-of-the-art MiDaS method for the case of a video sequence leveraging information about neighboring frames.

Simple convolution was trained for several data representation options. To ensure that the values are non-negative, the RELU activation function was used. Two versions of the loss function were used. The first one was identical to MiDaS's:

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^N \mathcal{L}_{ssi}(c^n, (c^t)^n) + \alpha \mathcal{L}_{reg}(c^n, (c^t)^n), \quad (1)$$

where

$$\mathcal{L}_{reg}(c, c^t) = \frac{1}{M} \sum_{k=1}^K \sum_{i=1}^M (|\nabla_x R_i^k| + |\nabla_y R_i^k|),$$

and

$$\mathcal{L}_{ssi}(c, c^t) = \frac{1}{2M} \sum_{i=1}^M (c - c^t)^2,$$

where N is the size of the training set, M – the size of the image, $R_i^k = c_i - c_i^t$ at k scale. c_i, c_i^t are predicted and ground true inverse depth maps correspondingly which were transformed to have equal scales and translations.

However, during the experiments, it was noticed that the structure of the loss function is such that a larger penalty is given to closer objects. To balance the difference in values, an additional version of the loss function is proposed:

$$\mathcal{L}_{proximity}(c, c^t) = \frac{1}{2M} \sum_{i=1}^M \left(\frac{c_i - c_i^t}{c_i^t} \right)^2 \quad (2)$$

Thus objects farther away (with smaller values) are given a larger penalty for deviating from real values.

The median transformation was used as an approach to align predicted and original scale and translation parameters with each other. For each true inverse depth map and each prediction shift values were calculated as the median value and scale values as the mean absolute differences between the inverse depth map values and the computed shift values.

Three types of input data were considered:

- MiDaS's inverse depth maps for C_i, C_{i+k}, C_{i-k} (ND_k , neighbor depth).
- Inverse depth map for C_i , averaged inverse depth maps for k steps forward and backward (AD_k , average depth).

- Inverse depth map for C_i , differences between depth map for C_i and C_{i-k} , C_{i+k} (DD_k , depth difference).

A. Training details

The training was performed on 25000 images from a synthetic dataset [9] during 1 epoch. α parameter was set to 0.3, batch size was equal to 18 for ND and DD input types and 9 for AD.

B. Evaluation

Quality was estimated on the validation set (5500 images) for different values of k parameter: 1, 2, 4, and 8 for ND and DD, and 2, 3, 4 for AD.

Abs Rel, Sq Rel, RMSE, log RMSE, δ metrics were used for evaluation. The best result was achieved by the method with ND input type with k=4.

For further experiments, a simple U-Net-like model with skip-connection was trained on the same dataset for data input type. The architecture of the model can be seen in Fig. 1.

For the comparison, Sintel [10] and ETH3D [11] datasets were used. Results are presented in Table 1. Better performance using all metrics for the Sintel dataset and almost all (except for δ) for the ETH3D dataset was achieved.

4. CONCLUSION

The best-performing methods for real-time depth map calculation are based on using neural networks. The State-of-the-art MiDaS method can be improved upon by using information from neighboring frames in tasks where video sequence is involved.

TABLE I. METHOD COMPARISON

Method	Metrics				
	Abs Rel	Sq Rel	RMSE	log RMSE	δ
ETH3D, ND_4	0.108	0.282	1.386	0.148	0.118
ETH3D, MiDaS	0.112	0.431	1.61	0.152	0.117
Sintel, ND_4	0.276	2.546	5.111	0.438	0.38
Sintel, MiDaS	0.295	3.254	5.401	0.468	0.386

In this paper, several options for presenting input data were proposed and it was shown how their form affects the final quality. The link between the quality of the prediction and the distance between adjacent frames during training is shown. Additionally, a modified loss function was proposed, which made it possible to make the model more consistent for distant objects. The algorithm obtained as a result of combining the approaches made it possible to improve the quality of prediction on datasets that did not participate in

training and validation. At the same time, the possibility of using it in real-time was retained.

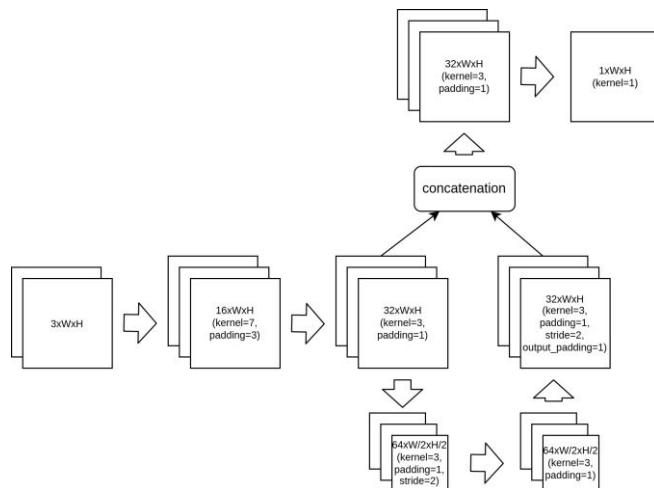


Fig. 1. Improved model architecture

REFERENCES

- [1] Couprie, C. Indoor semantic segmentation using depth information / C. Couprie, C. Farabet, L. Najman, Y. LeCun // ArXiv preprint: 1301.3572, 2013.
- [2] Qi, C.R. Frustum pointnets for 3d object detection from rgb-d data / C.R. Qi, W. Liu, C. Wu, H. Su, L.J. Guibas // Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. – P. 918-927.
- [3] Liu X.-C. Depth-aware neural style transfer / X.-C. Liu, M.-M. Cheng, Y.-K. Lai, P.L. Rosin // Proceedings of the Symposium on Non-Photorealistic Animation and Rendering. – 2017. – P. 1-10.
- [4] Liao, R. Depth-preserving style transfer / R. Loao, Y. Xia, X. Zhang // Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. – 2016.
- [5] Lasinger, K. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer / K. Lasinger, R. Ranftl, K. Schindler, V. Koltun // ArXiv preprint: 1907.01341, 2019.
- [6] Godard, C. Digging into self-supervised monocular depth estimation / C. Godard, O. Mac Aodha, M. Firman, G.J. Brostow // Proceedings of the IEEE international conference on computer vision. – 2019. – P. 3828-3838.
- [7] Luo, X. Consistent video depth estimation / X. Luo, J.-B. Huang, R. Szeliski, K. Matzen, J. Kopf // ArXiv preprint: 2004.15021, 2020.
- [8] Wang, C. Web Stereo Video Supervision for Depth Prediction from Dynamic Scenes / C. Wang, S. Lucey, F. Perazzi, O. Wang // CoRR. – 2019.
- [9] Krahenbuhl, P. Free supervision from video games / P. Krahenbuhl // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – P. 2955-2964.
- [10] Butler, D.J. A naturalistic open source movie for optical flow evaluation / D.J. Butler, J. Wulff, G.B. Stanley, M.J. Black // European Conf. on Computer Vision (ECCV). – 2012. – P. 611-625.
- [11] Schops T. A Multi-View Stereo Benchmark with High-Resolution Images and Multi-Camera Videos / T. Schops, J.L. Schonberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, A. Geiger // Conference on Computer Vision and Pattern Recognition (CVPR). – 2017.

Метод адаптивного управления дорожными сигналами на основе максимального взвешенного потока

А.А. Агафонов
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
ant.agafonov@gmail.com

А.С. Юмаганов
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
yumagan@gmail.com

В.В. Мясников
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
vmyas@geosamara.ru

Аннотация—В работе предлагается метод адаптивного управления дорожными сигналами (фазами светофоров). Предлагаемый метод включает два этапа, за каждый из которых отвечает отдельный алгоритм. На первом шаге оценивается прогнозируемый «поток» транспортных средств через перекресток по заданной полосе на разрешающий сигнал светофора. На втором шаге формируется оценка «взвешенного» потока, которая учитывает время ожидания транспортных средств на перекрестке. В итоге выбор фазы светофора определяется из критерия максимизации взвешенного транспортного потока через перекресток.

Ключевые слова— управление сигналами светофора, обучаемый метод, адаптивное управление, интеллектуальные транспортные системы, поток транспортных средств.

1. ВВЕДЕНИЕ

Идеи цифровой экономики накладывают отпечаток на все аспекты современной жизни. Существенное влияние они оказывают на транспорт. В частности, де-факто стандартом в настоящее время являются транспортные средства (ТС) с электронными помощниками, которые контролируют полосу движения, выполнение дорожных знаков и правил, предупреждают столкновения и т.п. Указанные нововведения являются лишь одной из составных частей создаваемых интеллектуальных транспортных систем (ИТС), о которых говорится в утверждённой Указом Президента Российской Федерации от 1 декабря 2016 г. № 642 Стратегии научно-технологического развития Российской Федерации, в частности, направления "Связанность территории Российской Федерации за счёт создания интеллектуальных транспортных и телекоммуникационных систем, а также занятия и удержания лидерских позиций в создании международных транспортно-логистических систем, освоении и использовании космического и воздушного пространства, Мирового океана, Арктики и Антарктики". Не меньшее, а скорее, большее значение при построении ИТС играют решения, позволяющие обеспечивать комплексное управление транспортным движением на территории целого населенного пункта. Хотя проектирование «с нуля» городской/областной транспортной инфраструктуры с ИТС позволяет сделать это максимально эффективно (начав с проектирования мест проживания и работы, обеспечив тем самым управление масштабами транспортного спроса и предложения на территориях, выполнив проектирование

транспортных потоков, и т.п.), дороговизна, а порой и принципиальная невозможность изменения существующей транспортной топологии и жилой инфраструктуры делает значительно более важным решение прагматичной задачи – оптимизации движения ТС в рамках существующей инфраструктуры. И здесь на первый план выступает задача *адаптивного управления сигналами/фазами светофоров*, поскольку ее решение допускает быстрое и относительно недорогое внедрение в существующую транспортную инфраструктуру при значительном (как правило в разы) росте ее эффективности (пропускной способности, затрат горючего и т.п.).

2. ПОСТАНОВКА ЗАДАЧИ

Под *перекрестком* понимается пересечение проезжих частей. Каждая из проезжих частей делится *разметкой* на полосы движения, часть из которых отвечает за въезд на перекресток, а другая – за выезд. При этом разметка определяет *допустимые направления движения*, а сигналы светофора(ов) определяют возможность перемещения по соответствующей полосе в текущий момент времени. Пример перекрестка приведен на рис. 1.

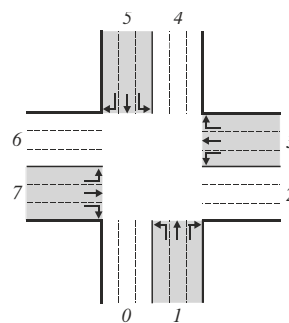


Рис. 1. Пример перекрестка (по материалам работы [1])

Для этого примера пара "16" означает движение ТС с левой полосы проезжей части, помеченной цифрой "1", на проезжую часть, помеченную цифрой "6". Далее, *фаза светофора* – это множество непротиворечивых сигналов светофора [1]. Фаза может выбираться в рамках некоторой предопределённой последовательности фаз, т.н. *светофорного цикла*, а может выбираться произвольно. При этом, если выбор фазы светофорного регулирования происходит на основании данных движения транспортных средств, то говорят о *системах адаптивного управления дорожными сигналами* [1]. В свою очередь, системы/алгоритмы адаптивного управления можно разделить на *эмпирические детерминированные* (где

правила изменения фаз от движения прописаны заранее) и системы/алгоритмы на основе машинного обучения. В последнем случае правило, связывающее обстановку на перекрестке с конкретной светофорной фазой, настраивается в результате работы системы. Авторское решение, представленное в работе [1], относится к эмпирическим детерминированным алгоритмам. Другими примерами являются SCOOT и SCATS [2], метод MaxPressure [3] и др. К решениям на базе методов обучения относятся методы на основе нечёткой логики, интеллекта роя, генетические алгоритмы, RL-методы и др. [4,5]. В данной работе предложена модификация эмпирического метода, описанного в [1], для учета времени ожидания транспортных средств на перекрестке.

3. МЕТОД МАКСИМАЛЬНОГО ПОТОКА

Детальное описание базового метода представлено в работе [1]. Суть подхода – в качестве следующей фазы светофора выбирается та, при которой оценка (прогноз) потока транспортных средств через перекресток будет максимальной, то есть P – множество допустимых фаз):

$$\operatorname{argmax}(\{PredFlow(\text{phase}) \text{ for phase in } P\}), \quad (1)$$

где

$$PredFlow(\text{phase}) = \sum_{l \in L_{\text{phase}}^{income}} \sum_{c \in C_l} I(t(c) < \tau) \quad (2)$$

здесь $I(val)$ – индикатор, принимающий значение «1» при $val=True$, и значение «0» – иначе; τ – длительность временного интервала в секундах между фазами светофора. Ключевым моментом метода является оценка (прогноз) времени $t(c)$ достижения ТС c перекрестка. Оценка производится на основании явно заданных моделей движения, известных из классического курса физики/кинематики, описана детально в [1]. В качестве входных данных функции $t(c)$ выступают всего два параметра: текущая скорость ТС $v=v(c)$ и расстояние от этого ТС до перекрестка $S=S(c)$.

4. ПРЕДЛАГАЕМЫЙ МЕТОД

Базовый метод не учитывает времени ожидания (время простоя), которое транспортные средства затрачивают на перекрестке. Как следствие, может возникнуть ситуация, при которой время ожидания транспортных средств на второстепенной дороге (с маленьким потоком) будет расти неограниченно. Для исправления этого недостатка предлагается рассчитывать «взвешенный» транспортный поток, учитывающий время ожидания следующим образом:

$$PWFlow(\text{phase}) = \sum_{l \in L_{\text{phase}}^{income}} \sum_{c \in C_l} I(t(c) < \tau) * (1 + \alpha * \text{delay}(c, l)) \quad (3)$$

где $\text{delay}(c, l)$ – время простоя ТС c на полосе l .

Способ выбора следующей фазы светофора аналогичен базовому методу [1]: выбирается та фаза, для которой взвешенный транспортный поток оказывается максимальным:

$$\operatorname{argmax}(\{PWFlow(\text{phase}) \text{ for phase in } P\}). \quad (4)$$

5. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Экспериментальное исследование направлено на сравнение эффективности предложенного метода с

классическим алгоритмом MaxPressure [3] и современными RL-методами [4,5]. Эксперименты проводились в среде моделирования SUMO с использованием сценариев для управления светофорами [6]. Для оценки эффективности использовались синтетические (4x4) и реальные сценарии движения на отдельном перекрестке, магистрали и в городской области. Результаты сравнения по критерию среднего времени движения, представленные в Таблице I, подтвердили высокую эффективность предложенного метода.

Таблица III. КРИТЕРИЙ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ ДОРОЖНЫМ ДВИЖЕНИЕМ С ИСПОЛЬЗОВАНИЕМ СИГНАЛОВ СВЕТОФОРОВ

Модель	4x4	Перекресток	Магистраль	Область
IDQN	145,76	43,85	58,94	85,40
IPPO	156,37	48,33	62,01	87,64
Max Pressure	161,02	47,35	60,27	98,64
Max PFlow	141,82	44,08	56,74	85,30

6. ЗАКЛЮЧЕНИЕ

В работе предложен детерминированный метод адаптивного управления дорожными сигналами, основанный на максимизации взвешенного транспортного потока. Показана высокая эффективность разработанного метода по сравнению с базовыми эмпирическими методами и методами на основе машинного обучения с подкреплением на синтетических и реальных сценариях движения.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Российского научного фонда (проект № 21-11-00321, <https://rscf.ru/en/project/21-11-00321/>).

ЛИТЕРАТУРА

- [1] Мясников, В.В. Детерминированная прогнозная модель управления сигналами светофоров в интеллектуальных транспортных и геоинформационных системах / В.В. Мясников, А.А. Агафонов, А.С. Юмаганов // Компьютерная оптика. – 2021. – Т. 45, № 6. – С. 917-925. DOI: 10.18287/2412-6179-CO-1031.
- [2] Lowrie, P.R. SCATS—a traffic responsive method of controlling urban traffic / P.R. Lowrie. – Roads and Traffic Authority NSW, 1990. – 28 p.
- [3] Varaiya, P. The Max-Pressure Controller for Arbitrary Networks of Signalized Intersections / P. Varaiya // Advances in Dynamic Network Modeling in Complex Transportation Systems: Complex Networks and Dynamic Systems. – New York, NY: Springer, 2013. – P. 27-66.
- [4] Gregurić, M. Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data / M. Gregurić, M. Vujić, C. Alexopoulos, M. Miletić // Applied Sciences. – 2020. – Vol. 10(11). DOI: 10.3390/app10114011.
- [5] Qadri, S.S.S.M. State-of-art re-view of traffic signal control methods: challenges and opportunities / S.S.S.M. Qadri, M.A. Gökçe, E. Öner // European Transport Re-search Review. – 2020. – Vol. 12(55). DOI: 10.1186/s12544-020-00439-1.
- [6] RESCO Web Site [Electronic resource]. – Access mode: <https://github.com/Pi-Star-Lab/RESCO> (15.02.2022).

Восстановление пространственно неоднородных изображений на основе дважды стохастических фильтров

К.К. Васильев
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
vkk@ulstu.com

В.Е. Дементьев
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
dve@ulntc.ru

А.В. Белянчиков
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
friedlemon73@gmail.com

Аннотация—В работе рассматриваются вопросы восстановления изображений, когда в распоряжении имеется только часть пораженных аддитивным шумом наблюдений, регулярно размещенных на исходном изображении. Для решения данной задачи предлагается использовать модификации нелинейных фильтров, синтезированных на основе глубоких гауссовых моделей. Полученные результаты позволяют сделать вывод об эффективности работы фильтров в сравнении с линейными методами и нейронными сетями.

Ключевые слова— восстановление изображений, нелинейная фильтрация, глубокие модели.

1. ВВЕДЕНИЕ

Среди множества задач, связанных с обработкой изображений, особую роль играют фильтрация и восстановление изображений [1-5]. Классическим подходом к решению этих задач является использование различных линейных моделей, позволяющих сформировать прогноз в зашумленную или отсутствующую область изображения. Наличие такого прогноза в рамках каузальных [1-3] и некаузальных [2] фильтров позволяет с высокой степенью точности обрабатывать однородный по пространству и стационарный по времени материал. Однако изображения реальных объектов отличаются характерной пространственной неоднородностью, что в случае использования линейных фильтров приводит к значительным ошибкам обработки. Одним из выходов здесь может быть применение нелинейных математических моделей изображений адекватных реальным сигналам и соответствующих этим моделям фильтров. В настоящей работе рассматривается возможность применения для обработки пространственно неоднородных изображений дважды стохастических фильтров, способных приспосабливаться к локальным корреляционным характеристикам изображений и обеспечивающим хорошее качество обработки пространственно неоднородного материала [4].

При этом в фокусе работы находится задача восстановления исходного изображения в условиях, когда в распоряжении имеется только часть пораженных аддитивным шумом наблюдений, регулярно размещенных на исходном изображении. Такая задача возникает не только при восстановлении утраченных фрагментов изображений, но и при улучшении качества исходного изображения за счет увеличения его

разрешения (задача «суперразрешения») [5] или при оценивании OFDM сигнала по регулярным пилот-сигналам [6].

2. ОПИСАНИЕ АЛГОРИТМА ВОССТАНОВЛЕНИЯ.

Для решения поставленной задачи будем считать восстанавливаемое изображение реализацией случайного поля (СП), заданного на прямоугольной сетке размером N_1 на N_2 . С учетом возможной пространственной неоднородности изображения, будем считать, что моделью случайного поля может быть глубокая гауссова модель (ГГМ), построенная, для определенности, на базе марковских конструкций [1,2,7]. Примером таких конструкций может быть двумерная дважды стохастическая модель:

$$x_{ij} = F_{\text{ARMR}(2,2)}(P_1 + \rho_{1ij}, P_2 + \rho_{2ij}, B\xi_{ij}),$$

где A, P_1, P_2, B – некоторые числа, характеризующие средние математическое ожидание, корреляционные свойства и дисперсию изображения; $F_{\text{ARMR}(K_1, K_2)}(\rho_1, \rho_2, b)$ – преобразование, соответствующее двумерной авторегрессионной модели с кратными корнями характеристических уравнений (АРКК модели) [7-9]; $K_1 = 2, K_2 = 2$ – коэффициенты, определяющие кратность модели; β – нормирующий коэффициент; $\rho_{1ij} = F_{\text{ARMR}(2,2)}(\gamma_{11}, \gamma_{12}, \gamma_1 \xi_{1ij})$, $\rho_{2ij} = F_{\text{ARMR}(2,2)}(\gamma_{21}, \gamma_{22}, \gamma_2 \xi_{2ij})$ – случайные величины, определяемые собственными АРКК моделями; $\xi_{ij}, \xi_{1ij}, \xi_{2ij}$ – гауссовские белые СП; $i = 1, \dots, M_1, j = 1, \dots, M_2$.

Тогда сформулируем задачу следующим образом. Пусть производятся наблюдения смеси $z_{ij} = x_{ij} + n_{ij}$ информационного СП и СП $\{n_{ij}\}$ с ковариационной функцией V_n . Решим задачу восстановления СП $\{x_{ij}\}$ по наблюдениям $\{z_{ij}\}$ предполагая, что нам доступны только отдельные такие наблюдения, связанные определенной регулярностью.

Используя особенности ГГМ и методы рекуррентной нелинейной фильтрации [4], возможно записать следующее семейство восстанавливающих фильтров в скользящих окнах:

$$\hat{X}_{i,j} = \hat{X}_{\Delta i,j} + P_{Dij} V_n^{-1} (z_{i+1,j+1} - \hat{X}_{\Delta i,j}), \quad (1)$$

где $\hat{X}_{\Delta i,j}$ – тензорный прогноз во фрагмент изображения с центром в точке (i, j) в соответствии с выбранной

моделью; $\hat{x}_{\Delta ij}$ – последний элемент $\hat{X}_{\Delta ij}$; $P_{Di,j} = P_{\Delta ij}(E + V_n^{-1}P_{\Delta ij})^{-1}$ – тензор ошибок восстановления; $P_{\Delta ij}$ – тензор ошибок экстраполяции [4]. Выражение (1) предполагает переоценку всех элементов вектора $\hat{X}_{i,j}$, за исключением последнего, на каждой итерации по результатам обработки одного единственного наблюдения $z_{i+1,j+1}$. В случае отсутствия наблюдения $z_{i+1,j+1}$ действий не производится и осуществляется переход к следующему шагу.

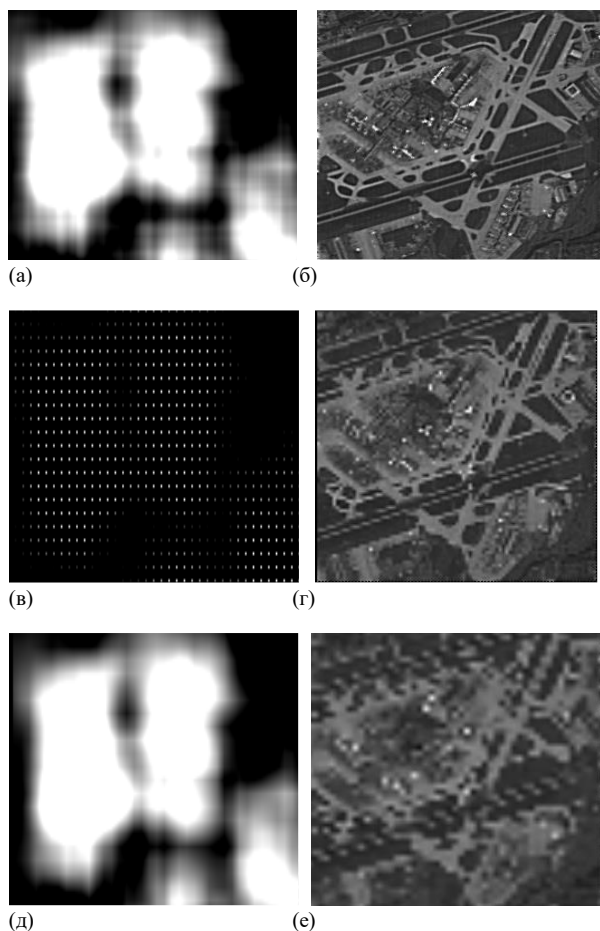


Рис. 1. Исходные изображения и результаты обработки

3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ АЛГОРИТМА

В рамках представленного подхода была выполнена обработка семейства искусственных изображений, полученных с помощью ГГМ, и реальных спутниковых изображений. В ходе подготовки эти изображения смешивались с аддитивным шумом, а значения яркости в их элементах обнулялись за исключением пикселей, расположенных друг от друга на заданном расстоянии l по строкам и столбцам соответственно. На рис. 1 представлены исходное искусственное изображение (1а), результат его подготовки (1б), результат восстановления (1в), фрагмент спутникового изображения (1г), результат его восстановления предлагаемым фильтром (1д) и, для сравнения, обычным билинейным интерполятором.

В таблице 1 приведены некоторые результаты сравнения работы предлагаемого фильтра с простой нейронной сетью, построенной на основе карт Кохонена [10] и обученной на 10 тыс. искусственных изображений,

и билинейным интерполятором в зависимости от отношения сигнал/шум $q = \sigma_x^2 / \sigma_n^2$ и шага l .

Таблица 1. ОШИБКИ ВОССТАНОВЛЕНИЯ ИЗОБРАЖЕНИЙ

Вида алгоритма	Оценка дисперсии ошибки восстановления $\sigma^2 = \sum (x_{\text{исх}}(i,j) - x_{\text{восст}}(i,j))^2 / N_1 N_2$		
	$q = 5, l = 3$	$q = 2, l = 3$	$q = 5, l = 6$
Нейронная сеть	0,56	0,93	0,83
Билинейный интерполятор	0,43	1,48	0,96
Нелинейный фильтр (1)	0,42	0,67	0,56
	$q = 2, l = 6$	$q = 5, l = 10$	$q = 2, l = 10$
Нейронная сеть	1,44	1,89	2,37
Билинейный интерполятор	1,97	1,92	3,98
Нелинейный фильтр (1)	0,81	0,93	1,13

Прямой анализ показывает, что предложенный фильтр предпочтительнее своих аналогов, при этом выигрыш растет по мере увеличения отношения сигнал/шум и шага l .

4. ЗАКЛЮЧЕНИЕ

Таким образом, в работе рассмотрены вопросы восстановления изображений на основе полукausalных нелинейных дважды стохастических фильтров. Полученные результаты позволяют сделать вывод об эффективности работы фильтров в сравнении с линейными методами и нейронными сетями.

БЛАГОДАРНОСТИ

Работа поддержана грантом РФФИ №19-29-09048.

ЛИТЕРАТУРА

- [1] Васильев, К.К. Статистический анализ изображений / К.К. Васильев, В.Р. Крашенинников. – Ульяновск: УлГТУ, 2014. – 214 с.
- [2] Bouman, C.A. Model Based Imaging Processing / C.A. Bouman. – Purdue University, 2013. – 414 p.
- [3] Васильев, К.К. Оптимальная обработка сигналов в дискретном времени / К.К. Васильев. – М.: Радиотехника, 2016. – 288 с.
- [4] Васильев, К.К. Дважды стохастическая фильтрация пространственно неоднородных изображений / К.К. Васильев, В.Е. Дементьев // Радиотехника и электроника. – 2020. – Т. 65, № 4. – С. 1-8.
- [5] Christopher, T. Deep learning based super resolution, without using a GAN / T. Christopher [Электронный ресурс]. – Режим доступа: [https://towardsdatascience.com\(01.02.2022\)](https://towardsdatascience.com(01.02.2022)).
- [6] Васильев, К.К. Прием дискретных сообщений в многочастотных каналах связи / К.К. Васильев, А.В. Белянчиков, В.Е. Дементьев // Автоматизация процессов управления. – 2020. – Т. 60, № 2. – С. 109-116.
- [7] Dementev, V.E. Doubly stochastic models of images / V.E. Dementev, K.K. Vasiliev, N.A. Andriyanov // Pattern Recognition and Image Analysis. – 2015. – Vol. 25, № 1. – P. 105-110.
- [8] Дементьев, В.Е. Представление и обработка спутниковых многозональных изображений / К.К. Васильев, В.Е. Дементьев. – Ульяновск, 2017. – 247 с.
- [9] Дементьев, В.Е. О связи между каузальным и некаузальным прогнозом при оценивании многомерных случайных полей / В.Е. Дементьев // Вестник нижегородского государственного инженерно-экономического университета. – 2018. – № 10. – С. 5-12.
- [10] Ларионов, И.Б. Карты Кохонена как способ восстановления мультимедийной информации / И.Б. Ларионов // Журнал радиоэлектроники. – 2010. – № 10. – 8 с.

Использование алгоритмов обработки изображения для решения задачи навигации беспилотных аппаратов

К.В. Святков
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
k.svyatov@gmail.com

В.Е. Дементьев
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
dve@ulntc.ru

М.А. Гапонова
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
barannikova-mari@mail.ru

Аннотация—Работа посвящена разработке и исследованию алгоритмов навигации беспилотных транспортных средств. Основой таких алгоритмов является комплексирование результатов обработки видеоизображений, получаемых с камер высокого разрешения, а также пространственного дальномера. В работе обосновывается состоятельность решения и приводятся характеристики его эффективности.

Ключевые слова— автономное управление, беспилотные аппараты, семантическая сегментация, нейронные сети, фильтрация данных.

1. ВВЕДЕНИЕ

В настоящее время беспилотные системы являются одним из наиболее перспективных трендов развития целого спектра научных и технических отраслей, связанных с искусственным интеллектом [1-4]. Важным требованием, возникающим при проектировании беспилотного транспорта, является возможность действовать в условиях быстро изменяющейся обстановки. Центральной частью современной системы управления беспилотным транспортным средством является навигационная система, предоставляющая информацию о местоположении объекта в произвольный момент времени.

В литературе описано большое количество различных подходов, применяемых при позиционировании [2-4 и др.]. Большая часть этих подходов связана с применением глобальных спутниковых навигационных систем (СНС) GPS/ГЛОНАСС/Beidou [3,4]. На открытом пространстве и при отсутствии помех такие системы позволяют определять собственное местоположение с точностью до нескольких метров. В то же время опыт практического применения СНС, например, в городской застройке показывает существенно большие ошибки позиционирования, связанные с погодными условиями, наличием препятствий, особенностями работы приемного комплекса и т.п. Немаловажным недостатком для позиционирования высокоподвижных беспилотных транспортных средств является относительно низкая периодичность получения координат с помощью СНС (до 1 секунды). Определенным выходом здесь является использование дополнительной информации с бортовых инерциальных навигационных систем (ИНС) в рамках комплексных систем позиционирования. Однако такое комплексирование не дает дополнительной информации об окружающей беспилотный автомобиль обстановке.

В настоящей работе для определения собственных координат беспилотного автомобиля и окружающих его объектов предлагается применять нелинейную фильтрацию данных, получаемых с синхронизированных бортовых камер высокого разрешения и пространственных дальномеров.

2. ОПИСАНИЕ АЛГОРИТМА ПОЗИЦИОНИРОВАНИЯ

Чтобы определить собственные координаты беспилотного автомобиля воспользуемся следующим алгоритмом. Вначале на изображениях, получаемых с бортовых камер высокого разрешения, выполним распознавание объектов, которые могут быть использованы как потенциальные ориентиры. Такими объектами в частности являются здания, отдельно стоящие деревья, характерные детали рельефа местности и т. д. Опыт нашего коллектива показывает, что приемлемые с точки зрения производительности и качества распознавания результаты обеспечивают современные искусственные нейронные сети. В частности в настоящей работе использовалась модификация сети Yolo5 [5], ключевой особенностью которой является возможность использования для каждой отдельной операции распознавания не только отдельного изображения с бортовой камеры, но и массива данных пространственного дальномера. При этом использование последовательных сверток позволяет относительно просто преодолеть проблему существенного различия в разрешении данных бортовой камеры и пространственного дальномера. Описанная модификация искусственной нейронной сети была обучена типовым образом на реальных парах изображений и данных пространственного дальномера, полученных с борта беспилотного автомобиля.

Вторым этапом обработки является уменьшение пространственных размеров распознанного материала в соответствии с разрешением пространственного дальномера и определение геометрических центров выделенных объектов. Наличие информации о таких центрах вкупе с данными данных пространственного дальномера позволяют определить текущие расстояния до соответствующих объектов. Тогда с помощью простых геометрических соотношений, основанных на использовании информации об особенностях регистрирующих устройств и размещении центров объектов на кадре исходного изображения, возможно

определять координаты опорных точек относительно текущего положения беспилотного автомобиля.

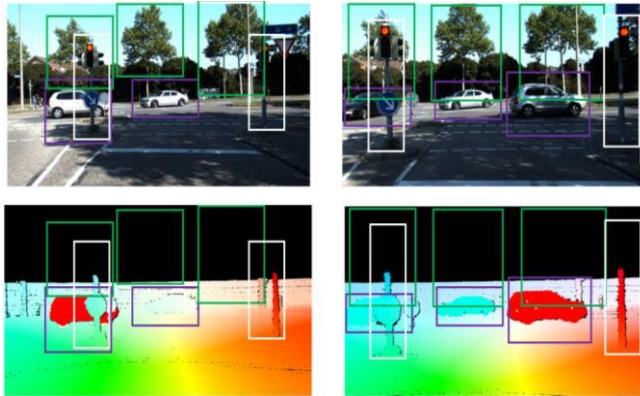


Рис. 1. Результаты совместной обработки двух временных кадров, каждый из которых представлен изображением с бортовой камеры и данными с пространственного дальномера

На третьем этапе, предполагая возможность перемещения автомобиля с переменным ускорением, формируют математическое описание движения этого автомобиля с помощью дважды стохастической модели [7]. Эта модель связывает компоненты оцениваемого навигационного вектора

$\bar{x}_i = (x_i \ y_i \ v_{xi} \ v_{yi} \ a_{xi} \ a_{yi} \ x_{oi}^1 \ y_{oi}^1 \ \dots \ x_{oi}^{M_i} \ y_{oi}^{M_i})^T$,
следующим образом:

$$\begin{aligned} x_i &= 2x_{i-1} - x_{i-2} + a_{xi}(x_{i-1} - x_{i-2}), \\ y_i &= 2y_{i-1} - y_{i-2} + a_{yi}(y_{i-1} - y_{i-2}) \end{aligned}$$

$$v_{xi} = v_{x(i-1)} + a_{x(i-1)}, \quad v_{yi} = v_{y(i-1)} + a_{y(i-1)},$$

$$a_{xi} = r_{ax}a_{x(i-1)} + \xi_{xi}, \quad a_{yi} = r_{ay}a_{y(i-1)} + \xi_{yi},$$

$$x_{oi}^k = x_{o(i-1)}^k, \quad y_{oi}^k = y_{o(i-1)}^k, \quad k = 1, \dots, M_i, \quad i = 2, \dots, n_2,$$

где $(x_{oi}^j, y_{oi}^j), j = 1, \dots, M_i$ - координаты j -го ориентира, наблюдаемого в i -ый момент времени; ξ_{xi}, ξ_{yi} - независимые гауссовские случайные величины с ненулевыми дисперсиями $\sigma_{\xi_{xi}}^2$ и $\sigma_{\xi_{yi}}^2$. Эти случайные величины дают возможность учесть внешние случайные воздействия на движущееся транспортное средство. Важным отличием представленных уравнений от известных вариантов [6, 8] является возможность описания движения объектов, перемещающихся в том числе с переменным ускорением.

Уравнения наблюдений компонент скорости и относительных координат объектов могут быть записаны в следующем виде [6,8]:

$$\begin{aligned} z_{vxi} &= v_{xi} + n_{vxi}, \quad z_{vyi} = v_{yi} + n_{vyi}, \\ z_{oxi}^k &= x_i - x_{oi}^k + n_{oxi}^k, \quad z_{oyi}^k = y_i - y_{oi}^k + n_{oyi}^k, \\ k &= 1, \dots, M_i, \quad i = 2, \dots, n_2, \end{aligned}$$

где $n_{oxi}, n_{vxi}, n_{oyi}, n_{vyi}, n_{oxi}^k, n_{oyi}^k, k = 1, \dots, M_i$ - гауссовские погрешности измерений. Представленные модели позволяют записать следующий вариант векторного дважды стохастического фильтра [7]:

$$\bar{\hat{x}}_i = \bar{\hat{x}}_{\xi i} + P_i C_i^T V_i (\bar{z}_i - C_i \bar{\hat{x}}_{\xi i}),$$

где $P_i = P_{\xi i} (E + C_i^T V_i C_k P_{\xi k})^{-1}$, $\bar{\hat{x}}_{\xi i} = \varphi_{i-1} \bar{\hat{x}}_{\xi i-1}$, $P_{\xi i} =$

$$\varphi_{i-1} P_{(i-1)} \varphi_{i-1}^T + V_{\xi i}; \quad V_{\xi i} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{\xi i}^2 \end{pmatrix}; \quad C_i =$$

$$\begin{pmatrix} 0 & 1 & 0 \end{pmatrix}.$$

Получаемая оценка $\bar{\hat{x}}_i$ навигационного вектора может быть уточнена с помощью обнаружения и привязки к значимым объектам, описание которых имеется в памяти навигационного устройства. Важно, что каждый шаг фильтрации предполагает расчет матрицы P_i , содержащей покомпонатные дисперсии ошибок позиционирования.

В рамках представленного подхода была выполнена обработка тестовых трех фрагментов данных, полученных с борта беспилотного автомобиля, с оценкой его координат в разные моменты времени. Установлено, что предлагаемый метод действительно позволяет выполнять восстановление траектории перемещения беспилотного автомобиля. При этом скорость накопления ошибки позиционирования составила порядка 1.5 м. за минуту при условии непрерывного наблюдения двух неподвижных объектов. Уменьшить и стабилизировать данную ошибку возможно путем комплексирования с результатами иных навигационных системы, например, спутниковых и инерциальных.

3. ЗАКЛЮЧЕНИЕ

Таким образом в работе представлен вариант навигации автономного транспортного средства по результатам совместной обработки данных, полученных с камеры высокого разрешения и пространственного дальномера. Особенности предлагаемого подхода предполагают простые процедуры его комплексирования с другими методами навигации.

БЛАГОДАРНОСТИ

Работа поддержана Фондом содействия развитию малых предприятий в НТС в рамках договора 8ГУКодИИС12-D7/72685 от 27.12.2021 (Программа КодИИ).

ЛИТЕРАТУРА

- [1] Prabhakar, N. Object Tracking Using Frame Differencing and Template Matching / N. Prabhakar, V. Vaithyanathan, A. Sharma, A.Singh, P. Singhal // Research Journal of Applied Sciences, Engineering and Technology. – Maxwell Scientific Organization, 2012.
- [2] Iqbal, A. Non-Linear Moving Target Tracking: A Particle Filter Approach / A. Iqbal, S. Waqar Shah, Sh. Khan // International Journal of Computer and Communication System Engineering (IJCCSE). – 2014. – Vol. 1(1).
- [3] Пат. 5398292 США, МКИ G 6K 9/46. Edge detecting apparatus. – № 049524; Заяв. 20.04.93; Опубл. 14.03.95; НПК 382/199. Англ.
- [4] Пат. 5627905 США, МКИ G 6K 9/00. Optical flow detection system. – № 353589; Заяв. 12.12.94; Опубл. 6.05.97; НПК 382/107. Англ.
- [5] Maturana, D. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition / D. Maturana // IEEE_RSJ International Conference on Intelligent Robots and Systems (IROS), 2015. – P. 922-928.
- [6] Dikshit, S.S. A Recursive Kalman Window Approach to Image Restoration / S.S. Dikshit // IEEE Trans. – 1984. – P. 125-139.
- [7] Dementev, V. Doubly stochastic models of images / V. Dementev, K. Vasiljev, N. Andriyanov // Pattern Recognition and Image Analysis. – 2015. – Vol. 25(1). – P. 105-110.
- [8] Васильев, К.К. Динамическое использование ориентиров для оценки координат автономных необитаемых подводных аппаратов / К.К. Васильев, А.В. Бобков // Информационно-измерительные и управляющие системы. – 2017. – Т. 15. – № 12. – С. 11-14.

Совместное применение нейронных сетей и псевдоградиентной привязки изображений при оценке дефектов мостовых конструкций

М.Н. Суевин
ФНПЦ АО «Научно-производственное объединение «Марс»
Ульяновск, Россия
source81@yandex.ru

Р.Г. Магдеев
Ульяновский государственный
технический университет
Ульяновск, Россия
radiktkd2@yandex.ru

А.Г. Ташлинский
Ульяновский государственный
технический университет
Ульяновск, Россия
tag@ulstu.ru

Аннотация—Работа посвящена исследованию, направленному на оценку динамики роста дефектов, выявленных на видеоизображениях металлических мостовых конструкций. Для оценки динамики дефектов использовано совместное применение нейронной сети и адаптивной псевдоградиентной привязки изображений. Приведены примеры полученных результатов.

Ключевые слова— дефект, псевдоградиент, совмещение, привязка, обработка изображений, нейросеть.

1. ВВЕДЕНИЕ

Мосты, переправы, эстакады имеют большое стратегическое и экономическое значение. При этом при воздействии железнодорожных составов или крупного автотранспорта на эти сооружения они подвергаются существенным динамическим нагрузкам, которые приводят к появлению тех или иных повреждений и дефектов. Развитие таких дефектов может привести к разрушению самой конструкции, причём с высокой вероятностью такие разрушения могут произойти в моменты прохождения железнодорожных составов или автотранспорта, что потенциально может привести к жертвам среди персонала и пассажиров и значительным финансовым потерям. Поэтому одной из важных проблем, стоящих в настоящее время перед организациями, обслуживающими такие конструкции, является мониторинг состояния целостности их основных элементов [1, 2].

Мониторинг предполагает, в том числе, проведение периодического осмотра элементов конструкции. При этом проведение такого осмотра требует привлечения высококвалифицированного персонала, что влечет за собой значительные временные и финансовые затраты. Выходом в этой ситуации является использование перспективных систем неинвазивного мониторинга, предполагающих минимальное участие человека. В настоящей работе рассматривается часть подобной системы [3], предполагающая возможность автоматизированной оценки динамики роста дефекта конструкции на основе сопоставления фотоизображений данного дефекта в разные периоды времени.

2. МЕТОДИКА И СОДЕРЖАНИЕ ИССЛЕДОВАНИЯ

Базовым компонентом системы неинвазивного мониторинга является регистрирующий комплекс, расположенный на борту компактного беспилотного

летательного аппарата (БЛА) и снабженный бортовой видеосъемкой высокого разрешения, размещенной на гироподвесе. БЛА опоясан легкой, прочной защитной решеткой, что позволяет ему перемещаться в непосредственной близости от элементов конструкции, не опасаясь столкновений с ними. Бортовой навигационный комплекс, состоящий из спутниковой и инерциальной навигационных систем, позволяет определять координаты аппарата и выполнять его многократное перемещение по заданной траектории. В результате формируются одновременные видеокдры одних и тех же элементов конструкции для текущей и постобработки [4].

При первичном обследовании диагностируемой конструкции оператор БЛА задает траекторию её облёта с учетом видеорегистрации всех потенциальных зон интереса. Затем полученный видеоматериал анализируется с помощью нейронной сети U-Net [5, 6], например, как это описано в [4, 6]. Выделяются места с дефектами, которые при постобработке фиксируются оператором на некоторой условной карте объекта, а также в базе данных запоминаются несколько изображений видеоряда с областью дефекта.

При повторном обследовании той же конструкции, проводимом с заданным диагностическим интервалом, требуется обнаружить не только новые дефекты, но и оценить динамику изменения обнаруженных ранее. В базу данных также запоминаются ряд изображений месторасположения вновь обнаруженных и «старых» дефектов. По собранному новым видеоданным и данным предыдущих исследований оценивается динамика изменения дефектов.

Для решения задачи оценивания изменения геометрических параметров дефектов применена привязка (совмещение) одновременных изображений этих дефектов. Необходимость этой операции обусловлена тем, что координаты съемки и направления оптической оси камеры при получении одновременных изображений отличаются, что приводит к различным геометрическим деформациям этих изображений. Кроме того, съёмка может вестись также разными камерами. Для привязки изображений в работе используются адаптивные безыдентификационные псевдоградиентные процедуры привязки [7], которые имеют высокую скорость сходимости оценок параметров взаимных пространственных деформаций, устойчивы в условиях априорной неопределённости и обеспечивают субпиксельную точность привязки изображений при приемлемых вычислительных затратах. Отметим, что в

процедурах этого класса могут быть заданы различные модели взаимных геометрических трансформаций, включая проективную [8]. Анализ показал, что в решаемой задаче целесообразно использовать релейные псевдоградиентные процедуры [7, 9] и достаточно применения аффинной модели деформаций. Эта модель требует меньше вычислительных ресурсов как для прямого, так и для обратного преобразований.

Однако недостатком псевдоградиентных процедур является небольшой диапазон оцениваемых параметров [10], что накладывает некоторые ограничения на рассогласование привязываемых изображений. Для нивелирования этого недостатка как предобработка изображений применялась гауссова фильтрация, эквализация яркостей и компенсация неравномерности освещённости. Критерием выполнения процедуры привязки служило превышение некоторого порога по коэффициенту корреляции привязываемых изображений.

3. ПРИМЕР ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Примеры результатов работы нейронной сети и процедуры привязки разновременных (интервал 6 месяцев) видеопоследовательностей приведены на рис. 1. Видно (рис. 1а,г – изображения из видеопоследовательности первого пролета БЛА, рис. 1б,д - второго), что нейронная сеть выделяет прямоугольниками заданные объекты интереса: типовые элементы конструкции (зеленый цвет), маркеры (желтый) и собственно трещины (красный). Но, несмотря на успешное выделение области расположения трещин, сделать корректный вывод о динамике их изменения сложно, поскольку в силу дестабилизирующих факторов (ветер, погрешность навигации и пр.) БЛА летит по заданной траектории вдоль мостового перехода с существенной пространственной погрешностью. Это приводит к тому, что изображения имеют разный ракурс и масштаб. Пространственное рассогласование устраняется процедурой псевдоградиентной привязки изображений, результат которой для первой пары изображений приведен на рис. 1в, для второй – рис. 1е.

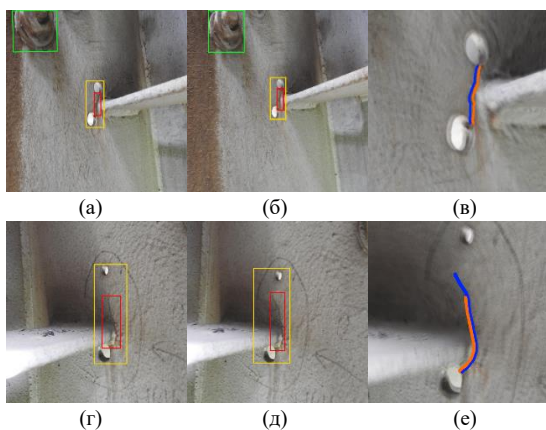


Рис. 1. Примеры дефектов, выявленных нейронной сетью на изображениях, и результат привязки разновременных снимков

Анализ приведенных в единую систему координат и пространственно-совмещенных изображений показал, что дефект (трещина в мостовой опоре), за время между съемками на изображениях рис. 1а и рис. 1б, практически не изменился. На рис. 1в в полупрозрачном режиме показаны совмещенные изображения в увеличенном

масштабе, при этом скелет трещины выделен цветом. Трещина же на изображении рис. 1д по сравнению с изображением рис. 1г увеличилась по размерам более, чем на 20%, что может служить сигналом соответствующим службам, для принятия мер по локализации или устранения данного дефекта.

4. ЗАКЛЮЧЕНИЕ

Таким образом, совместное использование нейронных сетей и псевдоградиентной привязки изображений для оценки динамики дефектов мостовых конструкций показало высокую эффективность и может служить основой автоматизированной системы диагностики и отслеживания динамики развития дефектов мостовых и других конструкций по разновременным видеопоследовательностям, полученным с промышленных БЛА.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ и Правительства Ульяновской области в рамках научных проектов № 19-29-09048 и № 19-47-730004.

ЛИТЕРАТУРА

- [1] Правила обследования несущих строительных конструкций зданий и сооружений: СП 13-102 – 2003. Введ. 2014-01-01. – М.: Государственный комитет Российской Федерации по строительству и жилищно-коммунальному комплексу (ГОССТРОЙ России), 2003. – 27 с.
- [2] ГОСТ 31937-2011. Здания и сооружения. Правила обследования и мониторинга технического состояния. – М.: Стандартинформ, 2014. – 55 с.
- [3] Dementiev, V.E. Detecting anomalies in temporal image sequences based on object identification by the stochastic gradient adaptation / V.E. Dementiev, R.G. Magdeev, A.G. Tashlinskii // IEEE Xplore. International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-5. DOI: 10.1109/ITNT52450.2021.9649175.
- [4] Dementiev, V. Improving The Quality of Detecting and Recognizing Defects in Images of Steel Structures / V. Dementiev, M. Suetin, M. Gaponova // IEEE Xplore. International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-5. DOI: 10.1109/ITNT52450.2021.9649166.
- [5] Nikitin, F. U-Net Based Architectures for Document Text Detection and Binarization / F. Nikitin, V. Dokholyan, I. Zharikov, V. Strijov // Lecture Notes in Computer Science. – 2019. – Vol. 11845 LNCS. – P. 79-88. DOI: 10.1007/978-3-030-33723-0_7.
- [6] Брагин, А.В. Распознавание дислокационной структуры эпитаксиальных слоев карбида кремния с использованием нейронной сети / А.В. Брагин, Д.В. Пьянзин, Р.И. Сидоров, Д.А. Скворцов // Компьютерная оптика. – 2020. – Т. 44, № 4. – С. 653-659. DOI: 10.18287/2412-6179-CO-660.
- [7] Tashlinskii, A.G. Pseudogradient Estimation of Digital Images Interframe Geometrical Deformations / A.G. Tashlinskii // Vision Systems: Segmentation & Pattern Recognition. – Vienna, Austria: "I Tech Education and Publishing", 2007. – P. 465-494. DOI: 10.5772/4975.
- [8] Ташлинский, А.Г. Оценивание параметров пространственных деформаций последовательностей / А.Г. Ташлинский. – Ульяновск: Издательство УлГТУ, 2000. – 139 с.
- [9] Magdeev, R.G. Efficiency of object identification for binary images / R.G. Magdeev, A.G. Tashlinskii // Computer Optics. – 2019. – Vol. 43(2). – P. 277-281. DOI: 10.18287/2412-6179-2019-43-2-277-281.
- [10] Magdeev, R. Efficiency of Stochastic Gradient Identification of Similar Shape Objects in Binary and Grayscale Images / R. Magdeev, A. Tashlinskii, G. Safina // CEUR Workshop Proceedings. – 2020. – Vol. 2665. – P. 25-28.

Исследование стойкости компактных дескрипторов векторной карты к трансформациям её содержимого

А.С. Асанов
Самарский национальный
исследовательский университет им.
акад. С.П. Королева
Самара, Россия
alanlox1882@gmail.com

Ю.Д. Выборнова
Самарский национальный
исследовательский университет им.
акад. С.П. Королева
Самара, Россия
vybornovamail@gmail.com

В.А. Федосеев
Самарский национальный
исследовательский университет им.
акад. С.П. Королева
Самара, Россия
vicanfed@gmail.com

Аннотация—Работа посвящена исследованию стойкости компактных дескрипторов векторной карты к трансформациям её содержимого. Данное исследование имеет целью выявление подходящих дескрипторов в сценарии zero watermarking при защите содержимого векторной карты от подделки или несанкционированного использования.

Ключевые слова— нулевое встраивание, защита векторной карты, ГИС, ЦВЗ.

1. ВВЕДЕНИЕ

На сегодняшний день одним из активно используемых подходов к защите векторных данных является так называемое «нулевое встраивание цифрового водяного знака (ЦВЗ)» (zero watermarking) [1, 2, 3]. Данная процедура представляет собой аналог процедуры хеширования: для защищаемой карты вычисляется некоторая идентифицирующая информация (дескриптор), которая впоследствии хранится отдельно. На этапе «извлечения» дескриптор вычисляется заново, после чего сформированный ЦВЗ сравнивается с исходным. Например, в [4] вычисляется отношение расстояний между вершинами (FVDR), после чего для большей секретности объединяется с битами ЦВЗ. В [5] ЦВЗ строится на основе триангуляции и вычисления локальных характеристик внутри каждого треугольника.

В зависимости от решаемой задачи ЦВЗ должен быть либо стойким ко всем трансформациям векторной карты (если целью является защита авторских прав), либо он должен разрушаться при определенных преобразованиях и таким образом сигнализировать о подделке данных. Первостепенной задачей настоящего исследования было определить, какие характеристики векторной модели полезно использовать при формировании ЦВЗ в зависимости от видов подделки карт, которые необходимо с помощью этого ЦВЗ детектировать. Например, ЦВЗ, построенный на основе отношения расстояний между характерными вершинами полигонов, позволит детектировать только добавление/удаление нехарактерных вершин объектов, а ЦВЗ, построенный на основе среднего расстояния от центра ограничивающего прямоугольника до каждой вершины, позволит детектировать ещё и добавление/удаление характерных вершин.

2. АЛГОРИТМ РАМЕРА-ДУГЛАСА-ПЕКЕРА И ЕГО МОДИФИКАЦИЯ

В рассматриваемых в рамках настоящей работы дескрипторах в качестве составной части используется алгоритм Рамера-Дугласа-Пекера [6, 7], предназначенный для сокращения избыточности описания векторного полигонального объекта путём уменьшения числа точек. Использование этого алгоритма в дескрипторах теоретически должно обеспечивать стойкость к наибольшему изменению объектов карты. Ниже рассмотрим более детально данный алгоритм, а также его модификацию, разработанную нами для устранения недостатков, проявляющихся при использовании данного алгоритма в составе дескрипторов.

Входом алгоритма Рамера-Дугласа-Пекера служат наиболее удалённые вершины в объекте (точки A и B на рис. 1). Если таких вершин несколько, то выбираются те, которые были найдены первыми. Далее алгоритм находит вершину, наиболее удалённую от отрезка, соединяющего вершины, выбранные на первом шаге (точка C на рис. 1). Далее рассчитывается отношение расстояния от точки до этого отрезка к длине самого отрезка (отношение CD к AB на рис. 1). Если полученное значение меньше некоторого порога: $CD/AB < \alpha$, то все ранее не отмеченные вершины считаются *нехарактерными* и могут быть выброшены из конечного набора вершин оптимизированного объекта. Если же $CD/AB \geq \alpha$, то алгоритм рекурсивно вызывает себя для двух новых отрезков – CA и CB . Результатом работы алгоритма будет объект, состоящий только из *характерных* вершин.

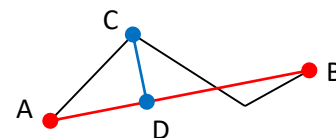


Рис. 1. Иллюстрация алгоритма Рамера-Дугласа-Пекера

Недостатком данного алгоритма применительно к рассматриваемой задаче построения информативных дескрипторов является сложность подбора значения α таким образом, чтобы в результате работы алгоритма оставалось достаточное количество характерных точек для сохранения формы объекта. По этой причине в алгоритм внесено косметическое изменение: при первой итерации точка C признаётся характерной всегда, а на дальнейших итерациях вместо постоянного для всех объектов порога α используется порог $\alpha \cdot CD/AB$, где AB

и CD в данном случае – отрезки очередного объекта, чьё отношение рассчитывается на первой итерации алгоритма. Подобная модификация позволяет повысить устойчивость множества характерных вершин к незначительным модификациям карты, что подтвердили результаты проведённых экспериментов.

3. ОПИСАНИЕ АНАЛИЗИРУЕМЫХ КОМПАКТНЫХ ДЕСКРИПТОРОВ

Чтобы получить возможность локализации изменений в цифровой карте, исходная карта делится на равные по размеру квадратные участки. На каждом из них вычисляется компактный дескриптор участка карты, учитывающий характеристики всех полигональных объектов, находящихся на данном участке. Перед расчётом дескриптора каждый объект подвергается оптимизации алгоритмом Рамера-Дугласа-Пекера.

В рамках работы исследовались следующие дескрипторы:

- 1) среднее отношение расстояний между характерными вершинами (FVDR) [4];
- 2) среднее отношений площадей ограничивающих прямоугольников;
- 3) среднее отношений расстояний между центрами масс объектов;
- 4) средняя доля характерных вершин внутри группы;
- 5) среднее отношение расстояний от центра до правого верхнего угла ограничивающего прямоугольника;
- 6) среднее отношение расстояний от центра ограничивающего прямоугольника до каждой вершины.

4. ПОРЯДОК ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Как отмечалось выше, ключевой задачей исследования являлся анализ стойкости выбранных дескрипторов к различным искажениям векторной карты. Для её решения был выполнен ряд вычислительных экспериментов. В ходе них карта разбивалась на равные фрагменты, после чего для каждого фрагмента вычислялись вышеуказанные дескрипторы. Далее выполнялась процедура искажения карты. При этом изменялись как тип искажения, так и его уровень, определяемый скалярным параметром. Далее по искажённой карте также рассчитывались дескрипторы, и оценивалось относительное изменение дескриптора. Дескриптор признавался устойчивым к определённому искажению, если относительная ошибка оказывалась ниже 1% при всех значениях параметра. Ниже перечислены применявшиеся искажения и в скобках их параметры:

- 1) Поворот карты (угол от 0 до 360°).
- 2) Добавление вершин без изменения формы объекта (процент изменения числа точек от 10 до 100%).
- 3) Добавление нехарактерных вершин, меняющих форму объекта (процент изменения числа точек от 10 до 100%).
- 4) Удаление произвольных вершин (доля удалённых вершин от 5 до 40%).

- 5) Удаление вершин, не являющихся характерными (доля удалённых вершин от 5 до 40%).
- 6) Изменение порядка вершин – циклический сдвиг (величина сдвига).
- 7) Добавление копий существующих объектов карты (процент изменения числа объектов от 10 до 100%).
- 8) Добавление новых четырёхточечных объектов карты (процент изменения числа объектов от 10 до 100%).
- 9) Удаление произвольных объектов карты (доля удалённых объектов от 10 до 90%).

5. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Объём тезисов не позволяет подробно описать полученные результаты, поэтому отметим лишь общие выводы.

Наиболее устойчивым среди исследуемых дескрипторов оказался дескриптор 2, что означает, что он является самым эффективным для защиты авторских прав. Остальные дескрипторы являются стойкими лишь к части искажений, поэтому они могут быть использованы для детектирования искажений тех видов, к которым они неустойчивы. Одним из способов обнаружения конкретного вида искажения является комбинирование нескольких видов дескрипторов, различающихся всего в одном искажении. Например, можно использовать дескрипторы 1 и 4 для обнаружения удаления небольшого количества объектов, поскольку это единственное искажение, по которому данные дескрипторы дают различный результат.

Также следует отметить, что при добавлении вершин, не изменяющих форму объекта, и удалении нехарактерных вершин, все дескрипторы оказались устойчивыми только благодаря использованию алгоритма Рамера-Дугласа-Пекера. В ином случае устойчивым к этим видам искажений являлся бы только дескриптор 2.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке РФФИ (проект 19-29-09045)

ЛИТЕРАТУРА

- [1] Ren, N. Copyright Protection Based on Zero Watermarking and Blockchain for Vector Maps / N. Ren, Y. Zhao, C. Zhu, Q. Zhou, D. Xu // ISPRS International Journal of Geo-Information. – 2021. – Vol. 10(5). – P. 294. DOI: 10.3390/ijgi10050294.
- [2] Zhou, Q. Zero Watermarking Algorithm for Vector Geographic Data Based on the Number of Neighboring Features / Q. Zhou, C. Zhu, N. Ren, W. Chen, W. Gong // Symmetry. – 2021. – Vol. 13(2). – P. 208. DOI: 10.3390/sym13020208.
- [3] Xi, X. Dual Zero-Watermarking Scheme for Two-Dimensional Vector Map Based on Delaunay Triangle Mesh and Singular Value Decomposition / X. Xi, X. Zhang, W. Liang, Q. Xin, P. Zhang // Applied Sciences. – 2019. – Vol. 9(4). – P. 642. DOI: 10.3390/app9040642.
- [4] Peng, Y. A Zero-Watermarking Scheme for Vector Map Based on Feature Vertex Distance Ratio / Y. Peng, M. Yue // Journal of Electrical and Computer Engineering. – 2015. – Vol. 2015. – P. 1-6.
- [5] Li, A. Study on copyright authentication of GIS vector data based on Zero-watermarking / A. Li, B.-X. Lin, Y.A. Chen // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. – 2008. – Vol. 37(B4). – P. 1783-1786.

- [6] Ramer, U. An iterative procedure for the polygonal approximation of plane curves / U. Ramer // *Computer Graphics and Image Processing*. – 1972. – Vol. 1(3). – P. 244-256. DOI: 10.1016/S0146-664X(72)80017-0.
- [7] Douglas, D.H. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature / D.H. Douglas, T.K. Peucker // *Cartographica: The International Journal for Geographic Information and Geovisualization*. – 1973. – Vol. 10(2). – P. 112-122. DOI: 10.3138/FM57-6770-U75U-7727.

Сравнение алгоритмов управления сигналами светофоров в крупномасштабном сценарии моделирования движения транспортных средств

А.А. Агафонов
Самарский национальный
исследовательский университет им. акад.
С.П. Королева
Самара, Россия
ant.agafonov@gmail.com

Е.Ю. Ефименко
Самарский национальный
исследовательский университет им. акад.
С.П. Королева
Самара, Россия
evgeniaefimenko27@gmail.com

Аннотация—Задача управления сигналами светофоров остается важной задачей, решаемой в интеллектуальных транспортных системах. Развитие методов машинного обучения привело к активному развитию адаптивных методов управления сигналами светофоров, основанных на обучении с подкреплением. В то же время, сравнение классических подходов к управлению с методами на основе обучения с подкреплением в существующих работах проводится либо на синтетических сценариях, либо на сценариях, включающих малое количество перекрестков. В данной работе мы проводим сравнение алгоритмов управления сигналами светофоров в крупномасштабном сценарии моделирования движения транспортных средств в системе моделирования SUMO. Для проведения исследований был разработан сценарий моделирования на основе сценария SUMO «TAPAS Cologne».

Ключевые слова— обучение с подкреплением, детерминированная модель, управление сигналами светофоров, SUMO.

1. ВВЕДЕНИЕ

Задача оптимального управления транспортными потоками является одной из ключевых задач в современных транспортных системах. Управление транспортными потоками позволит сократить транспортные расходы на совершение поездок в сети, снизить потребление топлива и уменьшить уровень загрязнения окружающей среды.

Развитие информационно-коммуникационных технологий, интернета вещей (IoT), подключенных и автономных транспортных средств привело к увеличению объема данных, которые могут использоваться для решения задачи адаптивного управления сигналами светофоров. В частности, активно развиваются методы управления, основанные на алгоритмах обучения с подкреплением.

Адаптивное управление транспортными потоками путем управления сигналами светофоров основывается на информации, полученной от детекторов транспортных средств, видеокamer, подключенных транспортных средств и др. для выбора следующей фазы светофора. Обзор и сравнение классических методов управления (SOTL, MaxPressure) на основе детерминированных моделей было проведено в [1, 2]. В работе [3] авторы предложили адаптивный метод управления на основе детерминированной прогнозной модели. Предлагаемый метод основывается на выборе фазы светофорного цикла, прогнозируемый поток для которой оказывается максимален. В последние два десятилетия наибольший интерес представляет разработка методов адаптивного

управления на основе методов машинного обучения, и, в частности, на основе обучения с подкреплением [4-6]. Однако эти методы могут работать нестабильно в сложных сценариях с зашумленными данными и слишком чувствительны к гиперпараметрам [7]. Более того, исследования разработанных алгоритмов часто проводятся на синтетических сценариях моделирования с перекрестками типовой структуры, либо рассматривают малое число перекрестков в реальных сценариях.

В данной работе было проведено сравнение алгоритмов адаптивного управления сигналами светофоров на разработанном крупномасштабном сценарии моделирования движения транспортных средств. В следующем разделе приводится краткое описание сравниваемых алгоритмов. В разделе 3 представлено описание сценария, постановка и результаты экспериментов. Заключение работы представлено в разделе 4.

2. АЛГОРИТМЫ АДАПТИВНОГО УПРАВЛЕНИЯ СИГНАЛАМИ СВЕТОФОРОВ

А. Детерминированные методы

Детерминированные методы используют заранее определенные правила для выбора следующей фазы светофорного цикла.

- SOTL [1] оценивает зеленые «запросы» от текущей фазы и других конкурирующих фаз, а затем решает, сохранить или изменить текущую фазу. В частности, сохранение текущей фазы светофора может определяться наличием непрерывного потока транспортных средств.
- MaxPressure [2] стремится сбалансировать длину очереди между соседними перекрестками за счет минимизации «давления» фаз на перекресток.
- MaxFlow [3] основывается на выборе фазы светофорного цикла, прогнозируемый поток для которой оказывается максимален.

Б. Методы на основе обучения с подкреплением

Обозначим через S множество состояний некоторого объекта. Пусть A обозначает множество возможных действий. Упрощенно задача состоит в построении отображения $S \rightarrow A$, которое для каждого конкретного состояния отображает действие, которое необходимо совершить в указанном состоянии в определенном смысле оптимальным образом. Для решения этой задачи обычно используют подход на основе марковского

процесса принятия решений $\langle S, A, P, R, \gamma \rangle$, где P определяет вероятности перехода между состояниями, величина R характеризует «награду», γ – коэффициент дисконтирования. Задача обучения с подкреплением заключается в нахождении оптимальной политики $\pi: S \rightarrow A$, которая максимизирует награду R .

В качестве алгоритмов на основе обучения с подкреплением в работе рассматриваются:

- IDQN [8] – алгоритм на основе Q-обучения; каждый агент контролирует отдельный светофор и обучается независимо.
- IPPO [8] – алгоритм на основе оптимизации политики.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Экспериментальные исследования разработанных алгоритмов были проведены в системе моделирования движения транспортных средств SUMO [9]. Для проведения исследований был разработан сценарий моделирования, основанный на сценарии движения SUMO «TAPAS Cologne» [10].

Сценарий содержит 2928 перекрестков различной конфигурации, в т.ч. 316 сигнализированных перекрестков, 5808 сегментов. Движение транспортных средств рассматривалось в утренний час пик в период с 6 до 8 утра. Дорожная сеть рассматриваемого сценария представлена на Рис. 1.



Рис. 1. Дорожная сеть сценария моделирования движения

Сравнение алгоритмов проводилось по двум метрикам: среднее время ожидания и среднее время движения. Среднее времени ожидания (в секундах) показывает среднее время, которое транспортные средства провели без движения на перекрестке, среднее время движения (в секундах) — это среднее время, затрачиваемое транспортными средствами на совершение поездки в сети. Результаты сравнения алгоритмов представлены в Таблице I.

4. ЗАКЛЮЧЕНИЕ

В работе проведено сравнение адаптивных детерминированных алгоритмов и алгоритмов на основе

обучения с подкреплением для решения задачи управления сигналами светофоров. Для проведения экспериментальных исследований был разработан крупномасштабный сценарий моделирования движения транспортных средств. Результаты показывают, что алгоритм IDQN на основе обучения с подкреплением, обученный независимо для каждого агента, превосходит остальные алгоритмы. Однако данный подход не является масштабируемым, что требует разработки новых подходов к управлению.

Таблица II. СРАВНЕНИЕ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ ТРАНСПОРТНЫМИ ПОТОКАМИ

	Среднее время движения	Среднее время ожидания
IDQN	319,21	19,54
IPPO	371,32	67,28
SOTL	361,14	44,04
MaxPressure	334,8	28,59
MaxFlow	327,69	23,68

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Российского научного фонда (проект № 21-11-00321, <https://rscf.ru/en/project/21-11-00321/>).

ЛИТЕРАТУРА

- [1] Papageorgiou, M. Review of road traffic control strategies / M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, Y. Wang // Proceedings of the IEEE. – 2003. – Vol. 91(12). – P. 2043-2065.
- [2] Varaiya, P. The Max-Pressure Controller for Arbitrary Networks of Signalized Intersections / P. Varaiya // Advances in Dynamic Network Modeling in Complex Transportation Systems: Complex Networks and Dynamic Systems. – 2013. – P. 27-66.
- [3] Мясников, В.В. Детерминированная прогнозная модель управления сигналами светофоров в интеллектуальных транспортных и геoinформационных системах / В.В. Мясников, А.А. Агафонов, А.С. Юмаганов // Компьютерная оптика. – 2021. – Т. 45, № 6. – P. 917-925. DOI: 10.18287/2412-6179-CO-1031.
- [4] Yau, K.-L.A. A survey on Reinforcement learning models and algorithms for traffic signal control / K.-L.A. Yau, J. Qadir, H.L. Khoo, M.H. Ling, P. Komisarczuk // ACM Computing Surveys. – 2017. – Vol. 50(3).
- [5] Greguric, M. Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data / M. Greguric, M. Vujic, C. Alexopoulos, M. Miletic // Applied Sciences. – 2020. – Vol. 10(11). – P. 4011.
- [6] Qadri, S.S.S.M. State-of-art review of traffic signal control methods: challenges and opportunities / S.S.S.M. Qadri, M.A. Gökçe, E. Öner // European Transport Research Review. – 2020. – Vol. 12(1). – P. 55.
- [7] Genders, W. An Open-Source Framework for Adaptive Traffic Signal Control / W. Genders, S. Razavi // ArXiv preprint: 1909.00395, 2019.
- [8] Ault, J. Learning an Interpretable Traffic Signal Control Policy / J. Ault, J.P. Hanna, G. Sharon // ArXiv preprint: 1912.11023, 2020.
- [9] Lopez, P.A. Microscopic Traffic Simulation using SUMO / P.A. Lopez // 21st International Conference on Intelligent Transportation Systems (ITSC). – 2018. – P. 2575-2582.
- [10] TAPASCologne – SUMO Documentation [Electronic resource]. – Access mode: <https://sumo.dlr.de/docs/Data/Scenarios/TAPASCologne.html> (21.11.2021).

Сравнение методов реконструкции промежуточных кадров видеопоследовательностей с динамической сценой

Ю.Х. Ганеева

Самарский национальный исследовательский университет им. акад. С.П. Королева

Самара, Россия

jganeeva99@gmail.com

Аннотация—Задача реконструкции промежуточных кадров видеопоследовательностей направлена на увеличение частоты кадров за счет синтеза новых кадров на основе информации о соседних кадрах видеопоследовательности. Более высокая частота кадров позволяет улучшить качество визуального восприятия. Также методы реконструкции промежуточных кадров получили распространение в задаче 3D-реконструкции объектов и сцен. В данной работе представлен результат сравнения качества работы современных (англ. state of the art) методов реконструкции промежуточных кадров на видеопоследовательностях с динамической сценой.

Ключевые слова—реконструкция кадров, динамическая сцена, XVFI, RRIN, CDFI, RIFE, AdaCof.

1. ВВЕДЕНИЕ

Реконструкция кадров видеопоследовательности — это процесс синтеза новых кадров видеопоследовательности на основе информации о соседних кадрах. Результат процедуры реконструкции кадров позволяет повышать частоту кадров видеопоследовательности, что в результате приводит к улучшению визуального восприятия. Важную роль процедура реконструкции кадров может играть при решении задачи 3D-реконструкции объектов, когда исходная видеопоследовательность, описывающая внешний вид объекта или сцены, имеет низкую частоту съемки. Процедура реконструкции кадров позволяет получить изображение объекта с ракурса, которого не было в исходной видеопоследовательности, что в результате приводит к более точной и реалистичной реконструкции объекта.

Целью данного исследования является сравнение качества работы state of the art методов реконструкции кадров видеопоследовательностей XVFI [1], RRIN [2], CDFI [3], RIFE [4], AdaCof [5] на наборе данных, состоящем из видеопоследовательностей с динамической сценой.

Структура работы следующая. Раздел 2 посвящен краткому описанию рассматриваемых методов. В разделе 3 описывается используемый набор данных, используемые метрики качества, порядок проведения экспериментальных исследований и результаты. Работа заканчивается заключением, в котором делаются выводы на основе результатов, полученных в ходе экспериментальных исследований.

2. ОПИСАНИЕ РАССМАТРИВАЕМЫХ ПОДХОДОВ РЕКОНСТРУКЦИИ КАДРОВ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ

XVFI [1] – это метод реконструкции кадров, основанный на использовании архитектуры нейронной сети XVFI-Net, предложенной авторами работы. XVFI-Net имеет рекурсивную многомасштабную структуру и состоит из двух каскадных модулей BIOF-I и BIOF-T, которые обучаются предсказывать двунаправленные оптические потоки между двумя входными кадрами и от истинного кадра к входными кадрам, соответственно. Результаты экспериментальных исследований, представленных в работе, показали, что предложенный авторами метод позволяет успешно фиксировать информацию об объектах, имеющих высокую динамику между соседними кадрами.

RRIN [2] – это нейросетевой подход реконструкции промежуточных кадров видеопоследовательностей, в котором используется остаточное уточнение и адаптивные веса для синтеза промежуточных кадров. Метод остаточного уточнения используется при работе с оптическим потоком и при генерации новых кадров для повышения точности и улучшения визуального вида. Адаптивная карта весов же объединяет кадры с прямой и обратной деформацией для уменьшения артефактов. Все подмодули в методе реализованы с архитектурой U-Net меньшей глубины (в отличие от оригинальной архитектуры), что обеспечивает эффективность работы метода.

CDFI [3] – это нейросетевой подход реконструкции промежуточных кадров. Авторы предлагают управляемую сжатием архитектуру сети, которая использует упрощение модели за счет оптимизации, вызывающей разреженность для значительного уменьшения размера модели при достижении хорошей производительности. Сначала авторы упрощают модель AdaCof и показывают, что 10-кратно сжатая модель AdaCof работает так же, как и оригинальная реализация, затем авторы дополнительно улучшают модель, вводя модуль деформации с несколькими разрешениями, который повышает визуальную согласованность на многослойных объектах.

RIFE [4] – это нейросетевой метод реконструкции промежуточных кадров видеопоследовательностей. Учитывая пару последовательно идущих RGB-кадров I_0 и I_1 и целевой шаг t ($0 \leq t \leq 1$), цель метода – синтезировать промежуточный кадр I_t . Авторы оценивают

промежуточные оптические потоки $F_{(t-0)}$, $F_{(t-1)}$ и карту слияния M путем подачи входных кадров и t как дополнительный канал в предложенную для оценки промежуточного оптического потока нейронную сеть IFNet. В результате используют сверточный энкодер RefineNet [6] для детализации высокочастотной области I_t^+ и снижения артефактов в результатах модели «ученика». Вычислительная стоимость аналогична использованию IFNet. RefineNet выдает остаточный результат реконструкции промежуточного кадра Δ ($-1 \leq \Delta \leq 1$). В результате получается реконструированное изображение $I_{t+\Delta}$.

AdaCof [5] – это нейросетевой метод реконструкции промежуточных кадров видеопоследовательностей. Авторы предлагают использовать полностью сверточную нейронную сеть, которая оценивает веса $W_{k,l}$, векторы смещения $(\alpha_{k,l}, \beta_{k,l})$ и карту окклюзии V . Первая часть нейронной сети представляет собой архитектуру U-Net [7], которая состоит из кодера и декодера, между слоями которых выполняется операция конкатенации карт признаков. За архитектурой U-Net следует 7 подсетей, которые окончательно оценивают выходные данные $W_{k,l}$, $\alpha_{k,l}$, $\beta_{k,l}$ для каждого кадра, а также карту окклюзии V .

3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Для проведения экспериментальных исследований был собран набор данных, состоящий из видеопоследовательностей с динамической сценой. Характер динамики в видеопоследовательностях следующий: несколько действий, одно медленное действие, одно быстрое действие. Видеопоследовательности были сняты на камеры GoPro Hero 9 и имеют частоту 60 FPS и 240 FPS. Для создания проверочного набора, по которому производилась оценка качества реконструкции кадров, из видеопоследовательностей с частотой 60 FPS посредством удаления кадров были сформированы видеопоследовательности с частотой 30 FPS; из видеопоследовательностей с частотой 240 FPS аналогичным образом были сформированы видеопоследовательности с частотой 60 FPS и 120 FPS.

На вход методов, указанных во введении, подавались пары кадров сформированных видеопоследовательностей. Результатом работы методов являлись восстановленные на местах пропусков кадры. Качество работы методов оценивалось с помощью метрик PSNR и SSIM.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (1)$$

где $MSE = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$.

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (2)$$

где μ_x, μ_y – математическое ожидание x, y , σ_x^2, σ_y^2 – дисперсия x, y , $\sigma_x\sigma_y$ – ковариация x и y , $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$, L – динамический диапазон пикселей, $k_1 = 0,01$, $k_2 = 0,03$.

Следует отметить, что в методах [3] и [5] отсутствует возможность восстановления нескольких подряд идущих кадров (более двух).

В таблице 1 представлены лучшие результаты для каждого типа действий.

Таблица III. Лучшие результаты для каждого типа действия

Тип действия	Входное количество кадров в секунду	Выходное количество кадров в секунду	Метод	Качество работы метода	
				PSNR	SSIM
Одно медленное действие	120	240	CDFI [3]	38.0691	0.9709
Одно быстрое действие	120	240	AdaCof [5]	37.0799	0.9684
Несколько действий	120	240	AdaCof [5]	37.9498	0.9712

4. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования было выполнено сравнение качества работы state-of-the-art методов реконструкции кадров видеопоследовательностей, а именно XVFI [1], RRIN [2], CDFI [3], RIFE [4], AdaCof [5], на собранном наборе данных, состоящем из видеопоследовательностей с динамической сценой. Полученные результаты показывают довольно хорошее качество реконструкции промежуточных кадров, однако с появлением сильной динамики визуальное и количественное качество методов снижается. В связи с этим в следующих работах планируется разработать метод, устойчивый к сильной динамике.

ЛИТЕРАТУРА

- [1] Sim, H. XVFI: eXtreme Video Frame Interpolation / H. Sim, J. Oh, M. Kim // arXiv: 2103.16206, 2021.
- [2] Li, H. Video Frame Interpolation Via Residue Refinement / H. Li, Y. Yuan, Q. Wang // ICASSP – IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2020. – P. 2613-2617. DOI: 10.1109/ICASSP40776.2020.9053987.
- [3] Ding, T. CDFI: Compression-Driven Network Design for Frame Interpolation / T. Ding, L. Liang, Z. Zhu, I. Zharkov // arXiv: 2103.10559, 2021.
- [4] Huang, Z. RIFE: Real-Time Intermediate Flow Estimation for Video Frame Interpolation / Z. Huang, T. Zhang, W. Heng, B. Shi, S. Zhou // arXiv: 2011.06294, 2021.
- [5] Lee, H. AdaCoF: Adaptive Collaboration of Flows for Video Frame Interpolation / H. Lee, T. Kim, T. Chung, D. Pak, Y. Ban, S. Lee // arXiv: 1907.10244, 2020.
- [6] Lin, G. RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation / G. Lin, A. Milan, C. Shen, I. Reid // arXiv: 1611.06612, 2016.
- [7] Ronneberger, O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // arXiv: 1505.04597, 2015.

Влияние реконструкции промежуточных кадров видеопоследовательности на результат 3D-реконструкции объектов

Ю.Х. Ганеева

Самарский национальный исследовательский университет
им. академика С.П. Королева

Самара, Россия

jganeeva@gmail.com

В.В. Мясников

Самарский национальный исследовательский университет
им. академика С.П. Королева

Самара, Россия

vmyas@geosamara.ru

Аннотация—Задача 3D-реконструкции объектов является одной из сложных задач в области компьютерного зрения. Как правило это связано с тем, что требуется большое количество входных данных, например изображений и параметров камеры. Часто для восстановления объектов данные имеются в виде ограниченного количества изображений из сети Интернет или кадров видеопоследовательности с низкой частотой. В данной работе производится оценка качества работы методов/инструментов 3D-реконструкции Akool, COLMAP и NeRF. Полученные результаты показали, что при недостаточном количестве входных изображений объекта, этап реконструкции промежуточных кадров позволяет повысить качество 3D-реконструкции.

Ключевые слова— 3D-реконструкция, NeRF, Akool, COLMAP, Alpha shapes, реконструкция кадров.

1. ВВЕДЕНИЕ

Задача 3D-реконструкции объектов/сцен [1] является одной из сложных задач в области компьютерного зрения. Результат реконструкции может быть использован в AR/VR приложениях, что является актуальным для создания реалистичного виртуального мира (например, Meta [2]). Для приближенной к реальности реконструкции необходимо достаточное количество изображений, позволяющее полностью описать внешний вид и форму объекта, что не всегда представляется возможным. В случае, когда требуется произвести 3D-реконструкцию динамического объекта/сцены, вопрос о количестве данных (как правило, частоте кадров видеопоследовательности или синхронизированных видеопоследовательностей) является основополагающим.

Реконструкция кадров видеопоследовательности — это процесс синтеза новых кадров видеопоследовательности на основе информации о соседних кадрах. В ситуации, когда набор данных расширяется за счёт реальных изображений с разных ракурсов – повышение качества реконструкции очевидно. В рамках исследования была выдвинута гипотеза, что синтез кадров с отсутствующих точек обзора, позволит повысить качество реконструкции. Однако, существующие лучшие среди известных (англ. state of the art, SOTA) методы реконструкции кадров по большей части не являются устойчивыми к динамике в кадре и движению камеры. В рамках работы для реконструкции промежуточных кадров видеопоследовательностей были выбраны методы XVFI [3], CDFI [4] и RRIN [5], показавшие SOTA качество реконструкции на популярном наборе данных Vimeo90K [6].

Целью данного исследования является оценка влияния этапа реконструкции кадров видеопоследовательности, описывающей внешний вид и форму объекта, на качество 3D-реконструкции. Для реализации 3D-реконструкции объектов было выбрано и оценено качество работы трёх методов/инструментов: Akool [7], COLMAP [8], NeRF [9].

Структура работы следующая. В разделе 2 приведено описание используемых для 3D-реконструкции методов. В разделе 3 представлены порядок проведения экспериментальных исследований и соответствующие результаты. Работа заканчивается заключением, в котором делаются выводы на основе результатов, полученных в ходе экспериментальных исследований.

2. ОПИСАНИЕ ИСПОЛЬЗУЕМЫХ МЕТОДОВ/ИНСТРУМЕНТОВ 3D-РЕКОНСТРУКЦИИ

Akool [7] – это инструмент, предназначенный для создания 3D-моделей высокого качества на основе SOTA алгоритмов искусственного интеллекта и алгоритмов SLAM [10]. Для получения 3D-модели необходимо лишь наличие RGB-изображений объекта. Все последующие вычисления производятся на удаленном(ых) сервере(ах). Результатом работы является полноценная 3D-модель (меш), описывающая форму и текстуру объекта. Стоит отметить, что предоставляется возможность одновременного создания нескольких 3D-моделей. Однако, конечная внутренняя реализация алгоритма является неизвестной.

COLMAP [8] – это инструмент, имеющий открытый исходный код и предназначенный для создания 3D-моделей из набора изображений. Результатом работы алгоритма реконструкции, реализованного в COLMAP, является облако точек. Однако полученное облако точек нуждается в постобработке, которая заключается в удалении фоновых точек с помощью инструмента MeshLab [11], и в создании на основе полученного облака точек меша с помощью алгоритма Alpha shapes [12].

NeRF [9] – это SOTA метод 3D-реконструкции объектов, который основан на синтезе новых видов для сложных сцен. Авторы получают представление сцены с помощью полносвязной (не свёрточной) глубокой нейронной сети, которая на вход принимает 5D-вектор координат (пространственные координаты (x, y, z) и радиальные координаты (θ, φ)). Нейронная сеть в результате выдает 4 числа: (R, G, B, σ) , где (R, G, B) – это значения для каждого цветового канала в заданной точке

пространства (x, y, z) , σ – значение в диапазоне $(0, 1)$, показывающее есть ли в заданной точке пространства объект.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Порядок проведения экспериментальных исследований следующий:

- Разбить исходную видеопоследовательность на кадры.
- Выполнить прореживание видеопоследовательности. При прореживании из исходной последовательности кадров удаляется каждый второй кадр.
- Выполнить увеличение частоты видеопоследовательностей в два раза путем синтеза промежуточных кадров одним из методов: XVFI, CDFI, RRIN.
- Произвести реконструкцию объекта по кадрам, полученным на шаге 2.
- Оценить результат 3D-реконструкции.

Для реализации процедуры оценки реконструкции было выполнено 3D-моделирование коробки на основе предварительных замеров, а также 3D-моделирование и 3D-печать конуса. Полученные 3D-модели были приняты за эталонные и использовались далее для оценки результата 3D-реконструкции.

Оценка качества 3D-реконструкции производилась с использованием метрики 3D IoU (Intersection-over-Union), которая вычисляется следующим образом:

$$IoU_{3D} = V_{\cap} / V_{\cup}$$

где V_{\cap} - объем пересечения мешей, V_{\cup} - объем объединения мешей.

Для проведения экспериментальных исследований были записаны две видеопоследовательности с частотой кадров 30 FPS, описывающие внешний вид и форму коробки и конуса, представленных на рис. 1. При съемке объекты находились в статическом положении, камера же двигалась вокруг объектов.



Рис. 1. Объекты для реализации 3D-реконструкции: коробка, конус

Для реализации реконструкции кадров и оценки качества 3D-реконструкции из видеопоследовательностей был вырезан каждый второй кадр. В результате были получены две видеопоследовательности с частотой кадров 15 FPS.

В таблице 1 представлены лучшие результаты 3D-реконструкции с синтезированными и оригинальными кадрами.

Таблица IV. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Метод 3D-реконструкции	Метод реконструкции кадров					
	XVFI (30 FPS)		Оригинал (30 FPS)		Оригинал (15 FPS)	
	Коробка	Конус	Коробка	Конус	Коробка	Конус
Akool	0.8026	0.7435	0.8548	0.8337	0.0000	0.0000
COLMAP + Alpha shape	0.5309	0.6332	0.6700	0.8409	0.4968	0.4801
NeRF	0.1145	0.5489	0.5964	0.5516	0.0000	0.0000

4. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования было выявлено, что кадры, синтезированные методами CDFI и RRIN не пригодны для использования в задаче 3D-реконструкции, так как результирующие кадры имеют большое количество артефактов. 3D-реконструкция инструментом Akool с использованием кадров, синтезированных методом XVFI, показала результаты близкие к результатам, полученным на оригинальных кадрах. Полученные результаты в дальнейшем планируется использовать в задаче 3D-реконструкции динамических объектов/сцены.

ЛИТЕРАТУРА

- [1] Гошин, Е.В. Реконструкция 3D-сцен по разноразмерным изображениям при неизвестных внешних параметрах съемки / В.А. Фурсов, Е.В. Гошин // Компьютерная оптика. – 2015. – Т. 39, № 5. – P. 770-776. DOI: 10.18287/0134-2452-2015-39-5-770-775.
- [2] Welcome to Meta | Meta [Электронный ресурс] — Режим доступа: <https://about.facebook.com/meta/> (06.02.2022).
- [3] Sim, H. XVFI: eXtreme Video Frame Interpolation / H. Sim, J. Oh, M. Kim // arXiv: 2103.16206, 2021.
- [4] Ding, T. CDFI: Compression-Driven Network Design for Frame Interpolation / T. Ding, L. Liang, Z. Zhu, I. Zharkov // arXiv: 2103.10559, 2021.
- [5] Li, H. Video Frame Interpolation Via Residue Refinement / H. Li, Y. Yuan, Q. Wang // ICASSP – IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2020. – P. 2613-2617. DOI: 10.1109/ICASSP40776.2020.9053987.
- [6] Xue, T. Video Enhancement with Task-Oriented Flow / T. Xue, B. Chen, J. Wu, D. Wei, W.T. Freeman // International Journal of Computer Vision. – 2019. – Vol. 127(8). — P. 1106-1125. DOI: 10.1007/s11263-018-01144-2.
- [7] Akool [Электронный ресурс]. — Режим доступа: <https://akool.com/> (31.01.2022).
- [8] COLMAP — COLMAP 3.7 documentation [Electronic resource]. — Режим доступа: <https://colmap.github.io/> (06.02.2022).
- [9] Mildenhall, B. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis / B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng // arXiv: 2003.08934, 2020.
- [10] Stachniss, C. Robotic Mapping and Exploration / Stachniss, C. – 2009. – Vol. 55.
- [11] Cignoni, P. MeshLab: an Open-Source Mesh Processing Tool / P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia // Eurographics Italian Chapter Conference. – 2008. – P. 129-136.
- [12] Fischer, K. Introduction to Alpha Shapes / K. Fischer. – 2009. – P. 17.

Влияние комплексирования спектральных каналов и сезонных космических снимков на различимость хвойного и лиственного леса

С.М. Зраенко

*Институт радиоэлектроники и информационных технологий - РТФ
Уральский федеральный университет
Екатеринбург, Россия
z_sm@mail.ru*

Аннотация—Произведено исследование по повышению различимости хвойных и лиственных пород лесной растительности для разных сезонов года при комплексировании спектральных каналов прибора ЕТМ+ спутника Landsat-7. Отличия анализируемых пород определялись расстоянием по яркости и спектральному углу между векторами спектральной яркости для каждого из их пикселей. При этом наибольшее отличие пород, наблюдавшееся для четвертого канала июльского снимка, увеличилось более чем в 1,35 раза после объединения его спектральных каналов. Наибольшая разница по спектральному углу была также получена для июльского снимка при объединении его каналов. Для большего увеличения различий между породами осуществлено одновременное комплексирование спектральных каналов и сезонных снимков. Наибольший выигрыш в различимости пород здесь был получен по яркости и составил 1,61 раза по сравнению с наилучшим результатом комплексирования только спектральных каналов. По сравнению с наилучшим результатом для одного канала выигрыш в различимости достиг 2,18 раза.

Ключевые слова— хвойная и лиственная растительность, снимки Landsat-7, спектральная яркость, спектральный угол, комплексирование спектральных каналов, комплексирование сезонных снимков.

1. ВВЕДЕНИЕ

Одной из задач при проведении лесоустроительных работ является определение видового состава растительности. При этом для определения формулы состава древостоя обычно оценивают соотношение видов, выделяя по 10 стволов в разных местах биотопа при полевых исследованиях. Трудоемкость такой работы обуславливает необходимость применения современных методов основанных на использовании мультиспектральных космических снимков. Указанные методы, наряду с используемой аэрофотосъемкой, в том числе и съёмкой с беспилотных летательных аппаратов позволяют существенно облегчить составление карт лесной растительности. При этом информативным признаком является спектральный коэффициент яркости [1]. Конечно, отражательная способность растительности зависит не только от состава, но и от ее фенологических фаз и состояния, определяемого погодными условиями сезона и года съёмки. Кроме этого на значения отражательной способности влияют радиометрическое, спектральное и пространственное разрешение съёмочной аппаратуры. Дополнительно к этому необходимо учитывать время и сезон съёмки с точки зрения изменения азимута и высоты Солнца, крутизну и экспозицию поверхности, характеристики прозрачности атмосферы,

скорость и направление ветра. Отмеченные особенности существенно усложняют применение данных дистанционного зондирования для построения карт лесной растительности. Актуальность проблемы подтверждается активными исследованиями по формированию спектральных библиотек растительных и природных объектов на основе спутниковых снимков [2–4], мониторингу состояния этих объектов и их классификации по мультиспектральным и гиперспектральным данным [5–7].

В настоящей работе представлены новые результаты относительно представленных ранее [8, 9] по формированию классификационных признаков хвойной и лиственной растительности. Эти признаки основаны на комплексировании не только спектральных яркостей, но и их значений, соответствующих разным сезонам года. Полученные при этом объединенные классификационные признаки могут быть использованы для повышения надежности выделения лесных пород.

2. МЕТОДЫ ИССЛЕДОВАНИЯ

Для проведения исследования использовались космические снимки спутника Landsat-7 лесных участков с хвойной и лиственной растительностью.

В качестве классификационных признаков этих объектов выбраны их спектральные яркости в шести спектральных каналах прибора ЕТМ+ спутника Landsat-7. Это все каналы ЕТМ+ с пространственным разрешением 30 метров. Их номера 1, 2, 3, 4, 5, 7.

По результатам проведенного анализа определены спектральные каналы, в которых отличия пород максимальны для разных сезонов. Эти отличия определялись в соответствии с соотношением (1) в котором использованы математическое ожидание и среднеквадратичное отклонение для разности яркостей:

$$\Delta_k = M\{D_k^{ij}\} - 2\sigma\{D_k^{ij}\}. \quad (1)$$

В этом выражении расстояние по яркости между каждым пикселем лиственной и каждым пикселем хвойной растительности в k -м спектральном канале рассчитывалось на основе соотношения:

$$D_k^{ij} = B_k^{Fi} - B_k^{Cj}, \quad i = 1, \dots, N_F, \quad j = 1, \dots, N_C, \quad (2)$$

где N_F и N_C – количество пикселей лиственных и хвойных пород в известных фрагментах изображения, а B_k^{Fi} и B_k^{Cj} соответственно – их яркости в k -м канале.

Для повышения различимости растительности осуществлено комплексирование спектральных каналов. При этом в качестве одного из новых информативных признаков используется модуль вектора в k -мерном пространстве яркостей, базис которого образован спектральными каналами ЕТМ+. Степень отличия лиственного от хвойного объекта при этом определяется модулем разности их векторов.

В качестве второго информативного признака используется спектральный угол. Это угол между векторами лиственной и хвойной растительности.

Следует отметить, что используемое пространство спектральных яркостей не является ортогональным. В этой связи необходимо говорить о разложении векторов объектов в некотором условно ортогональном базисе. Таким образом, условными являются модули векторов яркости растительностей, расстояние и угол между ними. Несмотря на это, значения этих параметров позволяют оценить целесообразность объединения спектральных каналов для повышения различимости хвойных и лиственных пород.

Дальнейшее повышение различимости пород можно обеспечить комплексированием снимков разных сезонов наряду с комплексированием спектральных каналов. Одновременное комплексирование сезонных снимков и снимков спектральных каналов осуществлялось так же на основании отличий растительных пород по яркости и спектральному углу. Далее осуществлялось вычисление статистических характеристик для этих параметров – математического ожидания и среднеквадратичного отклонения по совокупности всех пикселей анализируемых участков растительности.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Было показано, что наибольшее отличие между породами 14,36 получено для четвертого канала июльского снимка при представлении изображений в формате одного байта на пиксель [8].

При комплексировании спектральных каналов наилучшая комбинация увеличила различия растительности до 19,44. Наибольшие отличия между породами по спектральному углу составляют 5,72 градуса.

При объединении спектральных каналов и сезонных снимков рассматривалось два варианта. В первом при поиске наилучшей комбинации снимков разных сезонов выбирался спектральный канал, в котором отличие пород максимально для каждого из комплексированных месяцев. Во втором при комплексировании месяцев в них выбирались спектральные каналы уже давшие наилучшие результаты по отличиям пород при их объединении. В первом варианте при объединении сезонных снимков наибольшее отличие пород по яркости составило 28,05. Во втором варианте это значение увеличилось до 31,35. Таким образом, объединение спектральных и сезонных снимков дает выигрыш в различимости пород в 1,61 раза по сравнению с наилучшим результатом при комплексировании только спектральных каналов. Если же сравнивать с наилучшим результатом для одного канала, то выигрыш в различимости достигает 2,18 раза.

Что касается спектрального угла, то при объединении сезонных снимков выигрыша в различимости

растительных пород по сравнению с объединением каналов получено не было.

4. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования определены каналы ЕТМ+ спутника Landsat-7 с максимальными яркостными различиями хвойной и лиственной растительности для разных сезонов года. Для увеличения различимости хвойной и лиственной растительности, произведено объединение спектральных каналов. При этом наибольшая разница в яркости пород по сравнению с одним каналом увеличилась более чем в 1,35 раза. Наибольшая разница для растительности по спектральному углу составляет 5,72 градуса. Для большего увеличения различий между породами осуществлено одновременное комплексирование спектральных каналов и сезонных снимков. При этом рассмотрены два варианта комплексирования. Максимальный выигрыш в различимости пород составил 1,61 раза по сравнению с наилучшим результатом при комплексировании только спектральных каналов и 2,18 раза по сравнению с наилучшим результатом для одного канала. Полученные результаты позволяют повысить эффективность выделения хвойной и лиственной растительности на спутниковых снимках.

БЛАГОДАРНОСТИ

Работа поддержана РФФИ, контракт № 19–29–09022.

ЛИТЕРАТУРА

- [1] Книжников, Ю.Ф. Спектральная яркость / Ю.Ф. Книжников [Электронный ресурс]. – Режим доступа: <http://www.lomonosov-fund.ru/enc/ru/encyclopedia:0132521> (07.02.2022).
- [2] Jet Propulsion Laboratory, ECOSTRESS spectral library. Version 1.0. – 2018 [Electronic resource]. – Access mode: <https://speclib.jpl.nasa.gov> (07.02.2022).
- [3] Zomer, R.J. Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing / R.J. Zomer, A. Trabucco, S.L. Ustin // J. Environ. Management. – 2009. – Vol. 90(7). – P. 2170-2177. DOI:10.1016/j.jenvman.2007.06.028.
- [4] Boori, M.S. Comparison of hyperspectral and multi-spectral imagery to building a spectral library and land cover classification performance / M.S. Boori, R.A. Paringer, K. Choudhary, A.V. Kupriyanov // Computer Optics. – 2018. – Vol. 42(6). – P. 1035-1045. DOI: 10.18287/2412-6179-2018-42-6-1035-1045.
- [5] Варламова, А.А. Информационная технология обработки данных ДЗЗ для оценки ареалов растений / А.А. Варламова, А.Ю. Денисова, В.В. Сергеев // Компьютерная оптика. – 2018. – Т. 42, № 5. – С. 864-876. DOI: 10.18287/2412-6179-2018-42-5-864-876.
- [6] Yukun, L. A change detection method using spatial-temporal-spectral information from Landsat images / L. Yukun, Zh. Lifu, W. Nan // Int. J. Remote Sens. – 2020. – Vol. 41(2). – P. 772-793. DOI: 10.1080/01431161.2019.1648905.
- [7] Ghazaryan, G. Vegetation monitoring satellite time series: An integrated approach for user-oriented knowledge extraction / G. Ghazaryan, O. Dubovyk, V. Graw, S. Jchellberg // Proc. SPIE. – 2018. – Vol. 10783. DOI: 10.1117/12.2325762.
- [8] Zraenko, S. Increasing the distinctiveness of forest species composition by satellite images / S. Zraenko // CEUR-WS. – 2020. – Vol. 2665. – P. 109-112.
- [9] Zraenko, S.M. Combining Channels to Increase the Differences Between Coniferous and Hardwood Vegetation in Satellite Images / S.M. Zraenko // IEEE Xplore. – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649083.

Оптимизация траектории движения транспортных средств в задаче управления транспортными потоками на перекрестке

А.С. Юмаганов

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
yumagan@gmail.com

А.А. Агафонов

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
ant.agafonov@gmail.com

Аннотация—В работе представлен метод формирования оптимальной траектории движения транспортных средств на контролируемом перекрестке. Представлены результаты экспериментальных исследований разработанного метода в среде симуляции SUMO.

Ключевые слова— интеллектуальные транспортные системы, траектория движения, управление транспортным средством.

1. ВВЕДЕНИЕ

В настоящее время наблюдается неуклонный рост количества транспортных средств на дорогах. Это приводит к существенной нагрузке на транспортную инфраструктуру городов, которая выражается в увеличении числа "пробок" на дорогах, загрязнении окружающей среды, увеличении числа транспортных происшествий. Для снижения нагрузки на транспортную сеть города применяются различные методы управления движением транспортных средств на перекрестках. Выделяют несколько подходов для решения задачи эффективного управления дорожным трафиком на перекрестках.

Первый подход основан непосредственно на управлении сигналами светофора. Методы основанные на классических подходах управления сигналами светофоров [1, 2] используются для управления одним изолированным перекрестком. В последнее время в связи с увеличением числа различного рода датчиков, устанавливаемых на транспортные средства (ТС) и элементы дорожной инфраструктуры, достаточно большое развитие получили методы на основе машинного обучения [3], которые позволяют управлять в том числе и несколькими перекрестками.

С появлением подключенных транспортных средств (connected and automated vehicles, CAV) получили развитие методы эффективного управления дорожным трафиком на перекрестках, основанные на оптимизации траекторий движения ТС [4, 5]. В рамках данной группы методов для каждого ТС в процессе прохождения перекрестка формируется оптимальная траектория движения, полученная с учетом информации о ближайших ТС и объектах транспортной инфраструктуры. В качестве критерия оптимизации обычно используются следующие: потребление топлива, время ожидания, время прохождения перекрестка и т.д. В данной работе представлен метод формирования оптимальной траектории движения транспортных средств на контролируемом перекрестке.

Представленный в данной работе метод является модификацией метода [4], в основе которого лежит Shooting Heuristic алгоритм.

2. SHOOTING HEURISTIC АЛГОРИТМ

В рамках представленного в работе [4] метода, формирование траекторий движения транспортных средств осуществляется следующим образом. На первом этапе формируется исходная траектория движения с параметрами $(\bar{a}^f, \underline{a}^f, \bar{a}^b, \underline{a}^b, v)$, которая затем подается на вход разработанного авторами алгоритма оптимизации с заданной целевой функцией. В процессе работы алгоритма параметры $(\bar{a}^f, \underline{a}^f, \bar{a}^b, \underline{a}^b, v)$ определенным образом обновляются до тех пор, пока полученная на их основе траектория не будет являться оптимальной.

Для формирования траектории движения ТС используется Shooting Heuristic алгоритм. В рамках данного алгоритма для всех транспортных средств, въезжающих на заданный перекресток формируется вектор, содержащий начальное местоположение ТС, его скорость и время. Shooting Heuristic алгоритм построит физически возможную траекторию от начального местоположения ТС до перекрестка с предварительно определенным набором параметров, включающим ускорение и ограничение скорости.

Shooting Heuristic алгоритм содержит два основных процесса: forward shooting (FSP) и backward shooting (BSP). В ходе FSP для рассматриваемого транспортного средства с некоторым начальным состоянием формируется траектория, при движении по которой ТС разгоняется до скорости v с заданным ускорением \bar{a}^f . Если полученная траектория не пересекает траекторию впереди идущего ТС и находится на безопасном расстоянии от него, то FSP вернет безопасную траекторию. Если же полученная траектория не является безопасной, то FSP вернет траекторию, являющуюся результатом её плавного слияния с так называемой "теневой" траекторией впереди идущего транспортного средства. В процессе слияния двух траекторий используется отрицательное ускорение \underline{a}^f .

Если полученная в ходе процесса FSP траектория пересекает перекресток на красный сигнал светофора, то для данной траектории запускается BSP процесс. В результате процесса BSP участок траектории содержащий выход на перекресток смещается к началу следующей зеленой фазы, строится траектория

полученная путем плавного слияния смещенного участка и исходной траектории. В процессе построения такой траектории используется ускорение \bar{a}^b и отрицательное ускорение \underline{a}^b . Полученная траектория является результатом работы Shooting Heuristic алгоритма для рассматриваемого ТС.

Таким образом Shooting Heuristic алгоритм последовательно формирует траектории для каждого ТС на дорожной полосе.

3. ПРЕДЛАГАЕМЫЙ МЕТОД

Представленный в работе [4] метод формирования траектории движения ТС был предназначен для однопольного перекрестка и не учитывал возможные перестроения ТС. В данной работе представлена модификация метода [4], которая позволяет применять данный метод для многополосных дорог. Перестроения могут возникнуть после пересечения ТС текущего перекрестка, в том случае, если на следующем по ходу движения смежном перекрестке ТС необходимо проехать по полосе движения отличной от той, на которую ТС въехало с текущего перекрестка.

В рамках разработанной модификации для каждой дорожной полосы формируется набор траекторий ТС, находящихся на соответствующей полосе. В случае перестроения, в результате которого на одной из дорожных полос появилось / пропало ТС выполняется следующее.

- 1) Если на дорожной полосе появилось новое транспортное средство, то для него строится траектория с учетом «теневой» траектории ближайшего впереди идущего ТС.
- 2) Для всех ТС расположенных на большем расстоянии от перекрестка, чем появившееся / пропавшее в результате перестроения ТС, выполняется перестроение траекторий в порядке удаленности от перекрестка.
- 3) Наконец, с помощью полученного набора траекторий выполняется управление движением ТС на соответствующей полосе движения.

4. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Экспериментальные исследования разработанного метода были проведены в системе моделирования SUMO на трех сценариях.

Сценарий «1x1». Транспортная сеть представляет собой один регулируемый светофором перекресток, состоящий из двух контролируемых проезжих частей: однопольной и двупольной.

Сценарий «1x1 v2» Транспортная сеть представляет собой один регулируемый светофором перекресток, состоящий из четырех контролируемых двупольных проезжих частей.

Сценарий «cologne8» Транспортная сеть представляет собой участок транспортной сети города Кельн, содержащий 8 регулируемых перекрестков разного вида.

Экспериментальные исследования проводились при шаге симуляции, равном 0,1 секунде, и при общем времени симуляции 3600 секунд. В таблице 1 представлены результаты сравнения времени ожидания на светофоре при использовании разработанного метода управления траекторией движения ТС и без него. При этом переключение фаз светофоров осуществлялось часто используемым методом Uniform, при котором смена фаз светофора осуществляется через заранее фиксированные промежутки времени.

Таблица V. ОЦЕНКА ЭФФЕКТИВНОСТИ РАЗРАБОТАННОГО МЕТОДА

Сценарий	Общее количество во ТС	Среднее время ожидания, с	
		Предлагаемый метод + Uniform	Uniform
«1x1»	900	0,001	10,06
«1x1 v2»	1080	4,01	20,41
«cologne8»	2046	3,80	26,01

Как видно из полученных результатов применение предлагаемого метода позволяет существенно понизить среднее время ожидания ТС на перекрестке.

5. ЗАКЛЮЧЕНИЕ

В работе предложен метод формирования оптимальной траектории движения транспортных средств на контролируемом перекрестке. Представлены результаты экспериментальных исследований, подтверждающие работоспособность предложенного метода.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Российского научного фонда (проект № 21-11-00321, <https://rscf.ru/en/project/21-11-00321/>).

ЛИТЕРАТУРА

- [1] Мясников, В.В. Детерминированная прогнозная модель управления сигналами светофоров в интеллектуальных транспортных и геоинформационных системах / В.В. Мясников, А.А. Агафонов, А.С. Юмаганов // Компьютерная оптика. – 2021. – Т. 45, № 6. – С. 917-925. DOI: 10.18287/2412-6179-CO-1031.
- [2] Varaiya, P. The max-pressure controller for arbitrary networks of signalized intersections / P. Varaiya // Advances in dynamic network modeling in complex transportation systems. Vol. 2. Complex networks and dynamic systems. – New York: Springer, 2013. – P. 27-66. DOI: 10.1007/978-1-4614-6243-9_2.
- [3] Wei, H. CoLight: Learning network-level cooperation for traffic signal control. / H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, Z. Li // Proc 28th ACM Int Conf on Information and Knowledge Management. – 2019. – P. 1913-1922. DOI: 10.1145/3357384.3357902.
- [4] Ma, J. Parsimonious shooting heuristic for trajectory design of connected automated traffic part II: Computational issues and optimization / J. Ma, X. Li, F. Zhou, J. Hu, B. Brian Park // Transportation Research Part B: Methodological. – 2017. – Vol. 95(B). – P. 421-441. DOI: 10.1016/j.trb.2016.06.010.
- [5] Guo, Y. DRL-TP3: A learning and control framework for signalized intersections with mixed connected automated traffic / Y. Guo, J. Ma // Transportation Research Part C: Emerging Technologies. – 2021. – Vol. 132. DOI: 10.1016/j.trc.2021.103416.

Аппаратно-программный комплекс для составления паспорта огнестрельного повреждения

И.Г. Пальчикова

*Конструкторско-технологический
институт научного приборостроения
Сибирского отделения
Российской академии наук
Новосибирск, Россия
palchikova@gmail.com*

Е.С. Смирнов

*Конструкторско-технологический
институт научного приборостроения
Сибирского отделения
Российской академии наук
Новосибирск, Россия
evgenii.s.smirnov@yandex.ru*

И.А. Будаева

*Конструкторско-технологический
институт научного приборостроения
Сибирского отделения
Российской академии наук
Новосибирск, Россия
i.budaeva@g.nsu.ru*

И.В. Латышов

*Санкт-Петербургский университет
Министерства внутренних дел
Российской Федерации
Санкт-Петербург, Россия
latyшов@gmail.com*

В.А. Васильев

*Волгоградская академия
Министерства внутренних дел
Российской Федерации
Волгоград, Россия
v-vasiliev@inbox.ru*

А.В. Кондаков

*Санкт-Петербургская академия
Следственного комитета
Российской Федерации
Санкт-Петербург, Россия
akondakov77@mail.ru*

Аннотация—Для автоматизации решения задач по формированию паспорта повреждения разработан аппаратно-программный комплекс, состоящий из мобильного регистрирующего устройства с спектрозональным и мультиспектральным осветителем, специализированного программного обеспечения и алгоритмов обработки цифровых изображений мишеней для количественного определения и анализа криминалистически значимых характеристик объекта исследования и выявленных на нём следов выстрела.

Ключевые слова— паспорт повреждения, следы выстрела, мишень, аппаратно-программный комплекс.

1. ВВЕДЕНИЕ

Научно-исследовательские работы по сбору и систематизации сведений о стрелковом огнестрельном оружии [1, 2] привели к выделению строго определённых криминалистически значимых характеристик следов выстрела [3, 4]. Обычно в судебной баллистике применяются обобщённые оценочные подходы [5, с. 36–37]. Визуальная оценка характеристик следов на мишени зависит от способностей и функций эксперта, и вместе с тем она устанавливает контрольные точки для сравнения с инструментальным измерением. Визуальную оценку нелегко провести, а тем более воспроизвести (повторить) даже с тренированными экспертами, потому что органолептические критерии подвержены влиянию персональных предпочтений, освещения, недостатков зрения и иных факторов. Кроме того, различные экспертные сообщества применяют различные шкалы для описания и оценки некоторых характеристик, например, цвета. В то же время инструментальный метод, основанный на математической обработке цветного цифрового изображения, позволяет [6, 7] объективно выполнить измерения криминалистически значимых характеристик на цифровом изображении мишени и выразить результаты в количественном виде. Значимые характеристики определены и подробно описаны в [7, 8].

В настоящей работе представлены результаты разработки аппаратно-программного комплекса, состоящего из

мобильного регистрирующего устройства, специализированного программного обеспечения и алгоритмов обработки цифровых изображений мишеней для количественного определения и анализа криминалистически значимых характеристик объекта исследования, а также выявленных на нём следов выстрела.

2. НАЗНАЧЕНИЕ И СОСТАВ УСТРОЙСТВА

Структура аппаратно-программного комплекса, в котором для определения, расчета и систематизации криминалистически значимых характеристик мишени используются подходы компьютерного зрения, принципы получения и обработки цифровых изображений, представлена на рис. 1.

Для освещения исследуемой мишени предусмотрено использование любых из семи осветителей, встроенных в автономный спектрозональный и мультиспектральный осветительный прибор с функцией белого света с высоким индексом цветопередачи “Фотобокс 3138” [9]. Освещение и условия наблюдения выбираются по ГОСТ Р 52489-2005 с учетом свойств испытуемого образца и информации, которую необходимо получить при измерении.

Оптическая система объектива (рис.1) выполняет перенос изображения мишени в плоскость фоточувствительной матрицы регистратора. В качестве регистратора и основного элемента измерительного блока используется цифровая цветная фотокамера Canon EOS 500D (Canon Inc., Japan), характеристики фотокамеры и необходимые устанавливаемые параметры фотосъемки указаны в [6]. Работа регистратора обеспечивается управлением от компьютера.

3. ПОЛУЧЕНИЕ И ОБРАБОТКА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ

Процесс определения характеристик включает подготовку мишени, получение и регистрацию ее изображения с помощью цифровой камеры, которое затем пересылается в компьютер, где специально созданное десктопное приложение ImgOpinion выполняет его обработку

и представляет результаты на мониторе или в виде файла, пригодного для использования с базами данных.

Цифровые изображения регистрируются в режиме пользовательской настройки камеры в 14-битном формате RAW, преобразуются в адаптируемый 16-битный формат TIFF, который не ограничивает количество градаций яркости, содержащихся в 14-битном изображении. По нашему опыту, помимо формата TIFF, для работы с 16-битными цифровыми изображениями подходят также форматы PNG и DICOM. Эти форматы не вносят дополнительных искажений в исходное изображение.

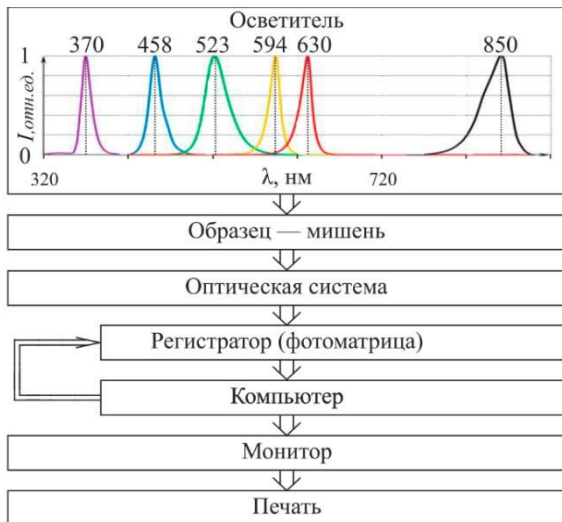


Рис. 1. Структура аппаратно-программного комплекса

Специализированное программное обеспечение *ImgOpinion* выполняет оптико-структурный анализ цифровых изображений, сегментацию изображения, расчет количественных характеристик для выявленных зон на следах выстрела, оценивает количество копоти выстрела [7] и ее распределение на поверхности мишени. *ImgOpinion* предназначено для использования в экспертных лабораториях. Приложение написано на языке программирования Java и работает на системах семейств Linux и Windows, системные требования приложения совпадают с [10]. Окно программы *ImgOpinion* в режиме "автоматического поиска зон на следах выстрела" показано на рис. 2, где в левой части рисунка расположены строки меню и переключатели окон, а справа – окошко для обработанного изображения, где показаны контуры зон отложения, определенные алгоритмом сегментации Otsu [7].

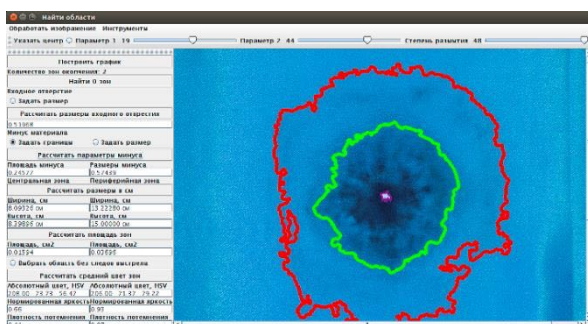


Рис. 2. Основное окно десктопного приложения *ImgOpinion*

В результате работы приложения *ImgOpinion* формируется паспорт повреждения [8], который представляет собой информационный ресурс, содержащий сведения о виде

поступившего объекта со следами выстрела, характеристиках его поверхности, морфологии огнестрельного повреждения и топографии отложения дополнительных следов выстрела.

Паспорт огнестрельного повреждения содержит следующие данные: характеристика объекта исследования; виды повреждения и следов выстрела; степень деструктивного действия повреждающих факторов выстрела; морфологические характеристики основного следа выстрела, наличие штандмарки и механизм образования повреждения; характеристики формы, размеров и ориентации разрушенных участков; количество, форма и размеры зон копоти; цвет копоти и ее колориметрические характеристики (по зонам); наличие, вид, цвет и локализация преимущественного отложения зерен пороха; размеры зоны; форма и размеры зон механического и термического воздействия газопороховой струи.

4. ВЫВОДЫ

Разработан аппаратно-программный комплекс для автоматизации решения задач по формированию паспорта повреждения, что позволит в дальнейшем создавать системы данных для ускорения процесса расследования происшествий и установления обстоятельств выстрела.

ЛИТЕРАТУРА

- [1] Латышов И.В. Стрелковое огнестрельное оружие и его следы на пулях, гильзах и преградах. Часть 1. 7,62 мм пистолет обр. 1933 г. (ТТ) (справочно-методическое пособие) / И.В. Латышов, И.И. Никитин, В.В. Сидоров, И.А. Чулков. – Волгоград: Перемена, 2001. – 83 с.
- [2] Карданов, Р.Р. Использование современных технологий для систематизации следовой картины применения огнестрельного оружия // Криминалистические средства и методы в раскрытии и расследовании преступлений: материалы VI Всероссийской научно-практической конференции по криминалистике и судебной экспертизе с международным участием. – М., 2014. – 216 с.
- [3] Беляков, А.Л. Оружиеведение: Часть 3. Судебная баллистика / А.Л. Беляков, А. Н. Матюшенко, Т. В. Попова. – Челябинск: Челябинский юридический институт МВД России, 2004. – 200 с.
- [4] Смирнов, В.Е. Закономерности развития газопороховой струи при выстреле из стрелкового оружия // Экспертная техника. – 1987. – № 65. – С. 34-45.
- [5] Латышов, И.В. Стрелковое огнестрельное оружие и его следы на пулях, гильзах и преградах. Часть 28. 5,45 мм автомат АК-105 (справочник) / И.В. Латышов, И.А. Чулков, А.С. Копанев, М.А. Оздоев. – Волгоград: ВА МВД России, 2017. – 76 с.
- [6] Пальчикова, И.Г. Цветовой анализ цифровых изображений при производстве экспертных исследований следов выстрела / И.Г. Пальчикова, И.В. Латышов, А.В. Кондаков, В.А. Васильев, Е.С. Смирнов // Доклады Академии наук Высшей школы Российской Федерации. – 2015. – № 2(27). – С. 88-101.
- [7] Palchikova, I.G. Computer vision in analyzing the propagation of a gas-powder jet / I.G. Palchikova, I.V. Latyshov, E.S. Smirnov, V.A. Vasiliev, A.V. Kondakov, I.A. Budaeva // Sensors. – 2022. – Vol. 22(1) [Electronic resource]. — Access mode: <https://www.mdpi.com/1424-8220/22/1/6> (06.02.2022). DOI: 10.3390/s22010006.
- [8] Латышов, И.В. Паспорт огнестрельного повреждения как интегративная часть инновационных аппаратно-программных комплексов / И.В. Латышов, И.Г. Пальчикова, А.В. Кондаков, В.А. Васильев, Е.С. Смирнов // Судебная экспертиза. – 2020. – № 2(62). – С. 58-65. DOI: 10.25724/VAMVD.NMNO.
- [9] Пальчикова, И.Г. Автономный спектрональный осветительный прибор с функцией белого света с высоким индексом цветопередачи / И.Г. Пальчикова, Е.В. Карамшук, Е.С. Смирнов, Е.И. Пальчиков, М.С. Самойленко // Приборы и техника эксперимента. – 2021. – № 3. – С. 155-157. DOI: 10.31857/S0032816221030241.
- [10] Photoshop system requirements [Электронный ресурс]. – Режим доступа: <https://helpx.adobe.com/photoshop/system-requirements.html> (15.12.2021).

Исследование методов сегментации для выделения областей интереса на рентгеновских снимках лёгких

Д.В. Машков
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
dima-mashkov00@mail.ru

Н.Ю. Ильясова
Институт систем обработки
изображений - филиал ФНИЦ
«Кристаллография и фотоника»
РАН
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
ilyasova.nata@gmail.ru

Н.С. Демин
Институт систем обработки
изображений - филиал ФНИЦ
«Кристаллография и фотоника»
РАН
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
volfgunus@gmail.com

Аннотация—В статье представлены различные подходы сегментации лёгких реализованные с использованием библиотеки Scikit-Learn и языка программирования Python, а также проведён сравнительный анализ их эффективности для повышения точности выделения областей интереса в задаче диагностики заболеваний лёгких. Набор данных, используемый для исследования, получен из базы данных пациентов клиники «Дорожная клиническая больница на ст. Самара ОАО РЖД».

Ключевые слова— биомедицинские изображения, зоны интереса, автоматическая сегментация, классификаторы, рентгеновские снимки лёгких.

1. ВВЕДЕНИЕ

Для того, чтобы предотвратить сложное течение болезни, необходима своевременная диагностика и точное установление диагноза. Рентгеновские снимки лёгких обычно используются в пульмонологии. Поскольку обследование с использованием рентгена грудной клетки является относительно быстрой и дешевой процедурой, его можно использовать в качестве надлежащего диагностического инструмента для скрининга заболеваний органов дыхания, таких как пневмония и туберкулез. Тема представляет практический интерес, так как в последние годы для ускорения проведения анализа медицинских изображений стало распространённым использование технических средств и цифровых технологий распознавания изображений [1]. Более того, сегментация рентгеновских снимков лёгких имеет важное значение при определении диагноза болезни [2]. Разработчики совместно с врачами представляют программное обеспечение для автоматической сегментации изображений с целью облегчить и сделать анализ снимков более удобным и быстрым.

На данный момент есть множество способов классификации медицинских изображений. Текстурные признаки и ранее показывали хорошие результаты при распознавании биомедицинских изображений и их дальнейшей диагностики [3]. Также в статье о клетках крови [4] проводится выявление патологий с использованием текстурных характеристик в различных цветовых спектрах. В этой работе предложена технология выделения областей интереса, разработанная на основе метода классификации изображений k – средних. Для

повышения точности изображения пользуются коэффициентами корреляции, которые описаны в работе [5].

Основная цель данной работы – выполнить сегментацию разными способами с акцентами на трёх областях: лёгкие, тело человека и задний фон. Для решения поставленной задачи мы имеем на руках сырую и не размеченную базу данных пациентов клиники «Дорожная клиническая больница на ст. Самара ОАО РЖД». База данных включала в себя рентгенограммы пациентов разных полов, старше 18 лет. Все изображения были в формате (.bmp) с разрешением 512x512. Для нашей задачи необходимо было выделить три области интереса: лёгкие, тело человека и задний фон. Для этого врачом-экспертом вручную были размечены данные в датасете.

Подобное исследование с учётом специфики данных, числа классов и способов сегментации проводилось впервые. В рамках работы также выявлялись факторы, которые необходимо учитывать для получения качественной сегментации изображений лёгких.

2. СЕГМЕНТАЦИЯ РЕНТГЕНОВСКИХ СНИМКОВ ЛЁГКИХ

В данной работе сегментация изображения проводится в две фазы: выделение признаков с предварительной подготовкой снимка и обучение модели с последующей сегментацией. Перед расчётом признаков выполняются некоторые предварительные операции: перевод изображений в оттенки серого и нормализация изображений. Результатом данного шага является более чёткая и резкая рентгенограмма лёгких в оттенках серого.

Для проведения сегментации использовался текстурный анализ на основе признаков Харалика [6]. Вычисление этих признаков выполнялось средствами библиотеки Scikit-Learn. На следующем этапе после получения необходимых численных значений нужно выполнить объединение масок с отдельными зонами в одну для каждого снимка.

Следующим шагом является разделение данных на тренировочные и тестовые, обучение модели и сегментация тестовых изображений. Проводим сегментацию рентгеновских снимков лёгких, используя

нижеперечисленные методы анализа изображений: дискриминантный анализ, алгоритм KNN, логистическая регрессия, дерево решений и Random forest.

Ниже приведена таблица 1, отражающая численные значения метрики precision, демонстрирующая способность отличать данный класс от других классов. На рисунке 1 представлено исходное изображение (рис. 1а) с результатами сегментации рентгеновского снимка лёгких. Из представленных расчётов можно сделать вывод, что дерево решений в экспериментальных условиях, в совокупности с представленными данными, является самым эффективным, так как по результатам сегментации метод показал себя быстрым и точным.

Дискриминантный анализ продемонстрировал точность близкую к значению 0.5, что является низким показателем. Такой же результат прослеживается и в методе логистической регрессии. Метод К-соседей показал неплохие результаты наравне с деревом решений и Random forest, но является несколько более медленным по времени анализа снимков, чем дерево решений. Random forest, демонстрируя точные результаты, является неэффективным из-за длительного времени работы.

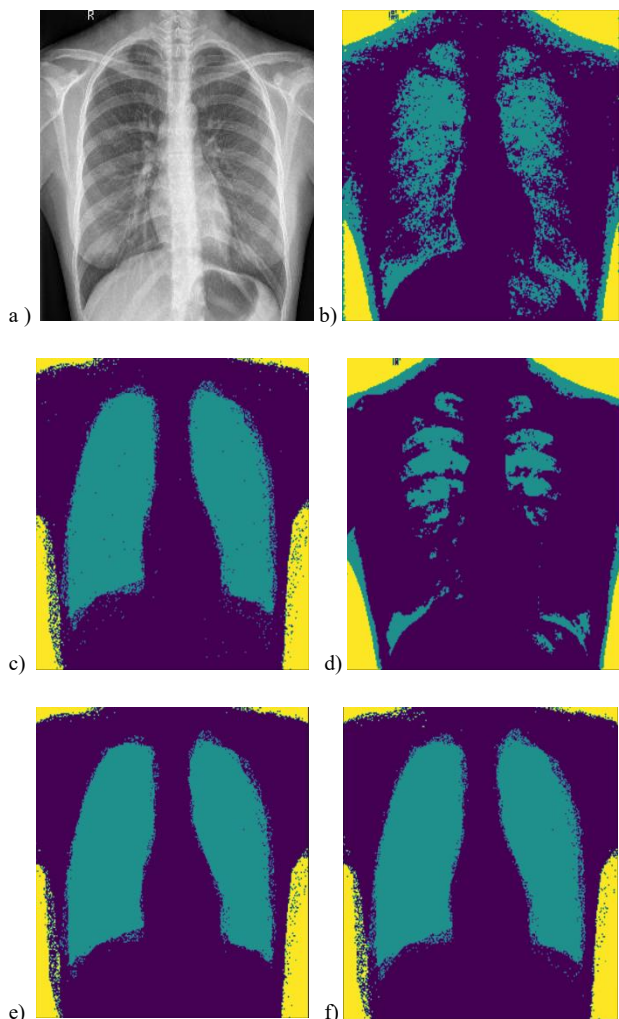


Рис. 1. Результаты экспериментальных исследований. а) Начальное изображение б) Дискриминантный анализ с) Метод К-соседей д) Логистическая регрессия е) Дерево решений ф) Random forest

Таблица 1. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

		Показатели эффективности			
		Время, с	Точность определения тела человека	Точность определения лёгких	Точность определения заднего фона
АЛГОРИТМЫ	Дискриминантный анализ	6.504	0.69	0.57	0.62
	Метод К-соседей	6.577	0.86	0.88	0.73
	Логистическая регрессия	14.651	0.64	0.56	0.61
	Дерево решений	4.683	0.86	0.9	0.74
	Random forest	116.50	0.87	0.89	0.73

3. ЗАКЛЮЧЕНИЕ

Методы компьютерного анализа изображений, являясь удобным инструментом, помогают человечеству во многих аспектах жизни, но особенно это заметно в медицине. За долгое время развития цифровых технологий человек сделал алгоритмы сегментации достаточно совершенными, чтобы проводить анализ биомедицинских изображений, опираясь на ограниченные объёмы данных. В ходе работы были реализованы несколько методов сегментации лёгких, а также проведён сравнительный анализ их эффективности. Из проведённых исследований можно увидеть разницу между самым точным и самым неточным из приведённых алгоритмов. Так, разница точности определения лёгких составляет 36%.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке РФФИ (грант № 19-29-01135), Министерства науки и высшего образования РФ, в рамках выполнения работ по государственному заданию ФНИЦ «Кристаллография и фотоника» РАН.

ЛИТЕРАТУРА

- [1] Бурсов, А.И. Применение искусственного интеллекта для анализа медицинских данных / А. И. Бурсов. – Альманах клинической медицины. – 2019. – № 47(7). – С. 630-633.
- [2] Zhou, B. Lung mass density analysis using deep neural network and lung ultrasound surface wave elastography / Boran Zhou, Xiaoming Zhang // Ultrasonics. – 2018. – № 89. – P. 173-177.
- [3] Ilyasova, N. Particular Use of BIG DATA in Medical Diagnostic Tasks / N. Ilyasova, A. Kupriyanov, R. Paringer // Pattern Recognition and Image Analysis. – 2018. – № 28(1) – P. 114-121.
- [4] Nikitaev, V. Sposob raspoznavaniya izobrazheniya tekstury kletok / V. Nikitaev, B. Flury // Biometrika. – 2010. – № 97(1) – P. 33-41.
- [5] Ilyasova, N. Formation of features for improving the quality of medical diagnosis based on discriminant analysis method / N. Ilyasova, A. Kupriyanov, R. Paringer // Computer Optics. – 2014. – Vol. 38(4). – P. 851-856. DOI: 10.18287/0134-2452-2014-38-4-851-855.
- [6] Haralick, R.M. Textural features for image classification / R.M. Haralick, K. Shanmugam, Dinstein // IEEE Transactions on Systems, Man, and Cybernetics. – 1973. – Vol. SMC-3. – P. 610-621.
- [7] Мясников, В.В. Методы обнаружения и распознавания объектов на цифровых изображениях: учеб. пособие / В.В. Мясников, Н.И. Глузов, В.В. Сергеев. – Самара: Изд-во Самара. гос. аэрокосм. ун-та, 2006. – 168 с.
- [8] Еременко, Ю.И. Применение метода k-средних для идентификации пользователя по клавиатурному почерку / Ю.И. Еременко, Ю.С. Олюнина // Вестник НГУ. – 2019. – № 17(2). – С. 30-38.

Повышение качества обработка видеоизображений на основе использования искусственной нейронной сети Yolo

В. Е. Дементьев
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
dve@ulntc.ru

А.Г. Ташлинский
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
tag54@mail.ru

М. Н. Суетин
Ульяновский Государственный
Технический Университет
Ульяновск, Россия
source81@yandex.ru

Аннотация—В работе рассмотрены вопросы обработки видеоизображений металлических конструкций, потенциально содержащих дефекты. Такую обработку предлагается выполнять на основе сверточных нейронных сетей. Для повышения качества обработки предлагаются варианты использования результатов обработки предшествующих кадров видеопоследовательности. Делается вывод о состоятельности и эффективности таких подходов.

Ключевые слова— Обнаружение аномалий, нейронные сети, обработка видеопотока, траекторное слежение.

1. ВВЕДЕНИЕ

В последнее время поиск объектов интереса на видеоизображениях находит широкое применение в различных областях и становится ключевой темой многих исследований [1-7]. В этих исследованиях обработка видеоизображений связана с решением таких задач как сопровождение и распознавание объектов интереса на последовательности кадров, которые основываются на детектировании объектов и предъявляют высокие требования к локализации объектов интереса на каждом кадре последовательности. Впечатляющие результаты по скорости обработки и качеству локализации объектов интереса демонстрируют подходы к детектированию связанные с применением нейросетевых архитектур (R-CNN, YOLO и т.д.) [1-4]. В работах [5-6] было показано, как на базе применения этих архитектур возможно получить приемлемое качество решения важной задачи обнаружения и оценки параметров дефектов на отдельных изображениях металлических конструкций. Привлекательной при этом выглядит идея совместной обработки нескольких изображений в видеопоследовательности для повышения качества детектирования и оценки параметров дефектов. Однако проведенные эксперименты показали, что перемещение регистрирующей камеры в ходе движения беспилотного аппарата накладывает свои ограничения на процесс обработки видеоизображений, связанные с искажением характеристик отдельных кадров, и ведет к потере качества детектирования. Искажение видеокadres приводит к пропаданию объекта интереса на очередном кадре и существенному ухудшению качества распознавания в целом.

2. ОПИСАНИЕ РЕШЕНИЯ

Для решения указанных проблем предлагается использовать информацию, полученную из предшествующих кадров видеопотока и содержащую результаты детектирования объектов интереса. Для этого выполняется последовательное совмещение кадров в последовательности на основе быстродействующих псевдоградиентных алгоритмов [7]. Оценки смещения могут быть использованы для прогнозирования области размещения объекта на последующих кадрах. Особенно удобным подобная процедура является в случае применения нейросетевых конструкций семейства Yola, поскольку выходной информацией сетей такой архитектуры являются прямоугольные окна, обрамляющие объекты интереса. Тогда, зная параметры деформаций, связывающих предшествующие и текущий кадр, возможно просто получить потенциальное место размещения дефекта на последующем кадре. Для этого достаточно выполнить расчет мест размещения четырех точек, являющихся углами выделенного нейронной сетью прямоугольника на предшествующих кадрах. В случае, если эти точки оказываются в пределах следующего кадра, соответствующий четырехугольник и будет местом потенциального размещения дефекта. Данный подход привлекателен в том числе и тем, что видеоизображения, получаемые с беспилотных платформ, обычно не содержит временных разрывов, и качество их совмещения обычно [5] очень хорошее.

Наличие оценок возможного размещения объекта на последующем кадре вкупе с дополнительной информацией о качестве выполненных псевдоградиентных релаксаций [7], позволяют варьировать коэффициент уверенности (confidence coefficient) нейронной сети в районе возможной локализации объекта. При этом в случае потери объекта на очередном кадре видеопоследовательности алгоритм повышает коэффициент уверенности нейронной сети при предсказании места объекта и тем самым стабилизирует результат работы детектора. Поиск необходимых зависимостей коэффициента уверенности от характеристик обрабатываемой видеопоследовательности и свойств детектируемого объекта также может быть выполнен на основе релейных псевдоградиентных процедур. Объект считается потерянным и исключается из алгоритма прогнозирования только в том случае, если он стабильно

не детектируется на заданном числе последовательных кадров.

3. ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

В ходе экспериментов, проведенных на тестовых видеоизображениях железнодорожных конструкций, была выполнена оценка стабильности детектирования областей внимания и трещин с помощью нейросетевой архитектуры YOLOv3 при различных коэффициентах уверенности нейросети в предсказаниях. В качестве гиперпараметров, влияющих на уверенность нейросети YOLOv3 в предсказаниях, были использованы пороговый коэффициент при отсеивании выявленных объектов интереса нейросетью (threshold) и пороговый коэффициент при отсеивании объектов интереса алгоритмом Non-Maximum Suppression (threshold NMS). При этом сравнивались результаты детектора, выполняющего независимую обработку кадров в видеопоследовательности (алгоритм 1), процедуры, предполагающей снижение коэффициентов threshold и threshold NMS в области прогнозируемого размещения дефекта на k процентов (алгоритм 2), процедуры, выбирающей коэффициенты threshold и threshold NMS в области прогнозируемого размещения дефекта в соответствии с предварительно проведенной процедурой псевдоградиентного поиска (алгоритм 3).



Рис. 1. Пример обработки кадра видеопоследовательности

В таблице 1 представлены зависимости средних вероятностей правильного обнаружения дефектов и вероятностей ложных тревог от вида используемого алгоритма.

Таблица 1. АНАЛИЗ РАБОТЫ АЛГОРИТМОВ

Алгоритм	Оценка вероятности правильного обнаружения	Оценка вероятности ложной тревоги
Алгоритм 1	0,56	0,01
Алгоритм 2, $k=10$	0,62	0,01
Алгоритм 2, $k=20$	0,66	0,02
Алгоритм 2, $k=40$	0,73	0,04
Алгоритм 2, $k=60$	0,79	0,13
Алгоритм 3	0,78	0,03

Предложенный в работе подход показал высокие результаты по стабилизации результатов детектирования видеоизображений. При этом заметно, что наилучшие результаты показывает процедура, предполагающая предварительный поиск зависимостей коэффициентов от свойств обрабатываемых изображений и типов детектируемых объектов. В то же время такая процедура сама по себе не является тривиальной и, по сути, также предполагает проведение некоторого обучения относительно свойств обрабатываемого материала. Поэтому на практике возможно ограничиться пропорциональным снижением соответствующих порогов уверенности, обеспечивая до 15% повышения вероятности верного обнаружения дефекта при незначительном увеличении вероятности ложной тревоги. Отметим особо, что рассматриваемые в работе процедуры сами по себе почти никак не сказываются на скорости обработки видеоматериала. Так на видеокarte Nvidia Quadro RTX 3000 обработка видеоизображения алгоритмами 1, 2 и 3 выполняется со скоростью примерно 30 кадров в секунду, что позволяет вести обработку видеопотока в режиме реального времени.

4. ЗАКЛЮЧЕНИЕ

Таким образом, в работе предложено семейство процедур, позволяющих поднять качество нейросетевых детекторов при обработке видеопоследовательностей. Выполнена оценка эффективности этих процедур на реальном материале, по результатам которой можно сделать вывод о повышении качества обработки без принципиального снижения производительности.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-09048.

ЛИТЕРАТУРА

- [1] Kalaivani, R. Object Detection in Video Frames Using Various Approaches / R. Kalaivani, C.R. Manicha // International Journal of Advanced Research in Computer and Communication Engineering. – 2013. – Vol. 2(9). – P. 157-160.
- [2] Redmon, J. You Only Look Once: unified, real-time object detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi [Electronical Resource]. – Access mode: <http://itnt-conf.org> (01.02.2022).
- [3] Padmanabula, S.S. Object Detection Using Stacked YOLOv3 / S.S. Padmanabula, V. Sistla, V.K.K. Kolli // International Information and Engineering Technology Association. – 2020. – Vol. 25(5). – P. 691-697.
- [4] Gai, W. An improved Tiny YOLOv3 for real-time object detection / W. Gai, Y. Liu, J. Zhang, G. Jing // System Science & Control Engineering: an open access journal. – 2021. – Vol. 9(1). – P. 314-321.
- [5] Dementiev, V.E. Detecting anomalies in temporal image sequences based on object identification by the stochastic gradient adaptation / V.E. Dementiev, R.G. Magdeev, A.G. Tashlinskii // IEEE Xplore. International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-5.
- [6] Dementiev, V. Improving The Quality of Detecting and Recognizing Defects in Images of Steel Structures / V. Dementiev, M. Suetin, M. Gaponova // IEEE Xplore. International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-5.
- [7] Ташлинский, А.Г. Оценивание параметров пространственных деформаций последовательностей / А.Г. Ташлинский. – Ульяновск : Издательство УлГТУ, 2000. – 139 с.

Исследование эффективности метода стохастической модуляции для стеганографического встраивания в тепловые видеоданные

А.Ю. Баврина

*Институт систем обработки
изображений - филиал ФНИЦ
«Кристаллография и фотоника»
РАН
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
bavrina@mail.ru*

Д.Д. Карнаухов

*Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
darzy1997@gmail.com*

В.А. Федосеев

*Самарский национальный
исследовательский университет им.
академика С.П. Королева
Институт систем обработки
изображений - филиал ФНИЦ
«Кристаллография и фотоника»
РАН
Самара, Россия
vicanfed@gmail.com*

Аннотация—Исследуется метод стеганографического встраивания информации в данные съемки тепловой камеры. После оценки особенностей шумовой составляющей тепловизионных видеоданных предлагается использование метода стохастической модуляции, заключающегося в замене шума камеры на шум с аналогичными параметрами, содержащий встраиваемую информацию. Исследуются такие показатели, как объем встраиваемой информации, соответствие встраиваемого шума исходному и стойкость к стегоанализу.

Ключевые слова— *стеганография, метод стохастической модуляции, тепловые видеоданные.*

1. ВВЕДЕНИЕ

Стеганография играет важную роль для безопасной передачи информации в современном цифровом мире. Значительную часть визуальной информации, передаваемой по открытым каналам связи, составляют данные оптических камер, полученные в видимом диапазоне света. Однако в последнее время активнее стали использоваться данные тепловых камер, имеющие свои особенности, поэтому задача разработки методов скрытой передачи информации с использованием тепловых видеоданных является актуальной.

Тепловые видеоданные используются во многих областях: медицина [1], промышленность (например, неразрушающий контроль материалов), системы видеонаблюдения [2], аутентификации, системы анализа при чрезвычайных ситуациях [3]. Встраивание в тепловые данные производится для проверки их подлинности, а также для передачи дополнительной информации (например, в медицине, данных о диагнозе пациента, параметрах области интереса, изображения этой области в видимом диапазоне).

Для обеспечения скрытности факта передачи, методы встраивания информации должны учитывать особенности контейнера. Для тепловых видеоданных характерны низкие значения отношения сигнал/шум и контрастности. Источниками искажений являются шумы в датчиках, особенности считывания заряда, условия окружающей среды, колебания температуры объекта, использование методов сжатия.

Исследованию качества тепловых видеосистем посвящен ряд работ. Некоторые исследователи [4] ограничиваются оценкой величины NETD (Noise Equivalent Temperature Difference - разница температур, эквивалентная шуму), другие же исследуют искажения более детально [5, 6]. На практике характер искажений конкретной камеры должен быть оценен по имеющимся видеоданным и для упрощения многие исследователи принимают аддитивную гауссову модель шумов на тепловых изображениях.

В настоящей работе также используется гауссова модель шумов на тепловых видеоданных, параметры которой оцениваются по отдельным кадрам. При стеганографическом встраивании предлагается использовать метод стохастической модуляции [7] для замены шума камеры на шум с аналогичными параметрами, содержащий встраиваемую информацию.

2. ПРИМЕНЕНИЕ МЕТОДА СТОХАСТИЧЕСКОЙ МОДУЛЯЦИИ ДЛЯ ВСТРАИВАНИЯ В ТЕПЛОВЫЕ ВИДЕОДАННЫХ

Стохастическая модуляция - это метод стеганографического встраивания, при котором в отдельные пиксели встраивается сообщение путем добавления к изображению шумового сигнала с вероятностным распределением, симметричным относительно нуля [7]. В работе исследовались два варианта метода: с использованием одной (SM - Stochastic Modulation) и двух (ISM - Improved Stochastic Modulation) шумовых последовательностей с нормальным законом распределения.

Для встраивания информации в тепловые видеоданные были выполнены следующие действия:

- Фильтрация кадров видеопоследовательности локальным фильтром (удаление шумовой составляющей).
- Вычисление шума как разницы между исходным кадром и результатом фильтрации.
- Оценка характеристик шума (мат. ожидания и дисперсии).

- Генерация шумовой последовательности с нормальным распределением и полученными на предыдущем этапе параметрами.
- Встраивание секретного сообщения методом стохастической модуляции в результат фильтрации с использованием сгенерированной шумовой последовательности.
- Оценка эффективности встраивания: ёмкости, соответствия стегошума исходному, способности быть обнаруженным с использованием широко применяющихся стегодетекторов.

Работоспособность и эффективность метода встраивания были исследованы с использованием натуральных тепловых видеопоследовательностей в формате без сжатия из тестового набора [8]. Для сравнения также были использованы простейшие методы встраивания, такие, как LSB (Least Significant Bit) и "±1 встраивание".

На рис. 1 представлены диаграммы ёмкости встраивания (количество бит встраиваемого сообщения на отсчет видеопоследовательности). Для исследуемых методов SM и ISM эти значения меньше единицы из-за ограничений метода и зависят от параметров шума.



Рис. 1. Ёмкость встраивания (бит на отсчет)

Рис. 2 показывает степень соответствия стегошума исходному (мера схожести гистограмм). В этой части экспериментов шумовые составляющие исходного видео и стеговидео были оценены межкадровым способом по статичному фрагменту.

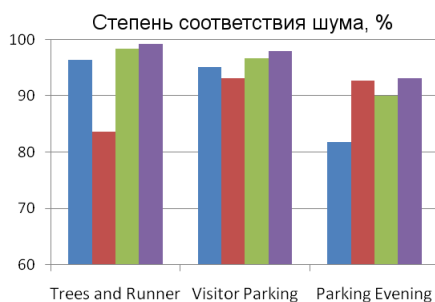


Рис. 2. Степень соответствия исходного шума и стегошума

Для исследования устойчивости метода встраивания были использованы стегодетекторы SPAM [9] (Subtractive Pixel Adjacency Matrix) и SP (Sample Pairs). SP ожидаемо показал детектируемость факта встраивания только для LSB. Признаки, вычисленные с помощью SPAM, использовались для обучения линейного классификатора SVM. Обучение производилось на базе оптических изображений BOSS [10] (так как авторы не располагали

полноценной базой тепловых видеоданных), тестирование – на базе BOSS и базе TERMO (составлена авторами из базы [8] и собственных тепловых видео). Результаты, представленные в Таблице 1, показывают, что встраивание в тепловые видеоданные не обнаруживается стегодетектором, настроенным на оптические данные.

Таблица 1. Точность обнаружения факта встраивания

обучение	тестирование		
	BOSS + SM	BOSS + LSB + "±1"	TERMO + SM
BOSS + SM	0,95	0,49	0,44
BOSS + LSB + "±1"	0,58	0,93	0,28

3. ЗАКЛЮЧЕНИЕ

В работе рассматривается перспектива применения метода стохастической модуляции для стеганографического встраивания в данные видеосъемки тепловых камер. К недостаткам метода можно отнести необходимость знания на принимающей стороне статистики распределения шумовой составляющей (вида распределения, его параметров, начального значения генератора). Дальнейшие исследования необходимо направить на составление полноценной базы тепловых видео и исследовании на ней эффективности и стегоустойчивости рассматриваемого метода встраивания.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке РФФИ (грант 19-29-09045 мк).

ЛИТЕРАТУРА

- [1] Velumani, R. A Semiblind Reversible Watermarking Scheme for Authentication and Fusion of Facial Thermograms / R. Velumani, V. Seenivasagam // Asian Journal of Scientific Research – 2013. – Vol. 6(1). – P. 78-87. DOI: 10.3923/ajsr.2013.78.87.
- [2] Lahmyed, R. A new thermal infrared and visible spectrum images-based pedestrian detection system / R. Lahmyed, M. El Ansari, A. Ellahyani // Multimedia Tools and Applications. – 2019. – Vol. 78. – P. 15861-15885. DOI: 10.1007/s11042-018-6974-5.
- [3] de Vries, J. Image processing and noise reduction techniques for thermographic images from large-scale industrial fires / J. de Vries // QIRT. – 2014. DOI: 10.21611/qirt.2014.042.
- [4] Redjimi, A. Noise Equivalent Temperature Difference Model for Thermal Imagers, Calculation and Analysis / A. Redjimi, D. Knezevic, K. Savic, N. Jovanovic, M. Simovic, D. Vasiljevic // Scientific Technical Review. – 2014. – Vol. 64(2). – P. 42-49.
- [5] Alparone, L. Noise modeling and estimation in image sequences from thermal infrared cameras / L. Alparone, G. Corsini, M. Diani // Proc. SPIE, Image and Signal Processing for Remote Sensing X. – 2004. – Vol. 5573. – P. 381-389. DOI: 10.1117/12.567998.
- [6] Евтихийев, Н.Н. Метод измерения шумов цифровых камер автоматической сегментацией голосовой сцены / Н.Н. Евтихийев, А.В. Козлов, В.В. Краснов, В.Г. Родин, Р.С. Стариков, П.А. Черёмхин // Компьютерная оптика. – 2021. – Т. 45, № 2. – С. 267-276. DOI: 10.18287/2412-6179-CO-815.
- [7] Fridrich, J. Digital image steganography using stochastic modulation / J. Fridrich, M. Goljan // Proceedings of the SPIE. – 2003. – Vol. 5020. – P. 191-202. DOI: 10.1117/12.479739.
- [8] INO's Video Analytics Dataset [Electronic resource]. – Access mode: <https://www.ino.ca/en/technologies/video-analytics-dataset>.
- [9] Feature Extractors for Steganalysis [Electronic resource]. – Access mode: http://dde.binghamton.edu/download/feature_extractors.
- [10] Bas, P. Break Our Steganographic System: the ins and outs of organizing BOSS / P. Bas, T. Filler, T. Pevný, S. Craver, A. Ker // Computer Science: Springer, Berlin, Heidelberg. – 2011. – Vol. 6958.

Метод защиты видео при помощи фазовых цифровых водяных знаков

Д.А. Шапиро
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
david-shapiro@mail.ru

В.В. Сергеев
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Институт систем обработки
изображений - филиал ФНИЦ
«Кристаллография и фотоника» РАН
Самара, Россия
vserg@geosamara.ru

В.А. Федосеев
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Институт систем обработки
изображений - филиал ФНИЦ
«Кристаллография и фотоника» РАН
Самара, Россия
vicanfed@gmail.com

Аннотация—В работе представлен новый метод встраивания стойких цифровых водяных знаков (ЦВЗ) в видеосигнал. Суть метода состоит в добавлении к каждому пикселю видеосигнала малозаметной синусоидальной последовательности, развивающейся во времени, и имеющей индивидуальную фазу. Двумерное поле этих фаз соответствует изображению встраиваемого ЦВЗ. Описаны простые алгоритмы встраивания и извлечения ЦВЗ из видео. Приведены результаты экспериментальных исследований, демонстрирующие высокое качество извлечения ЦВЗ при их визуальной неразличимости.

Ключевые слова— цифровой видеосигнал, цифровой водяной знак, фазовое встраивание, защита видео.

1. ВВЕДЕНИЕ

Быстрое развитие Интернета и беспроводной связи значительно упрощает доступ к таким продуктам, как изображения, видео и т. д. Однако такое развитие привело к незаконному использованию и распространению продуктов, защищенных авторским правом.

Одним из способов решения проблемы повышения защищенности видеoinформации является использование цифровых водяных знаков (ЦВЗ) [1-3].

В настоящей работе предлагается новый метод встраивания двумерных ЦВЗ в цифровой видеосигнал. Преимуществом метода является распределенный во времени характер ЦВЗ, обеспечивающий его незаметность и повышенную стойкость к различным искажениям видео. Суть метода состоит в добавлении к каждому пикселю видеосигнала малозаметной синусоидальной последовательности, развивающейся во времени, и имеющей индивидуальную (меняющуюся от пикселя к пикселю) фазу. Двумерное поле этих фаз соответствует изображению встраиваемого ЦВЗ. Из отдельного кадра видео такой ЦВЗ выделить невозможно, однако из любого достаточно продолжительного фрагмента видео он извлекается легко и с высоким качеством.

2. ОПИСАНИЕ ПРЕДЛАГАЕМОГО МЕТОДА

Для краткости изложения далее будем рассматривать одноканальный (черно-белый) видеосигнал. Пусть $x_0(m)$ – m -й кадр исходного видео, а $x_0(m, n_1, n_2)$ – значение пикселя с координатами (n_1, n_2) в m -м кадре. Метод встраивания фазового ЦВЗ состоит в добавлении к

исходному видео синусоидальной компоненты, начальная фаза которого определяется изображением ЦВЗ:

$$x(m, n_1, n_2) = x_0(m, n_1, n_2) + w(m, n_1, n_2), \quad (1)$$

$$w(m, n_1, n_2) = A \cdot \sin(\theta m + \varphi(n_1, n_2)), \quad (2)$$

где A – амплитуда, θ – опорная частота, $\varphi = \varphi(n_1, n_2)$ – фазовое изображение ЦВЗ.

Для извлечения из пикселя встроенной фазовой информации можно использовать любой узкополосный (настроенный на частоту θ) цифровой фильтр, например, линейную инвариантную к сдвигу систему [4] с комплекснозначной импульсной характеристикой (ИХ):

$$h(k) = \alpha^m e^{i\theta m} u(m), \quad (3)$$

где α – параметр затухания ИХ ($0 < \alpha < 1$), $u(m)$ – функция единичного скачка. Эта система преобразует входной вещественный сигнал $x(m)$ в выходной комплекснозначный:

$$y(m) = h(m) * x(m) = y_c(m) + iy_s(m), \quad (4)$$

где «*» – символ свертки. Обработка сигнала, т.е. выполнение пары преобразований для расчёта y_c и y_s реализуется при помощи разностных уравнений:

$$f(m) = 2\alpha \cos(\theta) f(m-1) - \alpha^2 f(m-2) + x(m), \quad (5)$$

$$y_c(m) = f(m) - \alpha * \cos(\theta) f(m-1), \quad (6)$$

$$y_s(m) = \alpha * \sin(\theta) f(m-1), \quad (7)$$

где $f(m)$ – промежуточная последовательность.

Вещественную и мнимую компоненты выходного сигнала, полученные в результате вычислений по формулам (5)-(7), можно интерпретировать как оценки синусной и косинусной составляющей колебания с частотой θ и фазой φ , аддитивно добавленного к сигналу в качестве ЦВЗ:

$$y_c(m) = B \cos(\theta m + \varphi), \quad (8)$$

$$y_s(m) = B \sin(\theta m + \varphi), \quad (9)$$

где B – амплитуда этих колебаний.

Далее решаем систему уравнений (8), (9) относительно фазовых сигналов:

$$C(m) = B \cos(\varphi) = \cos(\theta m) * y_c(m) + \sin(\theta m) * y_s(m), \quad (10)$$

$$S(m) = B \sin(\varphi) = \cos(\theta m) * y_s(m) - \sin(\theta m) * y_c(m). \quad (11)$$

И, наконец, выделяем фазу, в которой и будет содержаться необходимый ЦВЗ:

$$\varphi = \begin{cases} \arctg \frac{S(m)}{C(m)}, & C(m) \geq 0, S(m) \geq 0, \\ \arctg \frac{S(m)}{C(m)} + \pi, & C(m) < 0, \\ \arctg \frac{S(m)}{C(m)} + 2\pi, & S(m) < 0. \end{cases} \quad (12)$$

Вычисленная по (12) фаза синусоиды, встроенной в пиксель, в идеале не зависит от номера кадра m . Однако в реальных условиях она претерпевает случайные возмущения, связанные с динамикой основного видео. Поэтому полезно рассматривать эту фазу снова как функцию m и применить к ней некоторую процедуру сглаживания (усреднения) во времени. Также следует отметить необходимость учета периодического (с периодом 2π) характера фазы, которое может исказить результат усреднения.

3. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

В проведенных экспериментах использовались фрагменты тестовых однокомпонентных видеосигналов, обладающих яркостными и динамическими характеристиками, типичными для художественных видеофильмов. Пиксели видео имели байтовый формат (256 уровней яркости) Длина фрагментов составляла 3000 кадров (1 минута 40 секунд при частоте 30 кадров в секунду). В качестве ЦВЗ был взят условный QR-код, состоящий из 65×65 квадратных клеток, каждый размером в 16×16 пикселей. Клеткам QR-кода назначались случайные равновероятные значения 0 или 1, которые переносились во встраиваемые синусоиды с разностью фаз $\pi/2$.

При извлечении ЦВЗ полученные для выбранного кадра значения фаз усреднялись по пикселям каждой клетки, трансформировались обратно в бинарную форму и сравнивались с со значениями клеток исходного QR-кода. Большое число клеток QR-кода ($65 \times 65 = 4225$), позволяет с высокой точностью (0,024%) оценивать вероятность ошибок извлечения каждого элемента ЦВЗ.

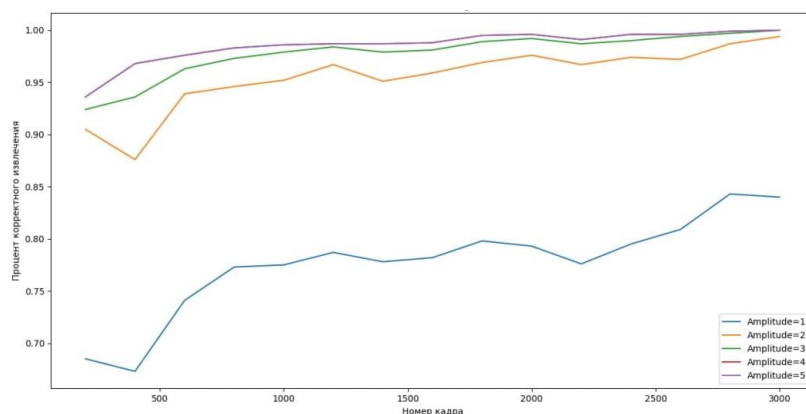


Рис. 1. График зависимости вероятности правильного извлечения элемента ЦВЗ от номера кадра

В ходе экспериментов варьировалась амплитуда встраиваемой синусоидальной компоненты (значение A) в диапазоне от 1 до 5, а попытка извлечения ЦВЗ осуществлялась каждые 100 кадров.

Пример полученных результатов для одного из тестовых видео приведён на рис. 1. Как свидетельствует рисунок, при амплитуде синусоиды $A > 1$ точность извлечения не опускается ниже 93%, что является приемлемым для практического использования. Подобная картина наблюдается и для других видеоизображений. Заметим, что малая амплитуда встраиваемых синусоид (на рисунке демонстрируются результаты для $A \leq 5$) заведомо обеспечивает визуальную неразличимость ЦВЗ.

4. ЗАКЛЮЧЕНИЕ

В работе предложен новый метод фазового встраивания стойких цифровых водяных знаков в видео. Результаты экспериментальных исследований демонстрируют работоспособность метода и высокое качество извлечения ЦВЗ при визуальной неразличимости защитной информации, вносимых в видеоизображения.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке РФФИ (проект 19-29-09045).

ЛИТЕРАТУРА

- [1] Zheng, J. A Novel Video Copyright Protection Scheme Based on Blockchain and Double Watermarking / J. Zheng, S. Teng, P. Li, W. Ou, D. Zhou, J. Ye // Security and Communication Networks. – 2021. – Vol. 2021. – P. 1-16. DOI: 10.1155/2021/6493306.
- [2] Yu, X. A Survey on Robust Video Watermarking Algorithms for Copyright Protection / X. Yu, C. Wang, X. Zhou // Applied Sciences. – 2018. – Vol. 8(10). – P. 1891. DOI: 10.3390/app8101891.
- [3] Egorova, A. QIM-Based Semi-Fragile Watermarking for H.264 Video Authentication / A. Egorova, V. Fedoseev // 8th International Symposium on Digital Forensics and Security (ISDFS). – 2020. – P. 1-6. DOI: 10.1109/ISDFS49300.2020.9116454.
- [4] Оппенгейм, А.В. Цифровая обработка сигналов / А.В. Оппенгейм, Р.В. Шафер. – М.: Связь, 1979. – 420 с.

Распознавание оптических мод Эрмита-Гаусса с использованием нейронной сети

В.В. Прудовская

Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
vvrnlk@gmail.com

Аннотация—В данной работе праведно исследование влияния аффинных преобразований на точность распознавания мод Эрмита-Гаусса. Показано, что наличие искажений снижает точность классификации более, чем на 10%.

Ключевые слова—нейронные сети, моды Эрмита-Гаусса, аффинные преобразования, машинное обучение.

1. ВВЕДЕНИЕ

Не так давно нейронные сети вызвали революцию в сфере исследований искусственного интеллекта, благодаря своим достижениям в таких областях, как интерпретация голоса в текст, а также распознавание образов на изображениях [1]. Благодаря своей гибкости нейронные сети смогли найти применение и в различных областях современной науки, в том числе физике и химии [2]. Одно из потенциальных приложений нейронных сетей – телекоммуникации [3]. Оптические коммуникации основаны на генерации, передаче и обнаружении состояний света для кодирования и передачи информации [4]. Для увеличения скорости передачи информации при помощи оптической связи были разработаны многочисленные протоколы [5]. Текущие препятствия для еще более высоких скоростей передачи битов носят чисто технический характер. Основной технической трудностью является точная классификация значений, обнаруженных на приемном конце коммуникационной платформы или на выходном слое установки. Традиционный метод распознавания требует сложного процесса оптического выравнивания и обеспечивает стабильно плохие результаты для сигналов, содержащих шум. Отличительной особенностью нейронных сетей, относительно традиционных методов распознавания изображений на выходном слое различных оптических установок, является возможность анализа изображения при большом количестве искажений.

Целью данной работы является исследовать возможность сверточной нейронной сети распознавать оптические моды Эрмита-Гаусса при наличии в наборе данных другого класса оптических мод – Лаггера-Гаусса, а также изучить влияние аффинных преобразований на точность распознавания.

2. ФОРМИРОВАНИЕ НАБОРОВ ДАННЫХ

Рассмотрим пучок Эрмита-Гаусса, комплексная амплитуда которого имеет вид [6]:

$$E_{n,m}(x, y, z) = i^{n+m} \left(\frac{w}{w(z)}\right)^2 H_n\left(\frac{\sqrt{2x}}{w(z)}\right) H_m\left(\frac{\sqrt{2y}}{w(z)}\right) \exp\left(-\frac{x+y^2}{w^2(z)} + \frac{ik(x+y^2)}{2R(z)}\right) \exp\left(-i(n+m+1)\arctg\left(\frac{z}{z_0}\right)\right) \quad (1)$$

где $w(z)$ – радиус гауссова пучка, $R(z)$ – радиус кривизны волнового фронта гауссова пучка, z_0 – длина Рэлея, k – волновое число света.

Для обучения нейронной сети были получены наборы изображений оптических мод - моды Эрмита-Гаусса в качестве основного класса, моды Лаггера-Гаусса – в качестве второго класса. Для формирования изображений была использована библиотека LightPipes языка Python. Полученные в результате работы программы изображения мод отличаются степенью и порядком. Также были определены неизменяемые параметры, такие как длина волны λ , размер области 15 мкм, размерность k и масштаб w . Отличие сгенерированных мод друг от друга заключается в их степени и порядке, именно поэтому другие параметры, не влияющие на итоговое изображение – постоянные величины. Результаты работы программы, генерирующей оптические моды, представлены на Рис. 1.

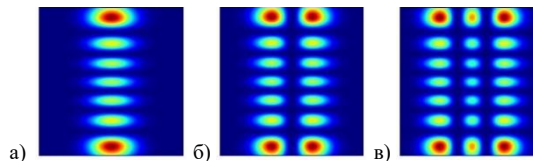


Рис. 1. Изображения мод Эрмита-Гаусса при заданных входных параметрах. а) Степень $l = 0$; б) Степень $l = 1$; в) Степень $l = 2$

Путем варьирования приведенных выше параметров были получены следующие наборы данных: четыре тысячи изображений мод Лаггера-Гаусса и столько же изображений мод Эрмита-Гаусса. Для оценки влияния аффинных преобразований на точность распознавания нейронной сети была реализована программа, осуществляющая преобразования. Программа была написана на языке Python с использованием библиотеки Keras. В результате работы программы получен набор данных из сдвинутых и повернутых под определенным углом изображений мод. Примеры изображений мод после применения аффинных преобразований представлены на Рис. 2.

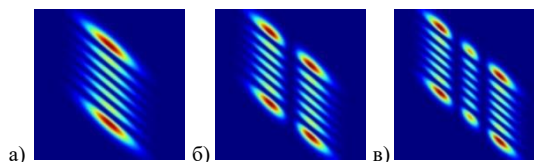


Рис. 2. Изображения мод Эрмита-Гаусса после применения к ним аффинных преобразований: а) Степень $l = 0$; б) Степень $l = 1$; в) Степень $l = 2$

3. МОДЕЛИРОВАНИЕ

Для однозначной классификации изображений мод Эрмита-Гаусса на языке Python была написана сверточная нейронная сеть. За основу обучающего набора была взята коллекция оптических мод Эрмита-Гаусса, отличающихся порядком и степенью. Нейронная сеть должна определять принадлежность поступившего на вход изображения к определенному классу.

Сверточная нейронная сеть основана на использовании последовательной модели Sequential. В структуре участвуют сверточные слои Conv2D с функцией активации ReLU, вспомогательные слои MaxPooling2D и Dropout, а также полносвязные слои Dense с функциями активации ReLU и Sigmoid. Вместе они формируют структуру сверточной нейронной сети, представленную на Рис. 3.

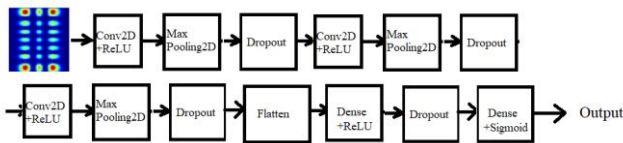


Рис. 3. Архитектура разработанной сверточной нейронной сети

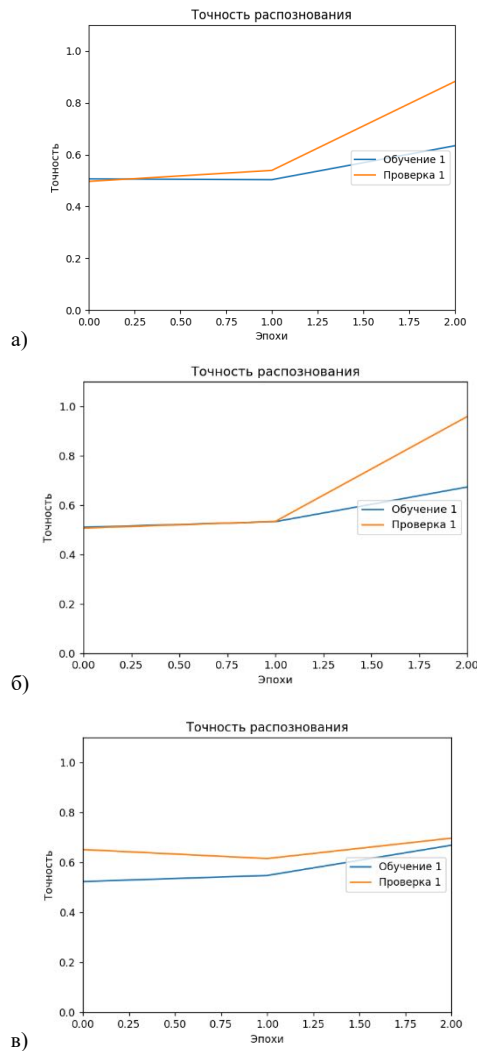


Рис. 4. Графики точности распознавания мод Эрмита-Гаусса на наборе изображений: а) без преобразований; б) с преобразованиями; в) из комбинации наборов с преобразованиями и без

Обучение и тестирование нейронной сети для оценки влияния аффинных преобразований на точность распознавания проводилось на разных наборах данных объемом 10 000 изображений. Было сформировано три набора данных: первый содержит только моды, сгенерированные с помощью библиотеки LightPipes, второй только изображения, подвергнутые преобразованиям, а третий набор является комбинацией. На Рис. 4 представлены графики точности распознавания, которые получились в результате проведенного исследования.

4. ЗАКЛЮЧЕНИЕ

В результате данной работы были сформированы наборы данных, содержащих изображения мод Эрмита-Гаусса и Лаггера-Гаусса, а также эти же изображения, подвергнутые аффинным преобразованиям. Была исследована зависимость точности распознавания смоделированной нейронной сетью от входных данных. Было установлено, что точность классификации мод Эрмита-Гаусса имеет высокий показатель, порядка 95% при распознавании изображений без аффинных преобразований. Также достаточно высокая точность достигается и при использовании смешанного набора данных – 88%. Точность классификации достигает меньшего значения в случае, когда все входные изображения подвергаются аффинным преобразованиям, однако она все равно составляет 69%. Данное исследование может применяться для дальнейших работ, направленных на увеличение пропускной способности оптических систем связи.

ЛИТЕРАТУРА

- [1] Harizi, R. Convolutional neural network with joint stepwise character/word modeling based system for scene text recognition / R. Harizi, R. Walha, F. Drira // Multimedia Tools and Applications. – 2021. – Vol. 80(7). – P. 34-45.
- [2] Ye, S. Symmetrical Graph Neural Network for Quantum Chemistry with Dual Real and Momenta Space / S. Ye, J. Liang, R. Liu, X. Zhu // The Journal of Physical Chemistry A. – 2020. – Vol. 124(34). – P. 6945-6953.
- [3] Yuan, X. Dual-Output Mode Analysis of Multimode Laguerre-Gaussian Beams via Deep Learning / X. Yuan, Y. Xu, R. Zhao, X. Hong, R. Lu, X. Feng, Y. Chen, J. Zou, C. Zhang, Y. Qin, Y. Zhu // Optics. – 2021. – Vol. 2(2). – P. 87-95.
- [4] Anbarasi, K. A review on channel models in free space optical communication system / K. Anbarasi, C. hermanth, R. Sangeetha // Optics & Laser Technology. – 2017. – Vol. 97(1). – P. 161-167.
- [5] Satrusalya, S. Review on free space optical communication / S. Satrusalya, L. Goswami // Materialstoday: proceedings. – 2021. – Vol. 45(6). – P. 1-25.
- [6] Котляр, В.В. Лазерные пучки Эрмита-Гаусса с орбитальным угловым моментом / В.В. Котляр, А.А. Ковалёв, А.П. Порфирьев // Компьютерная оптика. – 2014. – Т. 38, № 4. – С. 651-657. DOI: 10.18287/0134-2452-2014-38-4-651-657.

Распознавание оптических мод Лаггера-Гаусса с использованием сверточной нейронной сети

А.В. Бехтерев

Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
bekhterev98@mail.ru

Аннотация—В работе проведено исследование влияния аффинных преобразований на точность распознавания мод Лаггера-Гаусса сверточной нейронной сетью. Показано, что наличие искажений снижает точность классификации на восемь процентов.

Ключевые слова—нейронные сети, моды Лаггера-Гаусса, аффинные преобразования, машинное обучение.

1. ВВЕДЕНИЕ

Наука о данных сегодня претерпевает бурное развитие [1]. Причина заключается в росте объёмов связанной информации, которую необходимо быстро обработать в автоматическом режиме. Создание методов, с помощью которых вычислительные системы выполняют анализ таких данных, является объектом исследования машинного обучения [2]. В ходе развития теории в этой области науки появился новый инструмент с широкими возможностями в различных предметных областях - нейронная сеть [3]. Наиболее популярный класс задач, решаемых с его применением – анализ и классификация различного рода данных. В частности, для анализа и классификации изображений и видео используются сверточные нейронные сети [4].

Благодаря своей гибкости нейронные сети смогли найти применение в различных областях современной науки и техники [5]. В частности, в наши дни количество информации, передаваемой между цифровыми устройствами, достигает невероятного объема. Пропускная способность имеет ограничения, связанные с волоконно-оптическими нелинейными эффектами в оптических системах связи [6]. Одним из способов решения данной проблемы являются моды Лаггера-Гаусса [7], но распознавание поступившего сигнала зачастую затрудняется из-за помех и искажений. Целью данной работы является исследование возможности сверточной нейронной сети распознавать оптические моды Лаггера-Гаусса при наличии в наборе данных другого класса оптических мод – Эрмита-Гаусса, а также изучение влияния аффинных преобразований, как одного из способов искажения сигнала, на точность распознавания.

2. ФОРМИРОВАНИЕ БАЗЫ ДАННЫХ

Моды Лаггера-Гаусса – это решения уравнений Максвелла в свободном пространстве в параксиальном приближении, в цилиндрических координатах [8]. Нейронная сеть будет обучаться и тестироваться на изображениях мод. Для их формирования будет использоваться библиотека LightPipes высокоуровневого языка программирования Python. В результате работы программы будут формироваться изображения для мод Лаггера-Гаусса и Эрмита-Гаусса. Они будут отличаться

варьированием степени l и порядка p . Также определяются неизменяемые параметры, такие как длина волны, размер поля изображения, размерности и масштаб. Длина волны на протяжении всего эксперимента будет неизменной, так как ее значение не влияет на итоговое изображение интенсивности. В ходе исследования будут меняться в основном два параметра: степень и порядок. Именно они дают разительное отличие изображений мод друг от друга. Результаты генерации изображений оптических мод представлены на Рис. 1.

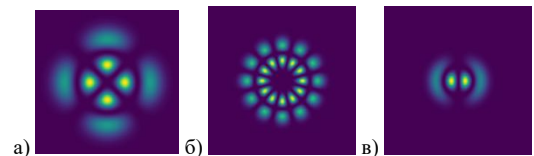


Рис. 1. Изображения мод Лаггера-Гаусса при заданных входных параметрах: а) $p=2, l=1$; б) $p=2, l=1$, в) $p=2, l=2$

Путем варьирования параметров были получены два набора данных с модами Лаггера- и Эрмита-Гаусса размером по четыре тысячи изображений каждый. Следующим этапом требовалось провести аффинные преобразования, которые производились также с использованием программных средств языка Python, а именно фреймворка TensorFlow. В результате были сформированы наборы из пяти тысяч изображений, которые получались в результате координатного сдвига под заданным углом. Примеры изображений мод после применения аффинных преобразований представлены на Рис. 2.

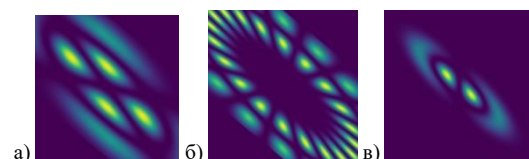


Рис. 2. Изображения мод Лаггера-Гаусса после применения аффинных преобразований

3. МОДЕЛИРОВАНИЕ

Для задачи распознавания мод Лаггера-Гаусса на языке Python с помощью фреймворка TensorFlow была создана сверточная нейронная сеть, которая принимает на вход изображения двух классов оптических мод, а на выходе определяет их вероятность принадлежности к определенному классу. В данном случае возможными классами являются моды Лаггера- и Эрмита-Гаусса.

Сверточная нейронная сеть основана на использовании последовательной модели Sequential. В структуре участвуют сверточные слои Conv2D с функцией активации ReLU, вспомогательные слои MaxPooling2D и

Dropout, а также полносвязные слои Dense с функциями активации ReLU и Sigmoid. Вместе они формируют структуру сверточной нейронной сети, представленную на Рис. 3.

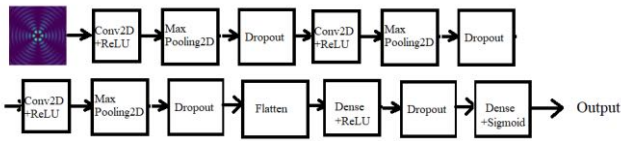


Рис. 3. Архитектура разработанной сверточной нейронной сети

Обучение и тестирование нейронной сети проводилось на различных данных. Для того, чтобы исследовать влияние аффинных преобразований на точность классификации мод, были сформированы по два набора размерностью по пять тысяч на каждый тип изображений: без преобразований, с преобразованиями и их комбинация. При одном запуске нейронной сети, обучение и тестирование производилось на двух наборах одного типа размерностью по десять тысяч, но содержащих различные изображения. На Рис. 3 представлены графики точности распознавания, которые получились в результате проведенного исследования.

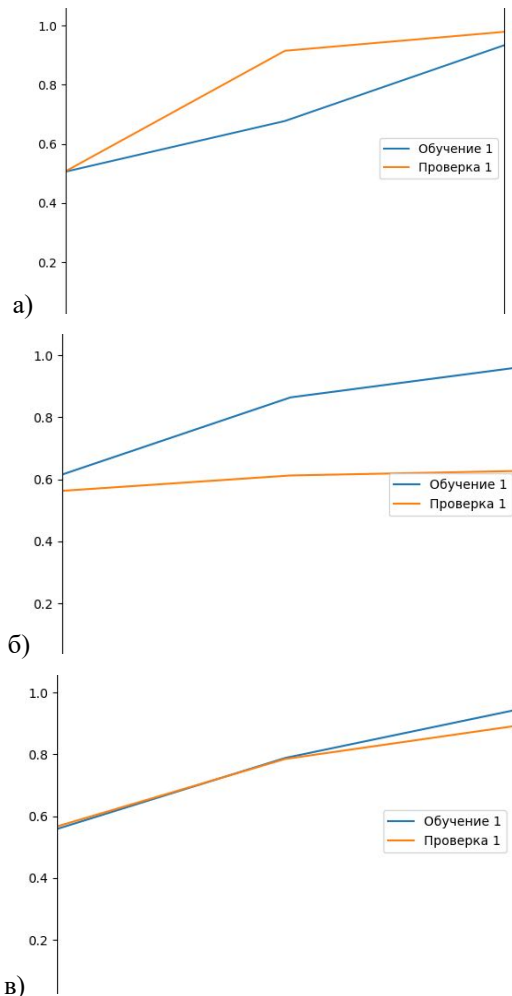


Рис. 4. Графики точности распознавания мод Лаггера-Гаусса на наборе изображений: а) без преобразований; б) с преобразованиями; в) из комбинации наборов с преобразованиями и без

Из графиков видно, что при наборе данных с изображениями мод Лаггера-Гаусса без использования аффинных преобразований точность распознавания сверточной нейронной сети достигает порядка 97,91%. При обучении и тестировании на наборе, содержащем только преобразованные моды точность значительно падает до 62,69%. На последнем наборе данных, состоящем из комбинации изображений точность, по сравнению с первым запуском, незначительно снижается до 89,1%.

4. ЗАКЛЮЧЕНИЕ

В результате данной работы были сформированы наборы данных с изображениями мод Лаггера- и Эрмита-Гаусса, которые также были подвергнуты применению аффинных преобразований. С помощью полученных наборов, а также смоделированной сверточной нейронной сети, была исследована зависимость точности распознавания от входных данных. Было установлено, что точность классификации мод Лаггера-Гаусса имеет высокий показатель, порядка 97%, при распознавании мод без аффинных преобразований. Достаточно высокая точность достигается и при использовании смешанного набора данных – 89%. Классификация сильно ухудшается в случае, если все входные изображения подвергаются аффинным преобразованиям, и составляет 62%. Данное исследование может применяться для дальнейших работ, направленных на увеличение пропускной способности оптических систем связи.

ЛИТЕРАТУРА

- [1] Sarker, H. Mobile Data Science and Intelligent Apps: Concepts, AI-Based Modeling and Research Directions / H. Sarker, M. Moshuiul, K. Uddin, T. Alsanoosy // *Mobile Networks and Applications*. – 2021. – Vol. 26(1). – P. 1-19.
- [2] Zhang, X. A Matrix Algebra Approach to Artificial Intelligence / X. Zhang // *Machine Learning*. – 2020. – Vol. 15(13). – P. 223-440.
- [3] Abiodun, I. State-of-the-art in artificial neural network applications: A survey / I. Abiodun, A. Jantan, E. Omolara, V. Dada, A. Mohamed, A. Humaira // *Heliyon*. – 2018. – Vol. 4(11). – P. 1-41.
- [4] Naranjo, J. A Review of Convolutional Neural Network Applied to Fruit Image Processing / J. Naranjo, M. Mora, R. Hernandez, R. Barrientos, C. Fredes, A. Valenzuela // *Applied Sciences*. – 2020. – Vol. 10(10). – P. 1-31.
- [5] Ye, S. Symmetrical Graph Neural Network for Quantum Chemistry with Dual Real and Momenta Space / S. Ye, J. Liang, R. Liu, X. Zhu // *The Journal of Physical Chemistry A*. – 2020. – Vol. 124(34). – P. 6945-6953.
- [6] Bozinovic, N. Terabit-Scale Orbital Angular Momentum Mode Division Multiplexing in Fibers / N. Bozinovic, Y. Yue, Yu. Ren, M. Tur, P. Kristensen, H. Huang, A.E. Willner, S. Ramachandran // *Science*. – 2013. – Vol. 340. – P. 1545-548.
- [7] Berry, M. Singularities in waves and rays / M. Berry // *Physics of defects. Les Houches Session XXXV*. – Amsterdam: North-Holland, 1980. – P. 453-543.
- [8] Ustinov, A.V. Local characteristics of paraxial Laguerre-Gaussian vortex beams with a zero total angular momentum / A.V. Ustinov, V.G. Niziev, S.N. Khonina, S.V. Karpeev // *Journal of Modern Optics*. – 2019. – Vol. 66(20). – P. 1961-1972.

Двухуровневый метод повышения качества сегментации изображений

Д.М. Мурашов

Федеральный исследовательский центр «Информатика и управление» Российской академии наук
Москва, Россия
d_murashov@mail.ru

Аннотация—Предложен новый двухуровневый метод комбинирования карт сегментации изображений на основе критерия минимума информационной избыточности. Метод позволяет улучшить результат сегментации. При этом комбинированные разбиения показали информационное отличие от эталонных разбиений, соизмеримое с отличием, порожденным результатами традиционных алгоритмов сегментации.

Ключевые слова— сегментация изображений, комбинирование карт сегментации, мера избыточности информации, энтропия, вариация информации.

1. ВВЕДЕНИЕ

Представляемая работа посвящена решению одной из задач, связанных с обеспечением качества сегментации цифровых изображений. Ряд работ по сегментации изображений посвящен подходу, который связан не с поиском на ансамбле разбиений изображения такого разбиения, которое обеспечивало бы оптимальное значение выбранной меры качества, а основан на комбинировании карт сегментации для получения нового разбиения, позволяющего улучшить значение меры качества. Задача комбинирования кластеров разбиений данных достаточно давно рассматривается в публикациях. Например, в работе [1] предложен метод построения консенсусного разбиения из ансамбля разбиений, полученных при кластеризации набора данных. Новое разбиение формируется при оптимизации величины средней нормализованной взаимной информации между консенсусным разбиением и анализируемым ансамблем разбиений. В работе [2] предложено использовать критерий на основе матрицы коассоциации (co-association matrix) элементов (образов) кластеров. Подход, связанный с комбинированием разбиений данных, успешно применяется при решении задачи сегментации изображений. В статье [3] представлен итерационный метод комбинирования множества грубо сегментированных изображений (карт сегментации или разбиений), полученных в разных цветовых пространствах при различных значениях параметров алгоритма сегментации. Для обеспечения качества используется критерий минимума средней вариации информации (mean variation of information criterion) между комбинированным изображением и каждой из грубых карт сегментации. Ряд статей посвящен применению различных показателей качества сегментации изображений [4] к задаче комбинирования карт сегментации.

В представляемой работе предлагается еще один метод комбинирования карт сегментации на стадии постобработки. Метод основан на критерии минимума информационной избыточности. Ранее этот критерий успешно применялся в задаче обеспечения качества

сегментации с помощью процедуры выбора из множества доступных разбиений входного изображения разбиения, минимизирующего информационную избыточность [5]. Была показана эффективность такого критерия.

2. ДВУХУРОВНЕВЫЙ МЕТОД КОМБИНИРОВАНИЯ КАРТ СЕГМЕНТАЦИИ

Комбинирование карт сегментации на стадии постобработки выполняется на двух уровнях. На первом уровне осуществляется комбинирование сегментов. На втором производится коррекция границ сегментов комбинированного изображения на пиксельном уровне. Двухуровневое комбинирование позволит получить более точное разбиение оригинального изображения и сохранить информационно важные области, которые могут быть утрачены при работе традиционных алгоритмов сегментации.

Задача формулируется следующим образом. Пусть из оригинального изображения U получено множество карт сегментации $V = \{V^1, V^2, \dots, V^q, \dots, V^Q\}$, где $V^q = \{S_1^q, S_2^q, \dots, S_{k_q}^q\}$ состоит из k_q сегментов. Пусть изображение $V^{q_{\min}}$ обеспечивает минимум меры информационной избыточности $R(U, V^q)$ [5]. Требуется сформировать разбиение V^* в виде комбинации Q разбиений V^q из V такое, что $R^* = R(U, V^*) < R(U, V^{q_{\min}})$.

А. Комбинирование сегментов

Процедура комбинирования на этом этапе реализуется следующим образом. В качестве начального приближения изображения V_0^* из множества карт сегментации V выбирается изображение $V^{q_{\min}}$. Далее выполняется последовательное сравнение сегментов из V^* и одного из изображений V^q , $q \neq q_{\min}$. Если какой-либо сегмент S_i^q из V^q отсутствует на V_0^* , то он копируется в V_0^* , и формируется новое изображение V_1^* . Если выполняется условие $R(U, V_{j+1}^*) < R(U, V_j^*)$, то полученное разбиение V_1^* сохраняется, и проверяется следующий сегмент, который в комбинации с изображением V_1^* порождает новое разбиение V_2^* . Если условие убывания избыточности информации не выполняется, то с использованием следующего сегмента формируется новое изображение V_2^* , и так далее. После проверки всех сегментов изображения V^q , выбирается следующее разбиение из множества V , и процедура повторяется.

Б. Коррекция границ сегментов

На втором уровне уточняются границы комбинированного изображения V^* . В качестве базового алгоритма использована итерационная процедура, предложенная в работе [3]. В предлагаемом алгоритме, в отличие от оригинала, вместо критерия минимума средней вариации информации применяется мера избыточности информации, содержащейся в сегментированном изображении. Процедура состоит в следующем. Последовательно пикселям, которые находятся на границе сегмента, присваивается метка соседнего сегмента. Вычисляется приращение меры избыточности. Если приращение отрицательно, то сохраняется новая метка сегмента. Если приращение положительно, метка восстанавливается. Далее меняется метка у следующего пикселя на границе сегмента, и описанные выше операции повторяются. Получены формулы для вычисления приращения меры избыточности.

$$\Delta R = R(U, V^{**}) - R(U, V^*) = \frac{H(U, V^{**}) - H(V^{**})}{H(V^{**})} - \frac{H(U, V^*) - H(V^*)}{H(V^*)}, \quad (1)$$

где $R(U, V^{**})$ и $R(U, V^*)$ - величины информационной избыточности, вычисленные для карт сегментации V^{**} и V^* ; V^{**} получена из V^* модификацией метки сегмента пикселя на границе сегментов; $H(U, V^{**})$, $H(V^{**})$, $H(U, V^*)$ и $H(V^*)$ - оценки совместных и маргинальных энтропий, вычисленные для карт сегментации V^{**} и V^* , соответственно:

$$H(U, V^{**}) = H(U, V^*) + \frac{n_k}{N} \log \frac{n_k}{n_k - 1} + \frac{n_m}{N} \log \frac{n_m}{n_m + 1} + \frac{1}{N} \log \frac{n_k - 1}{n_m + 1}, \quad (2)$$

$$H(V^{**}) = H(V^*) + \frac{n_k}{N} \log \frac{n_k}{n_k - 1} + \frac{n_m}{N} \log \frac{n_m}{n_m + 1} + \frac{1}{N} \log \frac{n_k - 1}{n_m + 1}, \quad (3)$$

где N - количество пикселей в изображениях U и V^* , l - значение уровня яркости пикселя изображения U с координатами, соответствующими координатам модифицируемого пикселя в V^* , k и m - метки сегментов разбиений V^{**} и V^* , которым принадлежал пиксель до и после модификации, n_k , n_m , n_k и n_m - количество пикселей в двумерных и одномерных гистограммах изображений U , V^{**} и V^* с уровнем яркости l и метками сегментов k и m , соответственно. Уравнения (1-3) позволяют быстро вычислить приращение избыточности ΔR при модификациях меток сегментов без полного сканирования изображений для построения гистограмм при вычислении энтропий.

В. Вычислительный эксперимент

Проведен вычислительный эксперимент на изображениях из базы BSDS500 [4]. Значения информационной избыточности, вычисленной на разных

стадиях комбинирования, представлены в Таблице 1, а значения нормированной вариации информации $VI(U, V)$ характеризующей информационное различие комбинированных разбиений и оригинала, - в Таблице 2.

Таблица I. ЗНАЧЕНИЯ ИНФОРМАЦИОННОЙ ИЗБЫТОЧНОСТИ НА РАЗНЫХ ЭТАПАХ КОМБИНИРОВАНИЯ РАЗБИЕНИЙ

Изображение	$R(U, V^{qmin})$	$R(U, V^*)$	$R(U, V^{**})$
35010.jpg	0,6784	0,6762	0,6476
124084.jpg	0,6623	0,6578	0,6223
42049.jpg	0,4660	0,4558	0,4335
67079.jpg	0,4810	0,4809	0,4477

Таблица II. ЗНАЧЕНИЯ ИНФОРМАЦИОННОГО РАЗЛИЧИЯ НА РАЗНЫХ ЭТАПАХ КОМБИНИРОВАНИЯ РАЗБИЕНИЙ

Изображение	$VI(U, V^{qmin})$	$VI(U, V^*)$	$VI(U, V^{**})$
35010.jpg	0,8846	0,8823	0,8758
124084.jpg	0,9182	0,9138	0,9133
42049.jpg	0,8805	0,8745	0,8725
67079.jpg	0,8886	0,8885	0,8867

3. ЗАКЛЮЧЕНИЕ

Предложен новый двухуровневый метод комбинирования карт сегментации изображений на основе критерия минимума информационной избыточности. Метод позволяет получить более точное разбиение оригинального изображения и сохранить информационно важные области, которые могут быть утрачены при работе традиционных алгоритмов сегментации. Результаты эксперимента показали, что предложенный метод позволяет улучшить результат сегментации изображений с точки зрения минимизации информационной избыточности и информационного различия с оригиналом. Комбинированные разбиения незначительно меняют величину информационного различия с эталонными сегментациями относительно различия разбиений, полученных традиционным алгоритмом.

ЛИТЕРАТУРА

- [1] Strehl, A. Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions / A. Strehl, J. Ghosh // Journal of Machine Learning Research. - 2002. - Vol. 3. - P. 583-617.
- [2] Fred, A. Robust data clustering / A. Fred, A.K. Jain // Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Madison, USA. - 2003. - Vol. 2. - P. 128-133.
- [3] Mignotte, M. A Label Field Fusion Model With a Variation of Information Estimator for Image Segmentation / M. Mignotte // Information Fusion. - 2014. - Vol. 20. - P. 7-20.
- [4] Martin, D. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics / D. Martin, C. Fowlkes, D. Tal, J. Malik // Proc. 8th Int. Conf. Comput. Vis. (ICCV), Vancouver, Canada. - 2001. - Vol. 2. - P. 416-423.
- [5] Murashov, D.M. An Information Model for Digital Image Segmentation / D.M. Murashov // Pattern Recognit. Image Anal. - 2021. - Vol. 31. - P. 632-645.

Алгоритмы сокрытия и извлечения информации в аудио-файлах

П.О. Волков
НИУ Высшая школа экономики
Москва, Россия
povolkov@miem.hse.ru

Аннотация—В работе рассмотрено два вида аудио-файлов, а именно: WAV, который является несжатым звуковым файлом, и MP3, полученный в результате сжатия аудиоинформации. В соответствии с их свойствами были разработаны алгоритмы сокрытия и извлечения информации.

Ключевые слова— цифровая стеганография, аудио-файлы, WAV-файлы, MP3-файлы.

1. ВВЕДЕНИЕ

Задача защиты информации от несанкционированного доступа решалась во все времена. С давних времен выделяется два пути решения данной задачи: криптография [1] и стеганография [2]. Целью криптографии является скрытие содержимого сообщений за счет их шифрования. В отличие от этого, при стеганографии (с греческого переводится как «тайнопись») скрывается сам факт существования тайного сообщения. Другими словами, скрываемое сообщение встраивается в некоторый не привлекающий внимание объект, а далее данный объект открыто отправляется адресату. Области применения стеганографии различны:

- защита от копирования;
- аутентификация;
- скрытая аннотация документов;
- скрытая связь.

Бурное развитие вычислительной техники и каналов передачи данных инициировало развитие нового направления сокрытия информации- цифровой стеганографии. В основе методов этого направления лежат особенности цифрового представления данных. Таким образом, цифровая стеганография – это наука о незаметном и надежном (или устойчивом к различного рода искажениям) скрытии одних битовых последовательностей в других, имеющих аналоговую природу.

Скрываемая информация внедряется в соответствии с ключом в те отсчеты, искажение которых не приводит к существенным искажениям контейнера. Эти биты образуют стегопуть.

В работе рассмотрен вопрос сокрытия информации в аудио-файлах. Данная проблема выбрана по той причине, что большинство контейнеров выбирают в качестве изображения. В данной работе рассмотрено 2 вида аудио-файла: WAV, который является несжатым звуковым файлом, и MP3, полученный в результате сжатия аудиоинформации. В соответствии с их свойствами был

разработан алгоритм сокрытия и извлечения информации.

2. СТРУКТУРА WAV-ФАЙЛА

Файлы WAV (или WAVE) были созданы компанией IBM и Microsoft [3]. Они содержат различные аудиоданные - звуки, звуковые эффекты, музыку, а также записи голоса. Файлы WAV значительно больше файлов MP3, поэтому они не пользуются популярностью. Отличительная черта файлов WAV - их технология сжатия звука без потерь. Полученные файлы отличаются высоким качеством. Отправка и загрузка таких файлов требует много времени и дискового пространства.

Отметим, что WAV-файлы используют универсальный RIFF формат (Resource Interchange File Format - формат файла для обмена ресурсами).

WAV-файл представляет собой две четко делящиеся области: заголовок файла и область данных. В заголовке файла хранится информация о:

- размере файла;
- количестве каналов;
- частоте дискретизации;
- количестве бит в сэмпле (глубине звучания), и т.д.

3. СТРУКТУРА MP3-ФАЙЛА

MP3 (или MPEG Layer3)- это один из потоковых форматов хранения и передачи аудио-сигнала в цифровой форме [4], [5]. В формате MP3 применяется алгоритм сжатия с потерями, разработанный для существенного уменьшения размера данных, необходимых для воспроизведения записи и обеспечения качества воспроизведения звука, близкого к оригинальному. В отличие от MP3-файла, WAV-файлу присуще несжатый объем данных.

MP3-файл, также как и WAV-файл, состоит из заголовка и области данных. Но заголовок MP3 не представляет такую важность, как заголовок WAV, вместо этого есть заголовок у каждого фрейма.

Фрейм- это условная часть аудио-данных. MP3-файл представляет последовательность фреймов, каждый из которых состоит из заголовка, служебной информации и данных. В 4-байтовом заголовке фрейма хранится информация о:

- частоте семплирования;
- битрейте;
- режиме стерео, и т.д.

4. АЛГОРИТМ СОКРЫТИЯ ИНФОРМАЦИИ В WAV-ФАЙЛАХ

Как было сказано ранее, область данных WAV-файла состоит из отсчетов. Изменив младший бит отсчета или даже удалив, человек не почувствует разницу между аудио-файлом с искаженным младшим битом и оригинальным аудио-файлом. В этом и состоит свойство каждого отсчета файла. Применив данное свойство, можно будет незаметно и хитро скрыть информацию.

Алгоритм сокрытия информации в WAV-файлах представлен в следующих шагах:

1. Скрываемое сообщение посимвольно разделить.
2. Каждый символ, характеризующийся числовым значением в диапазоне 1-255, переводится в биты с помощью стандартного преобразования из десятичного числа в двоичное.
3. Каждый бит встроить в младший бит отсчета.

Этот алгоритм характеризуется следующими свойствами:

- незаметностью при прослушивании;
- сложностью обнаружения стеганалитическими методами;
- высокой сложностью реализации и медленной работой.

Чтобы алгоритм сделать более надежным, можно биты скрываемого сообщения встраивать не в каждый младший бит отсчета аудио-сигнала, а в бит такого отсчета, номер которого подчиняется некоторому закону.

Извлечение информации происходит в обратном порядке. При этом учитывается количество символов скрываемого сообщения.

5. АЛГОРИТМ СОКРЫТИЯ ИНФОРМАЦИИ В MP3-ФАЙЛАХ

В силу того, что данные MP3-файла представляют сжатый вид (свободного места нет), алгоритм сокрытия информации в данном формате является затруднительной процедурой. Однако способы имеются: широкополосное кодирование, фазовое кодирование, замена шума [6], [7]. Общими характеристиками данных методов являются возможная заметность при прослушивании, пропускная способность порядка 40 бит в секунду и средняя скорость работы/сложность реализации. Так как методы вносят неестественные искажения в сигнал (белый шум определенной амплитуды, дискретные фазы сигнала и т.д.), вероятно создание надежных алгоритмов их обнаружения. Кроме того, указанные методы были разработаны применительно к процедурам обычного сжатия информации и не учитывали особенностей MP3 преобразований. Неопытные программисты кодируют сообщение в служебной информации MP3. Метод сокрытия информации, речь о котором пойдет далее, превосходит предыдущие способы сокрытия.

Как отмечалось, MP3-файл представляет последовательность фреймов. В заголовке файла содержится информация о скорости считывания данных (битрейт). Изменяя скорость считывания, можно увеличить размер области для аудио данных сверх

области, занимаемой полезной информацией контейнера (например, музыкой). Величина этой области данных файла-контейнера и ее смещение относительно заголовка определяется служебной информацией (СИ). Поскольку СИ не изменяется в процессе встраивания скрываемых данных, то эти данные не будут воспроизводиться в процессе проигрывания диска. Таким образом, MP3-файла будет выглядеть как на Рис. 1.



Рис. 1. MP3-файл со скрытыми данными

Алгоритм сокрытия информации в MP3 будет выглядеть следующим образом:

1. Увеличить параметр скорости (rate), находящийся в заголовке каждого фрейма. Вследствие чего увеличивается размер области данных.
2. В освободившееся место встроить сообщение.

Данный алгоритм обладает следующими преимуществами:

- абсолютной незаметностью при прослушивании;
- высокой пропускной способностью, порядка 20% от объема контейнера;
- высокой скоростью работы;
- простой схемной реализацией.

Процедура извлечения выполняется в обратном порядке.

6. ЗАКЛЮЧЕНИЕ

Информационная безопасность стала неотъемлемой частью современной системы связи. Основная задача информационной безопасности- защита целостности, доступности и конфиденциальности данных. Стеганография и криптография - это два направления цифровой безопасности, где стеганография скрывает сам факт существования сообщения, а криптография использует шифрование, чтобы сделать сообщение непонятным. Примером цифровой стеганографии является сокрытие и извлечение информации в аудио-файлах.

В данной работе представлено 2 алгоритма сокрытия информации в аудио-файлах. В первом алгоритме в качестве контейнера был выбран файл с разрешением WAV. Второй алгоритм реализует сокрытие данных в MP3. По итогу был разработан интерфейс пользователя на языке C++ в среде Visual Studio CLR Windows Forms, который позволяет выполнять работу по сокрытию

данных практически любому неподготовленному пользователю.

ЛИТЕРАТУРА

- [1] Введение в криптографию [Электронный ресурс]. – Режим доступа: https://www.rulit.me/data/programs/resources/pdf/Vvedenie-v-kriptografiyu_RuLit_Me_661516.pdf.
- [2] Гребенников, В.В. История Криптологии & Секретной Связи: глава 6.1, вступление в историю стеганографии / В.В. Гребенников, 2012 [Электронный ресурс]. – Режим доступа: <http://cryptohistory.ru/book/chast-6-istoriya-steganografii/61-vstuplenie/>.
- [3] Структура WAV-файла [Электронный ресурс]. – Режим доступа: <https://audiocoding.ru/articles/2008-05-22-wav-file-structure/>.
- [4] Структура MP3-файла [Электронный ресурс]. – Режим доступа: <https://audiocoding.ru/articles/2008-05-22-mp3-file-structure/>.
- [5] Структура MP3-файла [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/MP3>.
- [6] Кокорин, П.П. О методах стегоанализа в аудиофайлах / П.П. Кокорин // Труды СПИИРАН. – 2007. – № 4. – С. 239-246.
- [7] Забелин, М.А. Стегоанализ аудиоданных на основе методов сжатия / М.А. Забелин // Вестник СибГУТИ. – 2010. – № 1. – С. 41-46.

Применение взаимной информации в задаче стохастической привязки расфокусированных разноспектральных изображений

Р.М. Ибрагимов

Ульяновский государственный технический университет
Ульяновск, Россия
ibragimow.it@gmail.com

О.В. Шидиков

Ульяновский государственный технический университет
Ульяновск, Россия
oleg.shidikov@mail.ru

Аннотация—Рассмотрена эффективность применения релейных стохастических процедур привязки изображений в ситуации, когда одно из изображений расфокусировано. Произведено сравнение использования в качестве целевых функций привязки трех мер подобия изображений: взаимные информации Реньи, Шеннона и Тсаллиса.

Ключевые слова— привязка изображений, взаимная информация, Реньи, Шеннон, Тсаллис, расфокусировка, функция рассеяния точки.

1. ВВЕДЕНИЕ

Необходимость привязки изображений возникает при решении различных практических задач [1, 2], например, при обработке спутниковых и медицинских снимков, построении траектории автономных аппаратов и многих других. При этом привязываемые изображения могут быть получены в различных спектральных диапазонах, что дает дополнительную априорную неопределенность, поскольку помимо взаимных искажений пространственного характера для них характерны нелинейные яркостные искажения априорно неизвестного вида. Для таких условий хорошо себя зарекомендовали стохастические релейные процедуры привязки [3], синтезированные на базе теоретико-информационных, мер подобия изображений [4, 5]. На привязываемых изображениях могут присутствовать также дополнительные пространственные и яркостные искажения, в частности расфокусировка и смаз, вызванные, например, движением и различным временем экспозиции привязываемых изображений. В работе рассматривается влияние расфокусировки одного из изображений на качество привязки.

2. ПОСТАНОВКА ЗАДАЧИ

Привязка изображений $\mathbf{Z}^{(1)}$ и $\mathbf{Z}^{(2)}$ осуществляется по оцененным параметрам \mathbf{a} заданной модели деформаций. Параметры взаимных деформаций изображений оцениваются стохастической рекуррентной релейной процедурой [6]:

$$\hat{\mathbf{a}}_t = \hat{\mathbf{a}}_{t-1} \pm \Lambda_t \text{sign} \left(\beta \left(\mathbf{J} \left(\hat{\mathbf{a}}_{t-1}, \tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)} \right) \right) \right) \quad (1)$$

где $\beta(\mathbf{J}(\cdot))$ - стохастический градиент некоторой целевой функции $\mathbf{J}(\cdot)$; Λ_t - матрица обучения, определяющая скорость изменения оценок $\hat{\mathbf{a}}$ параметров на итерациях оценивания; $\tilde{\mathbf{Z}}_t^{(1)}$ - передискретизированное по оценкам $\hat{\mathbf{a}}_{t-1}$ изображение $\mathbf{Z}^{(1)}$; $t = \overline{1, T}$ - номер итерации. При этом стохастический градиент определяется не по всем

отсчетам изображений $\tilde{\mathbf{Z}}_t^{(1)}$ и $\mathbf{Z}^{(2)}$, а по небольшой их локальной выборке объема μ .

В работе в качестве модели деформаций использовалась модель подобия, включающая в себя параметры сдвига h_x и h_y по базовым осям, угла поворота φ и коэффициента масштаба k . В качестве ЦФ исследовались взаимные информации (ВИ) Реньи [7], Шеннона [8] и Тсаллиса [9]. Для имитации расфокусировки применялась функция рассеяния точки.

ВИ находятся через плотности и взаимные плотности распределения вероятностей (ПРВ) изображений локальных выборок отсчетов из $\tilde{\mathbf{Z}}_t^{(1)}$ и $\mathbf{Z}^{(2)}$. В свою очередь ПРВ в работе находились методом окон Парзена [10] как суперпозиция элементарных гауссовых ПРВ, центрированных на яркостях элементов, попавших в локальную выборку.

Для вычисления стохастического градиента ВИ Реньи, Шеннона и Тсаллиса были найдены аналитические выражения для частных производных по оцениваемым параметрам. Так, ВИ Тсаллиса:

$$S = H(\tilde{\mathbf{Z}}_t^{(1)}) + H(\tilde{\mathbf{Z}}_t^{(1)}) - (1-q)H(\tilde{\mathbf{Z}}_t^{(1)})H(\tilde{\mathbf{Z}}_t^{(1)}) - H(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) \quad (2)$$

где $H(\mathbf{Z}) = (q-1)^{-1} \sum_i p_z (1-p_z^{q-1})$ - энтропия Тсаллиса

порядка q изображения \mathbf{Z} , $H(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) =$

$$= (q-1)^{-1} \left(1 - \sum_i \sum_k p_{z_i, z_{ik}}^q \right) - \text{совместная энтропия}$$

Тсаллиса $\tilde{\mathbf{Z}}_t^{(1)}$ и $\mathbf{Z}^{(2)}$, p_z и $p_{z_i, z_{ik}}$ - оценки ПРВ и

совместной ПРВ яркостей, $i, k = \overline{1, \mu}$. Тогда после нахождения $\partial H(\tilde{\mathbf{Z}}_t^{(1)}) / \partial \bar{\mathbf{a}}$, $\partial H(\mathbf{Z}^{(2)}) / \partial \bar{\mathbf{a}}$,

$\partial H(\tilde{\mathbf{Z}}_t^{(1)}, \mathbf{Z}^{(2)}) / \partial \bar{\mathbf{a}}$, и подстановки в (2) получаем β .

3. ПРИМЕР РЕЗУЛЬТАТОВ РАБОТЫ АЛГОРИТМА

Экспериментальная проверка влияния расфокусировки одного из привязываемых изображений проводилась на разноспектральных изображениях, имеющих деформации сцены. При этом деформации задавались как моделированием, так и имели естественный характер. На рис. 1 приведен пример, когда изображение рис. 1а - оптического диапазона, рис. 1б - рис. 1г - инфракрасного. Инфракрасное изображение имеет относительно оптического следующие

деформации: $h_x = 3,4$, $h_y = 6,2$, $\varphi = 7,4^\circ$ и $k = 1,15$. При этом, рис. 1в не имеет расфокусировки, рис. 1в – расфокусировано с цилиндрической функцией рассеяния точки радиуса $\mathfrak{R}=1$, рис. 1в – $\mathfrak{R}=2$. На рис. 2а приведена зависимость оценки сдвига по оси x от числа итераций при использовании ВИ Тсалиса, здесь красная кривая соответствует отсутствию расфокусировки, зеленая – $\mathfrak{R}=1$, синяя – $\mathfrak{R}=2$. Видно, что расфокусировка приводит к смещенности оценок. Это подтверждает и зависимость евклидова расстояния рассогласования (интегральная оценка точности привязки изображений) от числа итераций при ВИ Шеннона (рис. 2б), а также зависимости ЦФ от параметров рассогласования (рис. 3).

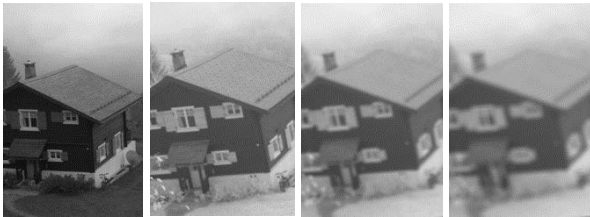


Рис. 1. Пример оптического и инфракрасного изображений с расфокусировкой.

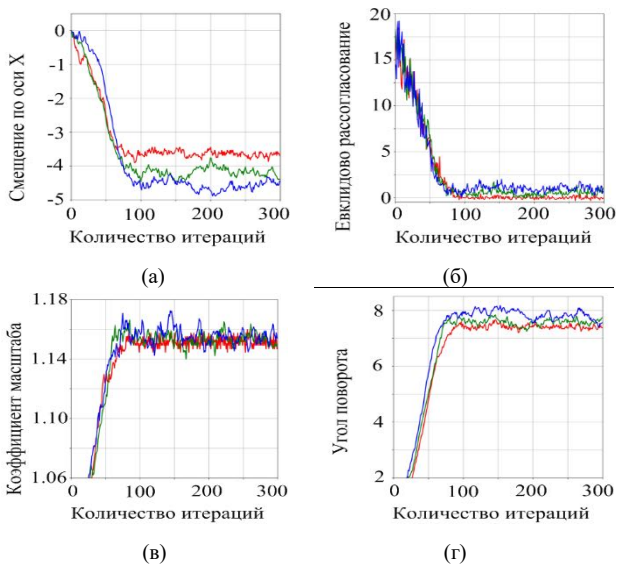


Рис. 2. Характеристики оценок параметров от числа итераций

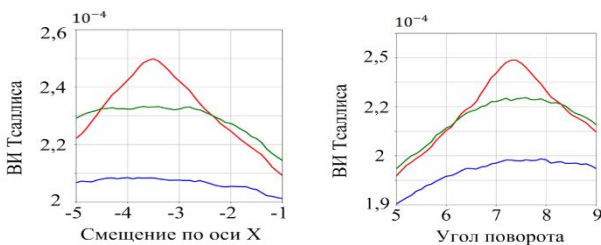


Рис. 3. Зависимости ЦФ от параметров рассогласования

4. ЗАКЛЮЧЕНИЕ

Анализ экспериментальных результатов показал, что расфокусировка одного из привязываемых разноспектральных изображений влияет на скорость

сходимости и погрешность привязки. При этом влияние на скорость сходимости незначительное, в отличие от влияния на погрешность. Кроме того, наблюдается смещенность отдельных оценок параметров деформаций, которая вызвана, скорее всего, тем, что расфокусировка приводит к анизотропности корреляционных свойств изображения, которые важны при использовании в качестве целевой функции ВИ. При этом отметим, что они ещё более важны при использовании такой меры подобия, как межкадровая корреляция. Однако суммарная погрешность всегда меньше, чем погрешность отдельных параметров, т.к. адаптивные стохастические алгоритмы оценивают вектор деформаций комплексно (рис. 2в,г).

В целом, все исследуемые ЦФ имеют близкие характеристики. Но ВИ Тсалиса показала более устойчивые результаты.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда № 22-21-00513, <https://rscf.ru/project/22-21-00513/>.

ЛИТЕРАТУРА

- [8] Камаев, А.Н. Визуальная навигация автономного необитаемого подводного аппарата с глобальным поиском связей между изображениями / А.Н. Камаев, Д.А. Карманов // Компьютерная оптика. – 2018. – Т. 42, № 3. – С. 457-467. DOI:10.18287/2412-6179-2018-42-3-457-467.
- [9] Magdeev, R.G. Efficiency of object identification for binary images / R.G. Magdeev, A.G. Tashlinskii // Computer Optics. – 2019. – Vol. 43(2). – P. 277-281. DOI:10.18287/2412-6179-2019-43-2-277-281.
- [10] Tashlinskii, A.G. Optimization of Goal Function Pseudogradient in the Problem of Interframe Geometrical Deformations Estimation / A.G. Tashlinskii // Pattern Recognition Techniques, Technology and Applications. – Austria, Vienna: I-Tech, 2008. – P. 249-280. DOI:10.5772/6244.
- [11] Chehade, W.E.H. Comparison of mutual information and its point similarity implementation for image registration / W.E.H. Chehade, P. Rogelj // International Journal of Electrical and Computer Engineering (IJECE). – 2021. – Vol. 11(3). – P. 2613-2620. DOI:10.11591/ijece.v11i3.pp2613-2620.
- [12] Tashlinskii, A.G. Similarity and dissimilarity measures as objective function in image registration/ A.G. Tashlinskii, S.V. Voronov // SWorld Journal. – 2015. – Vol. J11510. – P. 32-48.
- [13] Tashlinskii, A.G. Usage of mutual information as similarity measures for stochastic binding images / A.G. Tashlinskii, G.L. Safina, R.O. Kovalenko, R.M. Ibragimov // IEEE Conference Proceedings. – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649386.
- [14] Hongyang, J. Neonatal Fundus Image Registration and Mosaic Using Improved Speeded Up Robust Features Based on Shannon Entropy / J. Hongyang, G. Mengdi, Y. Kang, Z. Dongdong, M. He, Q. Wei, // Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). – 2021 – P. 3004-3007. DOI: 10.1109/EMBC46164.2021.9630593.
- [15] Pan, M. Medical Image Registration Based on Renyi's Quadratic Mutual Information / M. Pan, Z. Fen // IETE Journal of Research. – 2020. – P. 1-9. DOI: 10.1080/03772063.2020.1787873.
- [16] Sholehkerdar, A. Theoretical analysis of Tsallis entropy-based quality measure for weighted averaging image fusion / A. Sholehkerdar, J. Tavakolia, Z. Liub // Information Fusion. – 2020. – Vol. 58. – P. 69-81. DOI:10.1016/j.inffus.2019.12.010.
- [17] Mussa, H.Y. The Parzen Window method: In terms of two vectors and one matrix / H.Y. Mussa, J.B.O. Mitchell, A.M. Afzalb // Pattern Recognition Letters. – 2015. – Vol. 63. – P. 30-35. DOI: 10.1016/j.patrec.2015.06.002.

Оптимизация сверточных сетей с помощью квантизации и OpenVINO при распознавании снимков багажа

Н.А. Андриянов
Финансовый университет при Правительстве Российской Федерации
Москва, Россия
naandriyanov@fa.ru

Дж. Папакостас
Международный греческий университет
Салоники, Греция
grapak@cs.ihu.gr

Аннотация—Работа посвящена исследованию ускорения инференса нейронных сетей с помощью квантизации весов и применения инструментария Intel OpenVINO Toolkit. При этом в исследовании рассматриваются сверточные сети блочной архитектуры, обучаемые с нуля. Показано, что применение OpenVINO обеспечивает значительное ускорение без потери качества работы для таких сетей, в то время как квантизация приводит к существенным потерям качества.

Ключевые слова— компьютерное зрение, инференс, оптимизация, сверточные нейронные сети, квантизация, OpenVINO.

1. ВВЕДЕНИЕ

В последнее время с ростом сложности глубоких архитектур нейронных сетей, имеющих несколько миллиардов параметров [1], актуальным становится вопрос оптимизации работы таких моделей непосредственно при инференсе (логический вывод), т.е. при непосредственной работе сети. В частности, в задачах машинного зрения в ряде прикладных задач требуется вывод результатов практически в режиме реального времени.

Важной является задача распознавания и обнаружения объектов на изображениях [2]. Существует ряд подходов оптимизации, успешно применяемых в задаче ускорения работы глубоких сетей [3, 4], в том числе при обработке оптических изображений. В их числе прунинг, заключающийся в удалении весов и связей модели, квантизация весов и дистилляция, предназначенная для переноса знаний сети, например, с десктопного компьютера на мобильные устройства, не обладающие такими же большими вычислительными мощностями.

Другим решением является аппаратное ускорение или ускорение с учетом конкретных аппаратных платформ. К таким решениям относится ускоритель работы на процессорах Intel – OpenVINO Toolkit [5, 6]. Данный инструмент хорошо зарекомендовал себя при анализе оптических изображений. Однако применение таких алгоритмов при обработке изображений в других частотных диапазонах недостаточно изучено. В данной работе предлагается исследование ускорения инференса для рентгеновских снимков багажа и ручной клади на базе изображений аэропорта Баратаевка (г. Ульяновск). Задачи распознавания таких изображений более детально рассматривались в работе [7], а в настоящей работе основное внимание уделяется производительности сетей.

2. ОБУЧЕНИЕ И ИНФЕРЕНС

Рассмотрим задачу определения запрещенных к проносу предметов багажа и ручной клади. Очевидно, что важным показателем является именно полнота обнаружения таких предметов. При этом также важна и точность, поскольку при малой точности практически на каждый багаж будет требоваться ручная перепроверка. В исходной базе изображения были размечены под задачу распознавания. Задача детекции для снимков такого рода представляет дополнительные сложности. При сведении задачи к распознаванию удалось получить изображения отдельных предметов, которые были разделены на 2 класса. На рис. 1 представлены изображения запрещенного предмета (рис. 1а) и разрешенного предмета (рис. 1б).

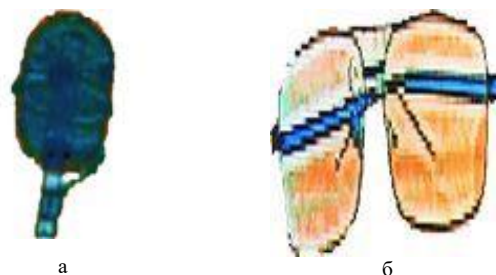


Рис. 1. Пример запрещенных (а) и разрешенных (б) объектов

В целом объем базы изображений составил 4000 изображений разрешенных объектов и 2000 – запрещенных. Для распознавания были обучены с нуля (с использованием библиотек Keras и TensorFlow) две сверточные сети. В первой, включающей 3 слоя свертки, использовались 128, 64 и 32 фильтра, оптимизация ADAM и 2000 эпох обучения, шаг обучения – 0,001. Вторая сеть состояла из 5 слоев свертки 256, 128, 64, 32 и 16 фильтров соответственно. Остальные параметры второй сети были выбраны, как и у первой. При этом размер обучающей выборки «разрешенных» изображений составил 2500 картинок, а «запрещенных» – 1200 картинок. Обучение происходило на видеокарте NVIDIA GeForce GTX 1060. Объем тестовой выборки составил 1500 и 800 изображений для «разрешенных» и «запрещенных» объектов соответственно. Размеры изображений 200 на 200 пикселей. Далее будем для простоты называть трехслойную сеть CNN-3, а пятислойную – CNN-5. В таблицах 1 и 2 представлены матрицы ошибок для CNN-3 и CNN-5.

Таблица I. МАТРИЦА ОШИБОК CNN-3

Истинный класс	Прогнозируемый класс		
	Разрешенный	Запрещенный	Всего
Разрешенный	1275	225 (FP)	1500
Запрещенный	168 (FN)	632 (TP)	800

Таблица II. МАТРИЦА ОШИБОК CNN-5

Истинный класс	Прогнозируемый класс		
	Разрешенный	Запрещенный	Всего
Разрешенный	1310	190 (FP)	1500
Запрещенный	127 (FN)	673 (TP)	800

Рассчитаем recall и precision для класса «запрещенных» предметов:

$$recall = \frac{TP}{TP + FN}, \quad precision = \frac{TP}{TP + FP} \quad (1)$$

где TP – количество «запрещенных» объектов, спрогнозированных, как «запрещенные»; FN – количество «запрещенных» объектов, спрогнозированных, как «разрешенные». Соответственно FP показывает количество «разрешенных» объектов, спрогнозированных, как «запрещенные».

В соответствии с выражением (1) $recall(3) = 0,79$, $recall(5) = 0,84$, $precision(3) = 0,74$, $precision(5) = 0,78$.

Инференс обученных нейронных сетей CNN-3 и CNN-5 будем выполнять на CPU Intel Core i7-8750. Выполним оценку показателя FPS (число кадров, обрабатываемых в секунду) на всей тестовой выборке. В связи с тем, что разные изображения в разных условиях могут обрабатываться за разное время, оценку производительности представим в виде $m_{FPS} \pm \sigma_{FPS}$, где m_{FPS} – среднее значение FPS по всем изображениям, а σ_{FPS} – стандартное отклонение по всем изображениям.

Далее выполним квантизацию весов до размеров INT8 (изначально сеть обучается в типе FP32), т.е. веса модели будут принимать только целочисленные значения в диапазоне [-128; 127]. Введем обозначения CNN-3q и CNN-5q соответственно для трехслойной и пятислойной сетей.

Наконец, реализуем Inference Engine с помощью инструмента Intel OpenVINO Toolkit. Следует отметить, что при первом запуске OpenVINO происходят конфигурационные настройки, другими словами, «разогрев нейросети». В связи с этим измерения производительности выполнялись уже после выполненной конфигурации под оптимальную работу на процессоре Intel. Оптимизированные с помощью OpenVINO сети из 3 и 5 слоев соответственно обозначим, как CNN-3ov и CNN-5ov.

В таблице 3 представлена оценка производительности сети, а также получаемые в условиях оптимизации характеристики полноты (recall) для тестовой выборки.

Анализ представленных результатов показывает, что использование OpenVINO целесообразно как для

ускорения сети, так и для обеспечения сохранения метрик эффективности.

Таблица III. ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ

Архитектура	FPS	Recall
CNN-3	2,23 ± 0,09	0,79
CNN-5	1,98 ± 0,12	0,84
CNN-3q	9,54 ± 0,11	0,68
CNN-5q	7,65 ± 0,10	0,72
CNN-3ov	29,03 ± 5,96	0,79
CNN-5ov	23,69 ± 6,74	0,84

3. ЗАКЛЮЧЕНИЕ

В работе предложено использовать ускорители для инференса нейронных сетей. Показано, что с помощью квантизации выполняется ускорение примерно в 3,5–4 раза. При этом полнота распознавания запрещенных объектов падает примерно на 10%. В случае использования инструментария Intel OpenVINO Toolkit удастся сохранить заявленные характеристики сети, при этом выигрыш для производительности в среднем составляет порядка 11-13 раз. Однако следует учитывать, что применение OpenVINO обеспечивает время инференса с большим разбросом.

БЛАГОДАРНОСТИ

Исследование выполнено при поддержке Совета по грантам Президента Российской Федерации в рамках реализации Проекта по Стипендии Президента РФ молодым ученым и аспирантам № СП-3738.2022.5 и частично при поддержке гранта РФФИ № 19-29-09048.

ЛИТЕРАТУРА

- [1] Romero, A. DeepMind Is Now the Undisputed Leader in Language AI with Gopher (280B) [Electronic resource]. – Access mode: <https://towardsdatascience.com/deepmind-is-now-the-undisputed-leader-in-language-ai-with-gopher-280b-79363106011f> (21.02.2022).
- [2] Андриянов, Н.А. Обнаружение объектов на изображении: от критериев Байеса и Неймана–Пирсона к детекторам на базе нейронных сетей EfficientDet / Н.А. Андриянов, В.Е. Дементьев, А.Г. Ташлинский // Компьютерная оптика. – 2022. – Т. 46, № 1. – С. 139-159. DOI: 10.18287/2412-6179-CO-922.
- [3] Kim, J. PQK: Model Compression via Pruning, Quantization, and Knowledge Distillation / J. Kim, S. Chang, N. Kwak // ArXiv preprint: 2106.14681, 2021.
- [4] Zhou, Y. Adaptive quantization for deep neural network / Y. Zhou, S.M. Moosavi-Dezfooli, N.M. Cheung, P. Frossard // ArXiv preprint: 1712.01048, 2017.
- [5] Andriyanov, N.A. Analysis of the Acceleration of Neural Networks Inference on Intel Processors Based on OpenVINO Toolkit // Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), 2020. – P. 1-5. DOI: 10.1109/SYNCHROINFO49631.2020.9166067.
- [6] Zunin, V.V. Intel OpenVINO Toolkit for Computer Vision: Object Detection and Semantic Segmentation // International Automation Conference (RusAutoCon), 2021. – P. 847-851.
- [7] Andriyanov, N. Automatic x-ray image analysis for aviation security within limited computing resources / N. Andriyanov, A.I. Volkov, A.N. Volkov, A. Gladkikh, S. Danilov // IOP Conference Series: Materials Science and Engineering, 2020. – Vol. 862. – P. 052009. DOI: 10.1088/1757-899X/862/5/052009.

Использование рецептивных полей в методах машинного обучения для обработки изображений, представленных выборками случайных отсчетов

В.Е. Анциперов

*Институт радиотехники и электроники им. В.А.
Котельникова РАН
Москва, Россия
anticperov@cplire.ru*

В.А. Кершнер

*Институт радиотехники и электроники им. В.А.
Котельникова РАН
Москва, Россия
vladkershner@mail.ru*

Аннотация—В работе предлагается и обосновывается способ обработки изображений, полученных путем формирования выборок случайных фотоотсчетов, с использованием рецептивных полей. Данный способ основан на известных фактах сенсорной нейробиологии, в частности в области функциональных особенностей нейронов зрительной системы человека. Предлагаемый метод позволит эффективно обрабатывать фотоотсчетные изображения средствами машинного обучения.

Ключевые слова— *Выборки случайных отсчетов, автоэнкодеры, рецептивные поля, сверточные нейронные сети, машинное обучение.*

1. ВВЕДЕНИЕ

Понятие рецептивного поля неразрывно связано с описанием функционирования соответствующих нейронов, в случае со зрительной системой такими нейронами являются ганглиозные клетки сетчатки глаза. Современные представления об этом классе зрительных нейронов достаточно расширились, что позволяет нам использовать уже имеющиеся модели зрительной системы на примере животных [1,2]. В свою очередь рецептивные поля связаны с глубокими сверточными сетями, которые широко используются в задачах по классификации изображений.

В ранних работах, посвященных новым способам представления данных для анализа изображений [3], было предложено использование выборок случайных фотоотсчетов для кодирования слабоинтенсивных изображений. На основе этих выборочных представлений с использованием принципов автокодирования был описан простой метод восстановления изображений. Стоит отметить, что автоэнкодеры являются относительно простой архитектурой нейронных сетей, направленной на обучение системы без учителя. Основная задача автокодировщика – преобразовать входные данные в подобные данные на выходе таким образом, чтобы максимально уменьшить объем входных данных и минимизировать любые искажения при соответствующих преобразованиях.

Однако для решения задач глубокого обучения, использование автоэнкодеров становится недостаточным, и требуется использование более эффективных методов машинного обучения, связанных с использованием более сложных алгоритмов, позволяющих не только преобразовывать данные из

одного представления в другое, но и выделять из них скрытую информацию - метаданные. В данной работе обсуждаются способы и методы обработки изображений для распознавания образов, обнаружения особенностей и т.д. с использованием рецептивных полей сенсорных нейронов.

2. СТРУКТУРА МОДЕЛИ РЕЦЕПТИВНЫХ ПОЛЕЙ ДЛЯ ОБРАБОТКИ ФОТООТСЧЕТНЫХ ИЗОБРАЖЕНИЙ

Поскольку подавляющее большинство известных методов обработки изображения были инспирированы зрительной системой животного, и в частности человека, стоит отметить некоторые сходства данных методов с этой системой. Основным объектом обработки информации в любой сенсорной системе человека является световое излучение, отраженное от поверхностей объектов, окружающих и находящихся в поле зрения наблюдателя. Регистрируемая входная информация, осуществляемая через зрительные нейроны сетчатки глаза (колбочки и палочки), проходит через биполярные клетки, передается в ганглиозные клетки и далее, уже в виде нервных импульсов, через зрительный нерв поступает в зрительную кору головного мозга (рис.1). Для наглядности эту систему можно упростить, представив в виде некоторой простой нейронной сети типа автоэнкодера с множеством слоёв. В таком случае входной слой будет состоять порядка из ста миллионов фоторецепторов, выходным слоем является зрительная кора, а набор из скрытых слоёв представлен наборами горизонтальных, биполярных, ганглиозных и других клеток. Стоит отметить, что ганглиозные клетки сетчатки глаза образуют мозаичную структуру из дендритов, покрывающую визуальное пространство сетчатки, и имеют соответствующие рецептивные поля.

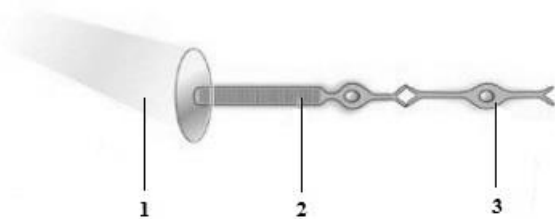


Рис. 1. Рецептивное поле зрительного нейрона: 1 – падающее световое излучение, 2 – фоторецепторная клетка, 3 – биполярная клетка

Понимания структуры сетчатки глаза и знания о нейронных сетях позволили нам разработать модель обработки изображения. Как и в случае со зрительной системой человека, модель обработки можно представить в виде аналогичной нейронной сети. В качестве предварительного этапа (первого слоя) используются выборки случайных фотоотсчетов. Использование выборочных представлений позволяет уменьшить объем исходного изображения, а значит более эффективно обрабатывать данные уже на начальном этапе. Следующим этапом является формирование параметрической модели, особенностью которой является упорядочивание компонент в области изображения. Такая модель имеет решетчатую структуру, что непосредственно связывает её с вышеуказанными фактами визуального восприятия человека.

На рис. 2 представлен пример обработки изображения. В качестве исходного изображения было взято цветное изображение 11860 (Country Pleasure ASB 3) из базы данных Emotive2000 [7]. Изображение было задано изначально в формате JPG с размерами 900 x 884 пикселей и имело глубину цвета 24 бита. Для уменьшения объема вычислений было произведено преобразование в формат GIF с размером 512 x 512 пикселей и в серую палитру с глубиной цвета 8 бит. Было сформировано выборочное представление с размером 500.000 отсчетов, осуществленное методом семплирования с отклонением. Количество использованных отсчетов было выбрано, исходя из двух параметров - конечного объема данных файла выборки, который используется в последующей обработке данных и отвечает за её скорость, и размера соответствующего изображения выборки, определяющего возможность интерпретировать обрабатываемые данные наблюдателю. На последнем этапе компоненты, представляющие собой отдельные отсчеты, были упорядочены в узлах решетчатой структуры, размер самих узлов и их шаг были определены таким образом, чтобы конечное изображение оставалось четким, но сама решетчатая структура визуально не воспринималась.

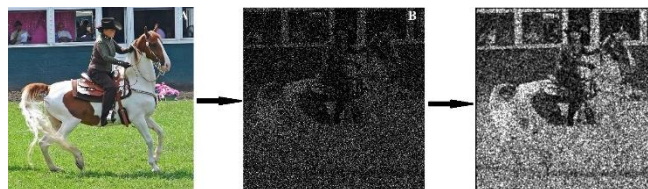


Рис. 2. Формирование параметрической модели: А – исходное изображение, В – выборка случайных фотоотсчетов, С – обработанное изображение

На основе этих данных была сформирована параметрическая модель исходного изображения, которая, как мы считаем, позволит использовать такие инструменты машинного обучения как сверточные нейронные сети для распознавания объектов, выделения особенностей и событий на изображениях, а также проведения оценок и анализа изображений.

БЛАГОДАРНОСТИ

Работа выполнена за счет бюджетного финансирования в рамках государственного задания в ИРЭ им. В.А. Котельникова РАН.

ЛИТЕРАТУРА

- [1] Antsiperov, V.E. New Generative Image Model for Variational Autoencoders Based on Counts Partition / V.E. Antsiperov // International Conference on Information Technology and Nanotechnology (ITNT). – 2021 – P. 1-8.
- [2] Hubel, D.H. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex / D.H. Hubel, T.N. Wiesel // The Journal of physiology. – 1962. – Vol. 160(1). – P. 106-154.
- [3] Hubel, D.H. Receptive fields of single neurones in the cat's striate cortex / D.H. Hubel, T.N. Wiesel // Brain Physiology and Psychology. – University of California Press, 2020. – P. 129-150.
- [4] Wienbar, S. The dynamic receptive fields of retinal ganglion cells / S. Wienbar, G.W. Schwartz // Progress in retinal and eye research. – 2018. – Vol. 67. – P. 102-117.
- [5] Fossum, E.R. Photon-Counting Image Sensors / E.R. Fossum, N. Teranishi. – MDPI, 2017.
- [6] Schwartz, G.W. The spatial structure of a nonlinear receptive field / G.W. Schwartz // Nature neuroscience. – 2012. – Vol. 15(11). – P. 1572-1580.
- [7] Robb, D.A. Crowdsourced Feedback With Imagery Rather Than Text: Would Designers Use It? / D.A. Robb, S. Padilla, B. Kalkreuter, M.J. Chantler // CHI'15: 33rd Annual ACM Conference on Human Factors in Computing Systems Proceedings, 2015.

Обнаружение вилочкового маркера уреаплазмоза на изображении фации биологической жидкости

М.Л. Албутова
Ульяновский государственный
университет
Ульяновск, Россия
albutova.m@yandex.ru

В.Р. Крашенинников
Ульяновский государственный
технический университет
Ульяновск, Россия
kvrulstu@mail.ru

О.Е. Маленова
Ульяновский государственный
технический университет
Ульяновск, Россия
nika-lilu@yandex.ru

Л.И. Трубникова
Ульяновский государственный
университет
Ульяновск, Россия
trubnikova-li@yandex.ru

Ю.Е. Кувайскова
Ульяновский государственный
технический университет
Ульяновск, Россия
u.kuvaiskova@mail.ru

Аннотация—Уреаплазмоз – распространенное инфекционное заболевание мочеполовой системы, вызываемое микроорганизмами. Диагностика в лабораторных условиях требует трудоемких и дорогостоящих методов. Значительно более быстрым и дешевым методом ранней диагностики является анализ изображения фации (тонкой пленки высушенной мочи). Признаком этого заболевания является наличие особых структур (маркеров) на изображениях. Один из этих маркеров имеет форму вилочки. В настоящей работе предложен алгоритм для обнаружения этого маркера на изображениях, введенных в компьютер. Испытания алгоритма на имеющихся изображениях фаций показали его высокую эффективность.

Ключевые слова— уреаплазмоз, изображение фации, маркер, обнаружение, алгоритм.

1. ВВЕДЕНИЕ

Уреаплазмоз – заболевание, вызываемое некоторыми видами уреаплазм и поражающее мочеполовую систему [1]. Лабораторная диагностика болезни требует применения трудоемких и дорогостоящих методов. Более быстрым методом ранней диагностики этого заболевания (и многих других) является анализ изображений фаций (тонких пленок) высушенных биологических жидкостей (в данном случае мочи). При высыхании жидкости содержащиеся в ней вещества кристаллизуются, образуя характерные структуры (маркеры). По виду этих маркеров можно с большой уверенностью судить о наличии многих заболеваний даже в самой ранней бессимптомной форме [2]. Плановые исследования в этом направлении были начаты и активно проводятся в НИИ геронтологии (Москва) под руководством академика В.Н. Шабалина. Однако при массовых обследованиях населения требуется анализ большого количества изображений, что утомляет зрение и снижает качество анализа. Поэтому актуальна задача разработки алгоритмов и программ для обнаружения и распознавания маркеров на изображениях, введенных в компьютер. К настоящему времени разработан ряд эффективных алгоритмов для распознавания маркеров некоторых видов, например, [3–5]. Исследования диагностики по фациям жидкостей ведутся в Ульяновском государственном университете коллективом под руководством профессора Л.И. Трубниковой. Этим коллективом предоставлены изображения фаций, использованные в настоящей работе, в которой предложен алгоритм для обнаружения одного

из видов этих маркеров, по форме напоминающих вилочку.

2. ВИЛОЧКОВЫЙ МАРКЕР УРЕАПЛАЗМОЗА

На Рис. 1 представлено исходное изображение части фации мочи с тремя маркерами, характерными при уреаплазмозе. Эти маркеры обведены красными кружками на Рис. 2. По форме они напоминают вилочку. На Рис. 3 представлено изображение без таких маркеров. Задача заключается в разработке алгоритма и программы для обнаружения вилочковых маркеров.

Для решения этой задачи сначала визуально анализировались имеющиеся изображения с этим маркером для выявления их характерных признаков. По наличию этих признаков на изображении фации в дальнейшем будет делаться вывод о присутствии маркера. Были выявлены следующие признаки: маркеры находятся на периферии фации, они темнее окружения и имеют два или три почти прямых зубца.



Рис. 1. Исходное изображение фации

3. ОБНАРУЖЕНИЕ МАРКЕРОВ

Обнаружение маркеров будем производить на изображении фации, преобразованном к оттенкам серого цвета путем усреднения составляющих RGB, так как цвета в данной задаче не несут полезной информации. Полученное изображение обрабатывается медианным фильтром для снижения влияния импульсных помех (посторонние примеси в жидкости и загрязнения предметного стекла).

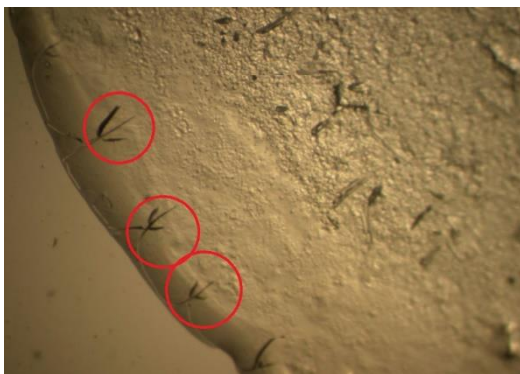


Рис. 2. Обнаруженные маркеры



Рис. 3. Изображение фации без маркеров уреоплазмоза

Для обнаружения маркеров сначала определяется «зона интереса», то есть область фации, где могут находиться искомые маркеры. Фация приблизительно круглая, так как получается высушиванием капли жидкости на предметном стекле. При этом фация может быть неполной, как на Рис. 1. Обмер изображений показал, что маркеры могут находиться между двумя окружностями: радиус внешней равен радиусу фации и радиус внутренней 75% радиуса фации. Для их построения следует определить радиус фации и ее центр. На краю фации имеется скачок яркости. Для обнаружения таких точек разность между максимальной и минимальной яркостью в скользящем окне сравнивается с порогом. Из полученного множества точек удаляются малые кластеры и формируются контур края фации. Координаты центра фации и ее радиус находится методом наименьших квадратов отклонений радиуса от расстояний между центром и точками края фации.

Искомые маркеры значительно темнее, чем фация в целом. Для их выделения находится средняя яркость изображения и выделяются все его части, яркость которых ниже средней. Участки с яркостью выше средней из дальнейшего рассмотрения удаляются.

Для выделения контуров был использован алгоритм Canny [6], с экспериментально подобранными значениями для двух порогов: порог минимума 100 и порог максимума 180. Они определяют значимость граничных точек, если значение градиента на данном фрагменте ниже минимального порога, то граница считается несущественной. Если же на фрагменте границы нет ни одного значения выше максимального порога, то он удаляется. Далее с помощью попиксельного обхода контура оставляются только границы достаточной длины

(не менее 50 пикселей). Обход реализован с помощью «алгоритма жука» [7].

На полученном изображении с помощью преобразования Хафа [8] найдем короткие приблизительно прямолинейные линии. Эмпирически установлено, что при увеличении $\times 400$ для изображения с разрешением 1080 длина искомых линий (зубцов «вилочки») около 60 пикселей.

Принятие решения о наличии маркера осуществляется на основании обнаружения пересечения двух прямых, из числа найденных на предыдущем этапе. Если на изображении обнаружено пересечение линий под углом более 20 градусов и менее 160, то принимается решение о наличии маркера.

4. ИСПЫТАНИЕ АЛГОРИТМА

Описанный алгоритм был тестирован на 154 реальных изображениях фаций. На 10 из этих изображений было 27 маркеров рассматриваемого типа, из которых было обнаружено 25. При этом все 10 изображений были идентифицированы как содержащие маркеры. Отметим, что для диагностики заболевания достаточно наличия хотя бы одного маркера на изображении, то есть не обязательно нужно обнаружить все маркеры. Поэтому в этом примере идентификация всех 10 изображений, содержащих маркеры, может считаться стопроцентным обнаружением, невзирая на 2 пропущенных маркера из 25 имеющихся.

На 144 изображениях маркеров уреоплазмоза не было, однако на 7 изображениях были ложные обнаружения, что составляет менее 5%.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-01-00613.

ЛИТЕРАТУРА

- [1] Трубникова, Л.И. Воспалительные заболевания специфической этиологии / Л.И. Трубникова, Н.В. Вознесенская, В.Д. Таджиева, Т.Ю. Корнилова, М.Л. Албутова, Н.Ю. Тихонова // Актуальные вопросы гинекологии». – Ульяновск: Ульяновский госуниверситет, 2019. – С. 306-341.
- [2] Шабалин, В.Н. Морфология биологических жидкостей человека / В.Н. Шабалин, С.Н. Шатохина. – М.: Христом, 2001. – 304 с.
- [3] Захарова, Г.П. Компьютерная идентификация отличительных признаков и аномалий в образцах многокомпонентных растворов биологических жидкостей / Г.П. Захарова, В.В. Шабалин // Российская оториноларингология. – 2014. – Т. 6, № 73. – С. 37-42.
- [4] Krasheninnikov, V.R. Algorithms for Markers Detection in Facies Images of Human Biological Fluids in Medical Diagnostics / V.R. Krasheninnikov, L.I. Trubnikova, A.S. Yashina, M.L. Albutova, O.E. Malenova // Intelligent Systems Reference Library 182. Springer International Publishing. – 2020. – P. 107-125.
- [5] Krasheninnikov, V.R. Algorithms for Automated Processing Images Blood Serum Fcies / V.R. Krasheninnikov, A.S. Kopylova // Pattern Recognition and Image Analysis. – 2012. – Vol. 22(4). – P. 583-592.
- [6] Canny, J. A Computational Approach to Edge Detection / J. Canny // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1986. – Vol. 6. – P. 679-698.
- [7] Потапов, Н.Н. Методы выделения контуров на изображениях / Н.Н. Потапов, П.А. Тупиков // 8-я международная конференция Распознавание образов и анализ изображений: новые информационные технологии. – 2007. – С. 105-108.
- [8] Ballard, D.H. Generalizing the Hough transform to direct arbitrary shapes / D.H. Ballard // Pattern Recognition. – 1981. – Vol. 13. – P. 111-122.

Новый алгоритм определения 3D координат центра роговицы глаза в системе бесконтактного айтрекинга

Г.И. Громилин
Институт автоматизации и
электрометрии СО РАН
Новосибирск, Россия
gromilin@iae.nsk.su

В.П. Косых
Институт автоматизации и
электрометрии СО РАН
Новосибирск, Россия
kosych@iae.nsk.su

Н.С. Яковенко
Институт автоматизации и
электрометрии СО РАН
Новосибирск, Россия
YakovenkoNS@iae.nsk.su

Аннотация—В ряде систем бесконтактного слежения за направлением взгляда по видеоданным, формируемым видеокамерой, удаленной от головы клиента, взаимное положение камеры и глаз клиента определяется по бликам, создаваемым на роговице глаз точечными источниками ИК излучения, жестко связанными с камерой. В работе предлагается новый алгоритм оценивания трехмерных координат центра роговицы по изображениям бликов от пары точечных источников излучения. Показано, что при расстоянии между глазом и камерой, значительно превышающем радиус роговицы, задача сводится к решению одного нелинейного уравнения, получаемому посредством быстро сходящейся итерационной процедуры. Приводятся результаты численного эксперимента, подтверждающего сохранение точности оценок.

Ключевые слова— бесконтактное определение направления взгляда, блики, координаты центра роговицы, трехмерная модель глаза.

1. ВВЕДЕНИЕ

Системы бесконтактного айтрекинга, т.е. слежения за направлением взгляда пользователя, находят разнообразное применение в различных областях нашей деятельности. Широкое распространение получила, например, линейка устройств Tobii Eye Tracker шведской фирмы Tobii [1], применяемая как в компьютерных играх, так и в качестве средства взаимодействия с компьютером пользователей с ограниченными физическими возможностями. Известны работы по привлечению систем бесконтактного айтрекинга к управлению различными подвижными механизмами [2], к контролю зрительного внимания водителя автомобиля [3], [4]. В маркетинге при разработке эффективных методов продвижения товаров с их помощью изучается порядок осмотра витрин покупателем [5], в психологии – характер движения глаз как результат влияния на поведение человека различных стрессовых ситуаций [6].

Основными элементами системы бесконтактного айтрекинга являются видеокамера, формирующая последовательность изображений лица клиента, и программно-алгоритмическое обеспечение, назначение которого заключается в выделении в изображении зрачков и определении по тем или иным признакам направления взгляда. Как правило, направление взгляда оценивается по положению центра зрачка относительно некоторых реперных меток, привязанных к системе координат камеры. Реперные метки в большинстве систем создаются искусственно точечными источниками инфракрасного (ИК) излучения, жестко связанными с видеокамерой и создающими блики на поверхности роговицы глаза.

Положение центра зрачка относительно этих бликов несет информацию об изменении направления взгляда. Методы определения направления взгляда по взаимному положению центра зрачка и бликов условно можно разделить на два класса. К первому относятся регрессионные методы, использующие эмпирически устанавливаемую в процессе калибровки зависимость между смещением зрачка относительно бликов и известным направлением взгляда [7]. Ко второму – методы, опирающиеся на трехмерную (3D) модель глаза. В методах, основанных на трехмерной модели глаза [8], направление взгляда вычисляется по трехмерным координатам центра роговицы глаза и центра зрачка, связанных соотношениями, установленными в модели «схематического глаза» Гульстранда [9].

2. АЛГОРИТМ ОПРЕДЕЛЕНИЯ 3D КООРДИНАТ ЦЕНТРА РОГОВИЦЫ ГЛАЗА

В ряде публикаций, посвященных бесконтактным системам определения направления взгляда, отмечается, что регрессионные методы, использующие только двумерную информацию о взаимном положении изображений центра зрачка и опорного блика, чрезвычайно чувствительны к изменению положения головы оператора по отношению к положению в момент калибровки [8], [10]. Методы, основанные на 3D модели глаза, дают более стабильную оценку направления взгляда при перемещении головы в пределах нескольких сантиметров как по дальности, так и в плоскости, ортогональной оптической оси камеры. Однако они требуют применения как минимум двух пространственно разнесенных источников реперных бликов и алгоритмов для оценивания 3D координат центра роговицы и зрачка. Первым этапом реализации этих методов является вычисление 3D координат центра роговицы по координатам изображений бликов от реперных источников. Оригинальный алгоритм вычисления координат центра роговицы по изображениям двух бликов, основанный на решении системы трансцендентных уравнений, был предложен авторами работы [10]. Система опирается на модель глаза [9], в которой поверхность роговицы аппроксимируется сферой. В нашей работе предложен более простой алгоритм решения этой задачи.

На рисунке 1 приведена схема формирования изображений бликов от двух пространственно разнесенных источников, отраженных сферической поверхностью роговицы глаза, в плоскости сенсора камеры. Здесь $OXYZ$ – система координат камеры, а P – плоскость ее сенсора.

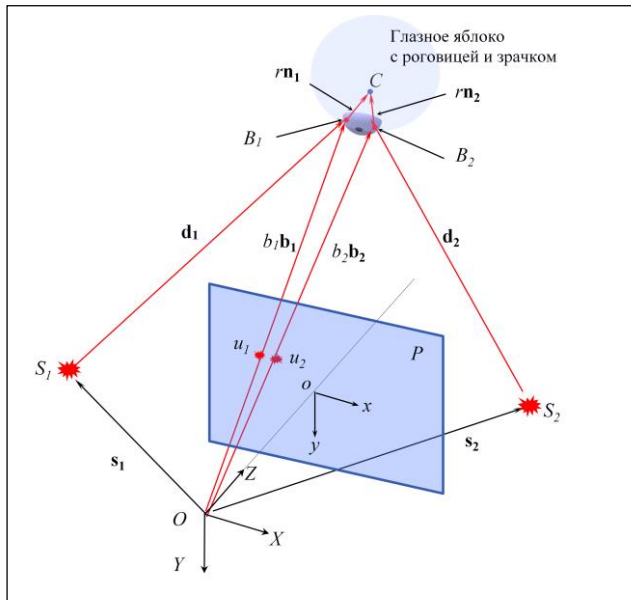


Рис. 1. Схема формирования изображений бликов

Начало системы координат камеры совмещено с оптическим центром объектива O , а ось OZ – с его оптической осью. Точечные источники ИК излучения, создающие опорные блики, находятся в точках S_1 и S_2 , центр сферы радиуса r , аппроксимирующей поверхность роговицы – в точке C . Излучение от источников создает блики на роговице в точках B_1 и B_2 , их изображения u_1 и u_2 формируются в плоскости P . Единичные векторы нормалей к сферической поверхности роговицы в точках бликов обозначены \mathbf{n}_1 и \mathbf{n}_2 . Положение источников задано конструкцией системы айтрекинга и определяется векторами \mathbf{s}_1 и \mathbf{s}_2 , а направление на блики задается единичными векторами \mathbf{b}_1 и \mathbf{b}_2 , которые определяются двумерными координатами изображений бликов u_1 и u_2 .

Как следует из рисунка, векторы источников бликов \mathbf{s}_i , направления падающего на роговицу \mathbf{d}_i и отраженного в объектив камеры \mathbf{b}_i лучей связаны соотношением

$$\mathbf{d}_i = b_i \mathbf{b}_i - \mathbf{s}_i, \quad i = 1, 2, \quad (1)$$

где b_i – расстояние от оптического центра объектива камеры до i -го блика на роговице. Показано, что нормаль \mathbf{n}_i к поверхности в точке отражения луча B_i и векторы \mathbf{b}_i и \mathbf{s}_i удовлетворяют условию:

$$\mathbf{n}_i = (|b_i \mathbf{b}_i - \mathbf{s}_i| \mathbf{b}_i + b_i \mathbf{b}_i - \mathbf{s}_i) / \|(b_i \mathbf{b}_i - \mathbf{s}_i) \mathbf{b}_i + b_i \mathbf{b}_i - \mathbf{s}_i\|. \quad (2)$$

С другой стороны, нормаль к сферической поверхности роговицы направлена в центр сферы, поэтому:

$$\mathbf{c} = b_1 \mathbf{b}_1 + r \mathbf{n}_1 = b_2 \mathbf{b}_2 + r \mathbf{n}_2, \quad (3)$$

где r – радиус роговицы. Из соотношения (3) следует нелинейное векторное уравнение относительно двух неизвестных, b_1 и b_2 :

$$b_1 \mathbf{b}_1 - b_2 \mathbf{b}_2 + r(\mathbf{n}_1(b_1) - \mathbf{n}_2(b_2)) = \mathbf{0}, \quad (4)$$

для решения которого можно воспользоваться методом наименьших квадратов (МНК).

Если расстояние от камеры до глаза значительно превышает радиус роговицы, можно положить $b_1 = b_2 = b$ и из (4) получить уравнение (5)

$$b(\mathbf{b}_1 - \mathbf{b}_2) + r(\mathbf{n}_1(b) - \mathbf{n}_2(b)) = \mathbf{0}. \quad (5)$$

Его МНК-решение \tilde{b} получается в результате итерационной процедуры и позволяет вычислить 3D координаты точки C :

$$\mathbf{c} = \sum_{i=1,2} [\tilde{b} \mathbf{b}_i + r(\tilde{b} \mathbf{b}_i - \mathbf{s}_i) \mathbf{b}_i + \tilde{b} \mathbf{b}_i - \mathbf{s}_i] / \|\tilde{b} \mathbf{b}_i - \mathbf{s}_i \mathbf{b}_i + \tilde{b} \mathbf{b}_i - \mathbf{s}_i\| / 2 \quad (6)$$

3. ЗАКЛЮЧЕНИЕ

Алгоритм был проверен посредством моделирования хода лучей от пары точечных источников излучения, отраженных имитирующей роговицу зеркальной сферой радиусом $r = 7.5$ мм в объектив камеры, формирующий изображения бликов в плоскости сенсора P . В системе координат камеры (все координаты указаны в мм) источники излучения размещались в точках $S_1(-120, 0, 0)$ и $S_2(120, 0, 0)$, центр сферы – в точке $C(0, 0, 480)$, т.е. расстояние от оптического центра камеры до поверхности сферы более чем в 60 раз превышало ее радиус. Формат сенсора камеры – 1280×1024 пикселей, размер пикселя – $0,0053 \times 0,0053$ мм², фокусное расстояние объектива – 25 мм. Координаты изображений бликов измерялись с ошибками, независимо распределенными вдоль осей ox и oy по нормальному закону с СКО 1 пикс. По 500 реализациям измерений координат изображений бликов среднеквадратичные ошибки МНК-оценок координат центра роговицы составили: по осям X и Y – 0.1 мм, по оси Z – 17.7 мм. Для решения уравнения (4) требовалось не менее 10 итераций, тогда как решение уравнения (5) получалось не более чем за 5 итераций.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Минобрнауки РФ (проект 121022000116-0).

ЛИТЕРАТУРА

- [1] URL: <https://gaming.tobii.com> (21/02/2022).
- [2] Eid, M.A. A novel eye-gaze-controlled wheelchair system for navigating unknown environments: case study with a person with ALS / M.A. Eid, N. Giakoumidis, A. El Saddik // IEEE Access. – 2016. – Vol. 4. – P. 558-573. DOI:10.1109/ACCESS.2016.2520093.
- [3] Said, Sh. Real Time Eye Tracking and Detection- A Driving Assistance System / Sh. Said, S. AlKork, T. Beyrouthy, M. Hassan, O.E. Abdellatif, M.F. Abdraboo // Advances in Science, Technology and Engineering Systems Journal. – 2018. – Vol. 3(6). – P. 446-454.
- [4] Nguyen, T.P. Eye tracking system to detect driver drowsiness / T.P. Nguyen, M.T. Chew, S. Demidenko // 6th International Conference on Automation, Robotics and Applications (ICARA). – 2015. – P. 472-477. DOI: 10.1109/ICARA.2015.7081194.
- [5] Chandon, P. Measuring the Value of Point-of-Purchase Marketing with Commercial Eye-Tracking Data / P. Chandon, J.W. Hutchinson, E.T. Bradlow, S.H. Young // INSEAD Business School Research. – 2007. – P. 46.
- [6] Барabanщиков, В.А. Айтрекинг в психологической науке и практике / В.А. Барabanщиков. – М.: Когито-Центр, 2015. – 410 с.
- [7] Cherif, Z. An adaptive calibration of an infrared light device used for gaze tracking / Z. Cherif, A. Nait-Ali, J. Motsch, M. Krebs // Proc. 19th IEEE Instrum. Meas. Technol. Conf. – 2002. – Vol. 2. – P. 1029-1033.
- [8] Hennessey, C. A Single Camera Eye-Gaze Tracking System with Free Head Motion / C. Hennessey, B. Noureddin, P. Lawrence // Proceedings of the Symposium on Eye Tracking Research & Applications. – 2006. – P. 87-94. DOI: 10.1145/1117309.1117349.
- [9] Goss, D.A. Introduction to the Optics of the Eye / D.A. Goss, R.W. West. – Boston, MA: Butterworth Heinemann, 2001. – 252 p.
- [10] Shih, S. A novel approach to 3-D gaze tracking using stereo cameras / S. Shih, J. Liu // IEEE Transactions on Systems, Man, and Cybernetics. Part B. – 2004. – Vol. 34(1). – P. 234-245.

Алгоритм разреженного представления в задаче устранения шумов на изображениях

Д.В. Архипова

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
mazyaikinadasha@gmail.com

Е.В. Гошин

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
goshine@ssau.ru

Аннотация— Настоящая статья посвящена исследованию восстановления изображений с использованием разреженного представления. Разреженное представление представляет собой описание изображения в форме коэффициентов для фрагментов, выбранных из заранее заданного словаря. В настоящей статье предлагается авторская интерпретация общего подхода к восстановлению изображений с использованием разреженного представления и приводятся результаты экспериментов для одной из реализаций этого подхода.

Ключевые слова— устранение шумов, разреженное представление, поиск совпадения

1. ВВЕДЕНИЕ

Задача устранения шумов на изображениях является классической проблемой обработки изображений. Тем не менее, устранение шумов считается сложной и открытой задачей. Основная причина этого заключается в том, что с математической точки зрения устранение шумов является обратной задачей и ее решение не единственно.

К настоящему времени исследователи уже предложили различные методы уменьшения шума. Каждый метод имеет свои преимущества и недостатки.

Методы пространственной области направлены на удаление шума путем вычисления значения серого для каждого пикселя на основе корреляции между пикселями/областями изображения в исходном изображении. В общем, методы пространственной области можно разделить на две категории: фильтрация в пространственной области [1] и методы вариационного шумоподавления [2]. К методам пространственной области относят: вариационные методы шумоподавления, разреженное представление, аппроксимация низкого ранга и т. д.

Методы шумоподавления изображения постепенно развивались от первоначальных методов пространственной области до современных методов в некотором новом пространстве, формируемом после преобразования. Первоначально такие методы были разработаны для преобразования Фурье [3]. Ключевая идея таких методов заключается в том, что характеристики преобразованных изображения и шума различны.

Рассмотренные выше методы вариационного шумоподавления относятся к схемам оптимизации на основе моделей, которые находят оптимальные решения

для восстановления изображения с шумоподавлением. Напротив, методы шумоподавления на основе CNN [4] пытаются оценить функцию отображения, оптимизируя функцию потерь на обучающем наборе, содержащем пары чистых и зашумлённых изображений.

В последние годы наблюдается рост интереса к устранению шумов на основе разреженного представления сигналов. Идея таких методов заключается в представлении каждого фрагмента изображения в виде линейной комбинации нескольких фрагментов сверхполного словаря.

В настоящей работе будет предложена авторская интерпретация подхода к восстановлению изображений с использованием разреженного представления.

2. ТРАДИЦИОННЫЕ ПОДХОДЫ

Математически задачу устранения шумов можно смоделировать следующим образом. Пусть изображение формируется в виде: $y = x + n$, где y — наблюдаемое зашумленное изображение, x — неизвестное исходное изображение, а n — аддитивный белый гауссовский шум.

Одним из наиболее распространённых подходов к восстановлению исходного изображения является байесовское оценивание. Этот подход может быть сформулирован в форме задачи минимизации функции:

$$f(x) = \frac{1}{2} \|\hat{x} - y\|_2^2 + G(x),$$

где \hat{x} — оценка исходного изображения, а $G(x)$ — штрафная функция гладкости, задаваемая из некоторых априорных соображений.

Другой подход, имеющий ценность в рамках настоящей работы, основан на применении некоторого фильтра (как правило, фильтра низких частот) после преобразования Фурье и может быть записан как последовательность преобразований:

$$\hat{x} = F^{-1}(G(F(y))),$$

где $F(\cdot)$ и $F^{-1}(\cdot)$ — прямое и обратное преобразование Фурье, соответственно.

3. РАЗРЕЖЕННОЕ ПРЕДСТАВЛЕНИЕ

В настоящем исследовании в качестве основы для разработки был использован алгоритм, использующий разреженное представление изображений.

Разреженное представление представляет процесс шумоподавления в следующей форме:

$$f(x) = \frac{1}{2} \|x - y\|_2^2 + \sum_{j=1}^K |a_j|^p, x = Da$$

где D – словарь (матрица) размером $N \times K$. N – размерность сигнала, который мы собираемся моделировать. K – размер словаря, a – вектор с малым числом ненулевых элементов.

Традиционно, решение задачи восстановления изображений с помощью разреженного представления формулируется следующим образом.

Каждый фрагмент изображения может быть представлен как линейная комбинация нескольких фрагментов из избыточного словаря D . Оценка MAP для шумоподавления этого фрагмента изображения для известного словаря строится путем решения выражения:

$$\hat{a} = \|a\|_0, \|Da - y\|_2^2 \leq T_0.$$

Нахождение наилучшего словаря для представления данных $\{y_i\}_{i=1}^N$ в виде разреженного представления осуществляется с помощью поиска минимума следующего выражения

$$\{\|Y - Da\|_F^2\}, \forall i \|a_i\|_0 \leq T_0.$$

Здесь T_0 – фиксированное количество ненулевых записей, а обозначение $\|a\|_F$ обозначает Евклидову норму.

4. ПРЕДЛАГАЕМЫЙ ПОДХОД

Рассмотренный выше традиционный подход не лишён недостатков. В частности, он требует одновременного решения задачи вычисления разреженного представления и минимизации априорного члена.

В настоящей работе предлагается подход, аналогичный устранению шумов, основанному на преобразовании Фурье.

Пусть $J(y, D)$ – обратимое точное разреженное представление изображения, то есть выполняется условие:

$$y = J^{-1}(J(y, D), D).$$

Рассмотрим класс преобразований разреженных представлений $G(\cdot)$ и сформулируем процедуру восстановления изображений в виде

$$\hat{x} = J^{-1}(G(J(y, D))).$$

В качестве простейшей реализации этого подхода по аналогии с традиционным подходом зададим функцию $G(\cdot)$ как ограничение на максимальное число ненулевых коэффициентов разреженного представления. Таким образом, T_0 наибольших коэффициентов сохраняется без изменения, а остальные – обнуляются.

5. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

В работе было проведено экспериментальное исследование с целью демонстрации преимуществ разреженного представления по сравнению с двумя рассмотренными подходами. Помимо этого, было

показано, что предложенный подход к использованию разреженного представления незначительно влияет на результат восстановления, вместе с тем открывая простор для выбора различных видов фильтрующих функций в пространстве коэффициентов разреженного представления.

В качестве алгоритмов для вычисления разреженного представления использован ортонормированный поиск совпадений из-за его простоты и эффективности.

Полученные изображения и сравнительная таблица результатов представлены ниже. Для сравнения качества восстановления изображения во всех вышеописанных случаях было использовано значение PSNR.



Рис. 1. (а) зашумленное изображение, (б) изображение, восстановленное с помощью разреженного представления, (в) изображение, восстановленное с помощью преобразования Фурье, (г) изображение, восстановленное с помощью байесовского подхода

Таблица I. ЗНАЧЕНИЯ PSNR

Метода	σ		
	15	25	50
Разреженное представление	34,78	33,39	31,42
Преобразование Фурье	28,59	27,89	25,52
Байесовское оценивание	26,12	24,06	22,79

6. ЗАКЛЮЧЕНИЕ

В настоящей работе рассматривается влияние степени зашумленности исходного изображения на результат восстановления, получаемый при помощи разреженного представления.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках государственного задания по теме 0777-2020-0017 и программы «Приоритет-2030».

ЛИТЕРАТУРА

- [1] Wiener, N. Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications / N. Wiener. – Cambridge: MIT Press, 1949. – P. 174.
- [2] Katsaggelos, A.K. Digital image restoration / A.K. Katsaggelos // IEEE Signal Processing Magazine. – 2012. – Vol. 14. – P. 24-41.
- [3] Gopinathan, S. Wavelet and FFT Based Image Denoising Using Non-linear Filters / S. Gopinathan, R. Kokila // International Journal of Electrical and Computer Engineering. – 2015. – Vol. 5. – P. 1018-1026.
- [4] Chen, Y.Y. Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration / Y.Y.Chen, T. Pock // IEEE Trans Pattern Anal Mach Intell. – 2017. – Vol. 39(6). – P. 1256-1272.
- [5] Gorodnitsky, I.F. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted norm minimization algorithm / I.F. Gorodnitsky, B.D. Rao // IEEE Trans. Signal Process. – 1997. – Vol. 45. – P. 600-616.
- [6] Lewicki, M.S. A probabilistic framework for the adaptation and comparison of image codes / M.S. Lewicki, B.A. Olshausen // J. Opt. Soc. Amer. A: Opt., Image Sci. Vision. – 1999. – Vol. 16. – P. 1587-1601.
- [7] Myasnikov, V.V. Reconstruction of functions and digital images using sign representations / V.V. Myasnikov // Computer Optics. – 2019. – Vol. 43(6). – P. 1041-1052. DOI: 10.18287/2412-6179-2019-43-6-1041-1052.

Исследование влияния шумов на результат сверхразрешения по цветным изображениям

Д.А. Аксенова

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
darinaksema@gmail.com

Е.В. Гошин

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
goshine@yandex.ru

Аннотация—В настоящей работе представлено исследование влияния различных типов шумов на результат восстановления изображения методом многокадрового сверхразрешения. В качестве видов шумов исследовались: аддитивный шум, шум вида «соль-перец» для цветных пикселей и шум вида «соль-перец» для отдельных слоёв. Для регуляризации использовались квадратичная функция и функция потерь Хьюбера.

Ключевые слова— сверхразрешение, восстановление изображения, метод максимального правдоподобия.

1. ВВЕДЕНИЕ

Задача восстановления изображений, в том числе методами сверхразрешения, распространена в современной научной литературе [1]. В обработке изображений и большинстве ее приложений требуются изображения с высоким или сверхвысоким разрешением. Основная причина данного требования – потребность в лучшем восприятии и интерпретации человеком или машиной для принятия решений [2]. Разрешение изображения обычно связано с количеством пикселей на дюйм (ppi) или точек на дюйм (dpi). Малое количество пикселей ведет к низкому разрешению, соответственно, но с увеличением плотности пикселей, растет качество изображения и его информативность. Однако, по мере увеличения количества пикселей, количество света, попадающего на каждый датчик, уменьшается, что приводит к появлению шумов.

Методы реконструкции изображений со сверхвысоким разрешением используются в разных областях для анализа и извлечения информации из изображений. Например, в медицинской визуализации, спутниковой съемке, распознавании образов, криминалистике [3-4].

Глобально существуют подходы, основанные на восстановлении изображений со сверхвысоким разрешением на основе одного изображения с низким разрешением и на основе последовательности изображений с низким разрешением [5]. В данной работе были исследованы методы реконструкции изображений со сверхвысоким разрешением на из последовательности изображений с низким разрешением.

2. ПОСТАНОВКА ЗАДАЧИ

Предположим, что имеется набор изображений с низким разрешением, которые были получены при помощи одного устройства и в одинаковых условиях. Для решения задачи сверхвысокого разрешения (СР) воспользуемся алгоритмами построения изображения с

высоким разрешением на основе набора искаженных системой визуализации изображений с низким разрешением.

Модель наблюдения будет выглядеть следующим образом:

$$y_k = DB_k^{cam} M_k B_k^{atm} x + \vartheta_k, \quad k = \overline{1, N}, \quad (1)$$

где y_k – набор изображений с низким разрешением; x – исходное изображение (изображение со сверхразрешением, рис. 1), M_k – проективное преобразование для k изображения, моделирует движение камеры относительно сцены, B_k^{cam} – размытие или функция рассеяния точки (ФРТ), B_k^{atm} – размытие за счет свойств атмосферы, D – понижающая дискретизация при помощи ПЗС-матрицы, а ϑ_k – аддитивный шум k изображения соответственно [6].

Модель можно использовать для инверсии процесса и получения изображение с высоким разрешением из набора изображений низкого разрешения путем оценки изображение высокого разрешения. Для восстановления изображения со сверхвысоким разрешением был использован метод максимального правдоподобия.

Оцененное изображение \hat{z} при помощи метода максимального правдоподобия выражается формулой:

$$\hat{z} = \arg \min_x [\sum_{k=1}^N \|DM_k z - y_k\|_p^p], \quad (2)$$

где k принимает значения в диапазоне от 1 до N , а p принимает значение 1 или 2. При $p = 2$, формируя итоговое изображение итеративно при помощи метода наискорейшего спуска, на каждом шаге оцененное изображение отыскивается по следующей формуле:

$$\widehat{z}_{n+1} = \widehat{z}_n - \beta \sum_{k=1}^N M_k^T D^T (DM_k \widehat{z}_n - y_k). \quad (3)$$

Для восстановления изображения высокого разрешения из последовательности кадров использовался алгоритм наискорейшего спуска, который состоит из следующих шагов:

1. Инициализация текущего изображения высокого разрешения (нулями или случайными значениями).
2. Формирование изображения низкого разрешения из изображения высокого разрешения применением операторов аффинного преобразования, размытия и прореживания.
3. Вычисление разностного изображения между полученным на предыдущем шаге изображением и текущим изображением с низким разрешением из набора.

4. Повышение разрешения разностного изображения с заполнением недостающих пикселей нулями.
5. Применение оператора размытия.
6. Применение обратного аффинного преобразования.
7. Обновление значений текущего изображения с учётом полученной разности и регуляризационного члена.
8. Повторение шагов 2-7 для различных изображений низкого разрешения в течение заданного числа итераций.



Рис. 1. Фрагмент исходного изображения высокого разрешения



Рис. 2. Результат восстановления для различных видов шумов и регуляризации: (а) аддитивный шум, L2 регуляризация, (б) аддитивный шум, функция Хьюбера, (в) импульсный шум послыбно, L2 регуляризация, (г) импульсный шум послыбно, функция Хьюбера, (д) импульсный шум, L2 регуляризация, (е) импульсный шум, функция Хьюбера

Подобный подход применим для восстановления изображения на основе набора монохромных изображений. В работе [7] в рамках данного подхода ранее было приведено экспериментальное исследование влияния шума на изображениях низкого разрешения и параметров минимизируемой функции на результат восстановления. В качестве функции потерь использовалась функция Хьюбера. В настоящей работе данный подход исследуется применительно к цветным изображениям. Также в настоящей работе помимо аддитивного шума исследуется импульсный шум типа «соль-перец». Рассмотрено два специфичных для цветных изображений варианта: когда шум применяется целиком ко всему пикселю или для каждого цветового слоя независимо.

3. ЗАКЛЮЧЕНИЕ

В ходе выполнения работы было проведено исследование влияния шумов на результат восстановления. Полученные результаты представлены на рис.2. Визуальная оценка показывает, что аддитивные шумы, присутствующие в 100% пикселей изображений низкого разрешения сглаживаются и не имеют значительного влияния на результирующее изображение, а импульсный шум, присутствующий на менее 1% пикселей не может быть подавлен ни одним из способов регуляризации, представленными в работе.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках государственного задания по теме 0777-2020-0017 и программы «Приоритет-2030».

ЛИТЕРАТУРА

- [1] Максимов, А.И. Метод оптимального линейного сверхразрешающего восстановления изображений / А.И. Максимов, В.В. Сергеев // Компьютерная оптика. – 2021. – Т. 45, № 5. – С. 692-701. DOI: 10.18287/2412-6179-СО-909.
- [2] Кузнецов, А.В. Исследование влияния размера камуфлируемой области на качество RS-изображений / А.В. Кузнецов, М.В. Гашников // Информационные технологии и нанотехнологии (ИТНТ-2021): сб. тр. по материалам VII Междунар. конф. и молодеж. шк. – 2021. – Т. 2. – С. 021712.
- [3] Yang, J. Image super-resolution: Historical overview and future challenges / J. Yang, T. Huang // CRC Press. – 2017. – P. 1-33.
- [4] Мясников, В.В. Реконструкция изображения как задача квадратичного программирования / В.В. Мясников, А.В. Веричев // Информационные технологии и нанотехнологии (ИТНТ-2020): сб. тр. по материалам VI Междунар. конф. и молодеж. шк. – 2020. – Т. 2. – С. 47-56.
- [5] Farsiu, S. Advances and Challenges in Super-Resolution / S. Farsiu // Wiley Periodicals, Inc. – 2004. – Vol. 14. – P. 47-57.
- [6] Khattab, M.M. Regularization-based multi-frame super-resolution: A systematic review / M.M. Khattab // Journal of King Saud University – Computer and Information Sciences. – 2000. – Vol. 3. – P. 755-776.
- [7] Goshin, Y. Noise Compensation in Super-Resolution Problem Using the Huber Loss Function / Y. Goshin, D. Arkhipova // Proceedings of ITNT 7th IEEE International Conference on Information Technology and Nanotechnology. – 2021.

Генеративная модель автокодировщиков изображений на основе рецептивных полей

В.Е. Агциперов

Институт радиотехники и электроники им. В.А. Котельникова РАН

Москва, Россия

antciperov@cplire.ru

Аннотация—В работе обсуждается новая концепция автокодировщиков, ориентированных на кодирование цифровых изображений. Проблема синтеза подобных автокодировщиков формализуется в рамках подхода машинного обучения, понимаемого здесь как обучение (без учителя) по данным самого же изображения. Особенностью рассматриваемого подхода является специальное представление изображений с помощью выборок отсчетов контролируемого размера –выборочных представлений. Основываясь на специфике представления формулируется генеративная (порождающая) модель автокодировщиков, которая конкретизируется затем до вероятностной параметрической модели отсчетов в виде смеси компонент экспоненциального скошенного семейства. По аналогии с концепцией рецептивных полей компоненты смеси выбираются в виде финитных вероятностных распределений, что обеспечивает синтез быстрых рекуррентных процедур кодирования. Обсуждается связь синтезированных автокодировщиков со схемами мягкой и жесткой брегмановской кластеризации, классическим алгоритмом K-средних и алгоритмом LBG векторного квантования.

Ключевые слова— машинное обучение, генеративная модель, автокодировщики, метод максимального правдоподобия, смеси экспертов, рецептивные поля.

1. ВВЕДЕНИЕ

На сегодняшний день методы машинного обучения, ориентированные на нейро-сетевые структуры глубокого обучения, достигли впечатляющих результатов, их широкое распространение воспринимается как старт эпохи искусственного интеллекта (ИИ) [1]. Среди наиболее впечатляющих применений следует отметить распознавание слитной речи, перевод текста между языками, распознавание рукописей, распознавание лиц на изображениях и т.д. Отметим, что реализованный в этих приложениях уровень ИИ на сегодняшний день вполне сопоставим с интеллектом живого человека. Профессиональное ИТ-сообщество, разделяя всеобщее воодушевление успехами ИИ, вместе с тем все чаще выражают некоторую озабоченность сложившимся положением дел. Это связано с тем, что до сих пор отсутствует ясное понимание того, почему ряд подходов, зачастую эвристических, оказался на практике существенно более эффективным, чем методы машинного обучения, разработанные до начала столетия.

Частичный ответ на этот вопрос может быть связан с главной особенностью методов машинного обучения – их ориентированностью на данные. Отсюда основные различия в характеристиках методов обусловлены, с одной стороны, спецификацией используемых обучающих данных, а, с другой – с выбором модели зависимости принимаемых по данным решений. Под спецификацией подразумевается наличие или отсутствие

в обучающих данных явной разметки (меток принимаемых решений). Это подразделяет методы машинного обучения на обучение контролируемое (с учителем) и неконтролируемое (без учителя). С точки же зрения используемой (статистической) модели, связывающей данные с разметкой, методы машинного обучения подразделяются на дискриминантное и генеративное, в зависимости от того, основана ли модель разметки на условном по отношению к данным вероятностном распределении меток или на совместном их распределении [2]. Машинное обучение знает примеры различных сочетаний обеих дихотомий – обучения с учителем / без учителя и дискриминантного / генеративного моделирования, что позволяет выделить в них, соответственно, четыре класса. Сравнений методов по отношению к приведенной классификации привело к тому, что суммарно, по совокупности характеристик наиболее успешными оказались подходы основанные на генеративных моделях обучения без учителя. К этим подходам, в частности, относятся генеративно-состязательные сети (GAN) [3], вариационные автокодировщики (VAE) [4], глубокие сети доверия (DBN) [5] и т.д.

2. МОДЕЛИРОВАНИЕ НА ОСНОВЕ ПРЕДСТАВЛЕНИЙ

Причина успешности генеративных моделей обучения без учителя не вполне ясна, однако, часто в качестве аргумента приводится тот факт, что они более последовательно моделируют механизмы естественного (человеческого / высших млекопитающих) интеллекта. Действительно, по мере развития перечисленных выше моделей [3-5], каждый новый этап разработки добавлял им новые функции / элементы, имитирующие, например, иерархическую архитектуру областей коры головного мозга, глубокое обучение с подкреплением, кратковременную память в корковых сетях, долговременную память о событиях (более подробно см. [6]). При этом, поскольку детали реализации информационных процессов в коре головного мозга ясны не до конца, соответствующие функции / элементы приходится моделировать на базе компьютерных структур *представления данных*. Эти представления необходимо должны быть достаточно гибкими и масштабируемыми, чтобы допускать многоуровневость, инвариантность, иерархические зависимости и пр., позволяя представлениям на более высоком уровне быть более абстрактными и инвариантными к вариациям, которые обычно присутствуют в обучающем (низкоуровневом) представлении, избегая при этом по возможности потери существенной информации [7]. В этой связи возникает интересный обратный вопрос – насколько выбор тех или иных представлений входных – промежуточных – выходных данных предопределяет синтез ожидаемых функций модели?

В данной работе, ориентируясь на генеративную модель обучения без учителя [2], применительно к задачам (авто)кодирования изображений, изложены результаты синтеза нового рекуррентного метода машинного обучения на основе представлений данных, сформированных по самим же изображениям.

3. АВТОКОДИРОВЩИК НА ОСНОВЕ ПРЕДСТАВЛЕНИЙ

В качестве представления входных данных использованы выборки контролируемого размера в k отсчетов, сэмплированных из кодируемых изображений (выборочные представления), детальное обсуждение которых можно найти в [8,9]. Коротко, выборочные представления состоят из k -наборов координат случайных отсчетов $X_k = (x_1, \dots, x_k)$, распределение которых по области изображения Ω задается плотностью распределения:

$$\rho(X_k|I(\vec{x})) = \prod_{j=1}^k \rho(\vec{x}_j|I(\vec{x})), \rho(\vec{x}_j|I(\vec{x})) = \frac{I(\vec{x})}{\iint_{\Omega} I(\vec{x}) ds} \quad (1)$$

где $I(\mathbf{x})$ – интенсивность формирующего изображение излучения, кратная, например, его битовой карте.

В рамках генеративной модели предлагаемого метода распределение вероятностей отдельного отсчета $\rho(x_j|I(\mathbf{x}))$ задается с помощью плотности распределения вероятностей из некоторого параметрического семейства смесей экспертов (смеси компонент) [10]:

$$\rho(\vec{x}|I(\vec{x})) = \sum_{z=1}^K w_z \eta(\vec{x} - \vec{\mu}_z; \vec{v}_z), \eta(\vec{x}; \vec{v}) = \exp\{\vec{v}\vec{x}^T - A(\vec{v})\} \rho_0(\vec{x}) \quad (2)$$

где K – число экспертов, $\{w_z\}$ – рейтинги экспертов, $\{v_z\}$ – экспертные оценки, задающие вид $\eta(\mathbf{x} - \boldsymbol{\mu}_z; \mathbf{v}_z)$ – локальных экспертных плотностей распределений \mathbf{x} . Каждое экспертное распределение задается как сдвинутый в точку $\boldsymbol{\mu}_z \in \Omega$ элемент параметрического экспоненциально-скошенного семейства плотностей $\{\eta(\mathbf{x} - \boldsymbol{\mu}_z; \mathbf{v})\}$. Выбранный вид (2) генеративной модели мотивирован концепцией рецептивных полей зрительной системы [11]. Именно, если под экспертами понимать ганглиозные клетки, а под $\boldsymbol{\mu}_z$ расположения контролируемых ими областей рецепторов сетчатки Ω , то $\rho(x_j|I(\mathbf{x}))$ задает модель зрительной нейронной активности. Данная ассоциация подсказывает вид $\rho_0(\vec{x})$ как симметричной плотности с компактным носителем Δ . При этом носителями экспертных плотностей будут Δ_z – смещенные в точки $\boldsymbol{\mu}_z$ компакты Δ , задающие систему (перекрывающихся) рецептивных полей.

Рассматривая параметры экспертов $\Theta = \{w_z, v_z\}$ в (2) как кодировку изображения (интенсивности $I(\mathbf{x})$), можно найти их оптимальные (наиболее вероятные) значения для представления X_k . Используя, например, метод максимального правдоподобия в отношении распределения отсчетов (1), можно получить, аналогично тому, как это делается при выводе рекуррентного EM-алгоритма [12], следующий метод оптимального кодирования:

Шаг I:

$$\rho_j^{(n+1)}(z) = \frac{w_z^{(n)} \exp\{-B_{A^*}(\vec{x}_j - \vec{\mu}_z, \vec{\zeta}_z^{(n)})\}}{\sum_{i \in \delta_{\vec{x}_j}} w_i^{(n)} \exp\{-B_{A^*}(\vec{x}_j - \vec{\mu}_i, \vec{\zeta}_i^{(n)})\}} \quad (3)$$

Шаг II:

$$w_z^{(n+1)} = \frac{1}{k} \sum_{\vec{x}_j \in \Delta_z} \rho_j^{(n+1)}(z), \quad (4)$$

$$\vec{\zeta}_z^{(n+1)} = \frac{\sum_{\vec{x}_j \in \Delta_z} \rho_j^{(n+1)}(z) (\vec{x}_j - \vec{\mu}_z)}{\sum_{\vec{x}_j \in \Delta_z} \rho_j^{(n+1)}(z)}$$

где $n = 1, 2, \dots$ – счетчик итераций, индексы $i \in \delta_{\vec{x}_j}$ экспертов задаются рецептивным окружением отсчета x_j , ζ_z – связанные преобразованием Лежандра с оценками v_z параметры средних, $B_{A^*}(\mathbf{x} - \boldsymbol{\mu}_z; \boldsymbol{\zeta}_z)$ – дивергенция Брегмана [13] для сопряженной $A(\mathbf{v})$ (2) по Лежандру функции $A^*(\boldsymbol{\zeta})$.

Следует отметить, что метод кодирования (3,4) с точностью до специфики, связанной с рецептивными полями, ограничивающей в (3) суммирование до δ_{x_j} , а в (4) до Δ_z , совпадает с алгоритмом мягкой брегмановской кластеризации, описанной в [13]. Более того, как показано там же, при переходе к жесткой брегмановской кластеризации и при соответствующем выборе функции $A^*(\boldsymbol{\zeta})$ получается целый спектр известных в области машинного обучения алгоритмов: классический алгоритм K -средних (K -means) [14], алгоритм LBG векторного квантования [15] и алгоритм теоретико-информационной кластеризации [13].

ЛИТЕРАТУРА

- [1] Sejnowski, T.J. The unreasonable effectiveness of deep learning in artificial intelligence. // Proceedings of the National Academy of Sciences. – 2020. – Vol. 117(48). – P. 30033-30038. DOI: 10.1073/pnas.1907373117.
- [2] Bishop, C.M. Generative or Discriminative? Getting the Best of Both Worlds / C.M. Bishop, J. Lasserre. // Bayesian Statistics 8, International Society for Bayesian Analysis. – Oxford University Press. – 2007. – P. 3-24.
- [3] Goodfellow, I. Generative Adversarial Networks. / I. Goodfellow // Communications of the ACM. – 2020. – Vol. 63(11). – P. 139-144. DOI: 10.1145/3422622.
- [4] Kingma, D.P. Auto-Encoding Variational Bayes. / D.P. Kingma, M. Welling // CoRR abs/1312.6114, 2014.
- [5] Hinton, G.E. A Fast-Learning Algorithm for Deep Belief Nets. / G.E. Hinton, S. Osindero, Y.-W. Teh // Neural computation. – 2006. – Vol. 18(7). – P. 1527-1554. DOI: 10.1162/neco.2006.18.7.1527.
- [6] Hassabis, D. Neuroscience-Inspired Artificial Intelligence / D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick // Neuron. – 2017. – Vol. 95(2). – P. 245-258. DOI:10.1016/j.neuron.2017.06.011.
- [7] Bengio, Y. Deep learning of representations for unsupervised and transfer learning // Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning workshop. – 2011. – Vol. 27. – P. 17-37.
- [8] Анциперов, В.Е. Статистическое представление изображений выборками случайных отсчетов фиксированного размера // Сб. трудов по материалам VII Межд. конф. и молодежной школы ИТНТ. – Самара: Самарский. нац. иссл. университет им. академика С.П. Королева. – 2021. – С. 24173.
- [9] Antsiperov, V.E. Image Coding by Count Sample, Motivated by the Mechanisms of Light Perception in the Visual System. / V.E. Antsiperov, V.A. Kershner. // Communications in Computer and Information Science. – 2022. – Vol 1534. DOI: 10.1007/978-3-030-96040-7_54.
- [10] Jacobs, R.A. Adaptive Mixtures of Local Experts / R.A. Jacobs, M.I. Jordan, S.J. Nowlan, G.E. Hinton // Neural Computation. – 1991. – Vol. 3(1). – P. 79-87. DOI: 10.1162/neco.1991.3.1.79.
- [11] Hyvärinen, A. Natural Image Statistics: Computational Imaging and Vision / A. Hyvärinen, J. Hurri, P.O. Hoyer. – London: Springer-Verlag, 2009. DOI: 10.1007/978-1-84882-491-1.
- [12] McLachlan, G.J. The EM Algorithm and Extensions / G.J. McLachlan, T. Krishnan. – New York: Wiley, 1997.
- [13] Banerjee, A. Clustering with Bregman Divergences / A. Banerjee, S. Merugu, I.S. Dhillon, J. Ghosh // Machine Learning Research. – 2005. – Vol. 6. – P. 1705-1749.
- [14] Lloyd, S. Least squares quantization in PCM // IEEE Transactions on Information Theory. – 1982. – Vol. 28(2). – P. 129-137. DOI: 10.1109/TIT.1982.1056489.
- [15] Linde, Y. An algorithm for vector quantizer design / Y. Linde, A. Buzo, R.M. Gray // IEEE Transactions on Communications. – 1980. – Vol. 28(1). – P. 84 95. DOI: 10.1109/TCOM.1980.1094577.

Разработка сервиса сегментации спутниковых данных с помощью алгоритмов глубокого обучения и библиотеки OpenVINO

Е.П. Васильев
Нижегородский Государственный
Университет им. Н.И. Лобачевского
Нижний Новгород, Россия
eugene.unn@gmail.com

А.А. Филатов
Нижегородский Государственный
Университет им. Н.И. Лобачевского
Нижний Новгород, Россия
dorfilfilatov@yandex.ru

И.С. Микерин
Нижегородский Государственный
Университет им. Н.И. Лобачевского
Нижний Новгород, Россия
mikerin-2002@mail.ru

Н.О. Ким
Нижегородский Государственный
Университет им. Н.И. Лобачевского
Нижний Новгород, Россия
ghj710@mail.ru

Д.А. Ермолаев
Нижегородский Государственный
Университет им. Н.И. Лобачевского
Нижний Новгород, Россия
ermolaev12danil@gmail.com

Аннотация—Целью работы является разработка сервиса семантической сегментации спутниковых снимков для областей лесных пожаров. В настоящее время идет активное развитие алгоритмов компьютерного зрения с применением глубокого обучения, что позволяет делать более точный и детализированный анализ спутниковых данных. С помощью алгоритмов глубокого обучения выполняется автоматическая сегментация природных пожаров, а также облаков на снимках Landsat-8. Разработан способ получения спутниковых данных для работы сервиса с портала EarthExplorer, для автоматической сегментации природных пожаров использована модель глубокого обучения на основе архитектуры UNet. Важной исследовательской составляющей данной работы является задача оптимизации производительности моделей глубокого обучения с помощью библиотеки инференса моделей глубокого обучения OpenVINO.

Ключевые слова— нейронные сети, семантическая сегментация, компьютерное зрение, обработка изображений, изображения поверхности Земли.

1. ВВЕДЕНИЕ

Использование моделей глубокого обучения для решения задачи семантической сегментации (задачи присвоения метки принадлежности к некоторому классу для каждого из пикселей изображения) стало широко используемой практикой в различных областях: в медицине для анализа рентгеновских снимков и данных компьютерной томографии [1], в анализе видео с видеорегистраторов [2], управлении роботизированными манипуляторами [3]. Остается актуальной тематика использования моделей глубокого обучения для сегментации спутниковых данных [4].

Как правило, различные спутниковые системы работают на разных высотах и оборудованы разными сенсорами, с разными рабочими длинами волн, поэтому алгоритмы определения пожаров подстраиваются разработчиками под конкретные спутники. Использование алгоритмов глубокого обучения позволяет получать более глубокое представление о данных, чтобы использовать один алгоритм для различных типов спутников с более высокой точностью распознавания.

В данной работе будут рассматриваться решения двух задач анализа данных дистанционного зондирования Земли (ДЗЗ): обнаружение природных пожаров на основе спутниковых данных Landsat-8, и автоматическая сегментация облаков.

А. Задача обнаружения природных пожаров с помощью алгоритмов глубокого обучения

Исследования в области обнаружения природных пожаров нацелены на разработку методов определения вероятности того, является ли конкретный пиксель мультиспектрального изображения изображением возгорания. В [5] была предложена архитектура сверточной нейронной сети для повышения разрешения изображения на снимках лесных пожаров по данным спутника Sentinel-2. Для построения собственных моделей глубокого обучения определения пожаров можно воспользоваться открытым набором данных activefire [6].

Б. Задача сегментации облаков на снимках ДЗЗ с помощью алгоритмов глубокого обучения

Обнаружение облаков на спутниковых изображениях является важным первым шагом во многих приложениях дистанционного зондирования. В [7] была использована архитектура ResNet, чтобы отличать дым от похожих объектов (облака, пыль, дымка) на 6225 спутниковых изображениях со спутников Aqua и Terra. Для создания собственных алгоритмов и моделей глубокого обучения можно использовать открытый набор данных 38-clouds-segmentation [8].

2. БИБЛИОТЕКА ИНФЕРЕНСА МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ OPENVINO

Библиотека Intel® Distribution of OpenVINO™ Toolkit разработана для повышения производительности моделей глубокого обучения при помощи оптимизации моделей встроенным оптимизатором и использованием доступного аппаратного ускорения вычислений [9]. Ускорение инференса моделей обеспечивается за счет анализа и оптимизации вычислительного графа, эффективного планирования обработки и векторизации данных, а также различных методов сжатия глубокой

модели. Библиотека Intel® Distribution of OpenVINO™ Toolkit ориентирована на разработку кросс-платформенных приложений для решения проблем компьютерного зрения и уделяет большое внимание оптимизации производительности при обработке изображений. OpenVINO имеет небольшое количество зависимостей, что облегчает интеграцию OpenVINO с существующим программным обеспечением.

3. РАЗРАБОТКА СЕРВИСА СЕГМЕНТАЦИИ СПУТНИКОВЫХ СНИМКОВ НА ОСНОВЕ ДАННЫХ LANDSAT-8

Данный сервис находится в стадии разработки. Исходный код всех модулей доступен на Github [10]. Сервис обладает модульной структурой, список модулей представлен ниже.

А. Выбор области получения данных

Для выбора области, на которой требуется сегментировать пожары, разработано графическое окно с внедренной картой `openstreetmap`, с помощью которой можно найти интересующую точку на карте. Для внедрения карты в приложение использована библиотека `leaflet`.

Б. Получение данных Landsat-8

Для получения снимков Landsat-8 можно воспользоваться Python пакетом `landsatxplore`, которые предоставляет интерфейс к portalу `EarthExplorer` для поиска и загрузки сцен Landsat Collections через интерфейс командной строки. Поиск данных в базе можно производить по координатам широты и долготы, временному промежутку, уровню облачности снимка.

В. Сегментация данных при помощи модели в формате OpenVINO

Сегментация данных в программе происходит с помощью модели, основанной на архитектуре UNet [4] (рисунок 1). Модели глубокого обучения являются вычислительно затратными, поэтому актуальным вопросом помимо качества работы является вопрос производительности решения.



Рис. 1. Снимок пожара со спутника Landsat-8 и результат сегментации [4]

Для измерения производительности используется модуль `DL Workbench`. Данный модуль позволяет проводить эксперименты по варьированию параметров запуска (количество потоков данных, обрабатываемых независимо; размер пачки данных, обрабатываемых одновременно; использование типа данных половинной точности). В таблице I представлены данные о производительности модели в различных режимах работы на CPU Intel i5 10600 (6 ядер 12 потоков 3.3GHz). Наилучшей производительностью обладает конфигурация, когда число параллельно работающих потоков равно числу физических ядер процессора, в

таком режиме каждое вычисления над одной пачкой не распараллеливаются на несколько ядер, а производятся на одном ядре.

Таблица I. ПРОИЗВОДИТЕЛЬНОСТЬ МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ С ПРИМЕНЕНИЕМ ФРЕЙМВОРКА OPENVINO

Размер пачки	Количество потоков	Тип весов	FPS, кадров в сек	Тип весов	FPS, кадров в сек
1	1	FP16	4.83	FP32	4.88
1	2	FP16	5.26	FP32	5.30
1	6	FP16	5.94	FP32	5.91
2	1	FP16	4.38	FP32	4.41
2	2	FP16	5.35	FP32	5.34
2	6	FP16	5.93	FP32	5.90
4	1	FP16	4.32	FP32	4.28
4	2	FP16	5.48	FP32	5.38
4	6	FP16	5.92	FP32	5.90

4. ЗАКЛЮЧЕНИЕ

В ходе исследования разработано программное приложение для автоматической сегментации природных пожаров и облаков по данным Landsat-8 с помощью инструмента `OpenVINO`. Отличительной особенностью приложения является использование моделей глубокого обучения для решения задачи анализа данных ДЗЗ. Приведены результаты анализа производительности в задаче. Исходный код приложения выложен в открытый доступ на Github [10].

БЛАГОДАРНОСТИ

Работа выполнена при поддержке компании Intel.

ЛИТЕРАТУРА

- [1] Lachinov, D.A. Glioma Segmentation with Cascaded UNet / D.A. Lachinov, E.P. Vasilyev, V.E. Turlapov // *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018* // LNCS. – 2019. – Vol. 11384. – P. 189-198.
- [2] Rethinking atrous convolution for semantic image segmentation [Electronic resource]. — Access mode: <https://arxiv.org/abs/1706.05587> (7.03.2022).
- [3] Sun, G. Robotic grasping using semantic segmentation and primitive geometric model based 3d pose estimation / G. Sun, H. Lin // *IEEE/SICE International Symposium on System Integration*. – 2020. – P. 337-342.
- [4] Pereira, G.H. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study / G.H. Pereira, A.M. Fusioka, N.B. Tomoyuki, R. Minetto // *ISPRS Journal of Photogrammetry and Remote Sensing*. – 2021. – Vol. 178. – P. 171-186.
- [5] Gargiulo, M. A CNN-Based Super-Resolution Technique for Active Fire Detection on Sentinel-2 Data / M. Gargiulo, D. Iodice, D. Riccio, G. Ruello // *Photonics Electromagnetics Research Symposium*. – 2019. – P. 418-426.
- [6] Active Fire Detection in Landsat-8 Imagery: a Large-Scale Dataset and a Deep-Learning Study [Electronic resource]. — Access mode: <https://github.com/pereira-gha/activefire> (7.03.2022).
- [7] Ba, R. SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention / R. Ba, C. Chen, J. Yuan // *Remote Sensing*. – 2019. – Vol. 11(14). – P. 1-22.
- [8] 38-Cloud: A Cloud Segmentation Dataset [Electronic resource]. — Access mode: <https://github.com/SorourMo/38-Cloud-A-Cloud-Segmentation-Dataset> (7.03.2022).
- [9] Intel® Distribution of OpenVINO™ toolkit [Electronic resource]. — Access mode: <https://docs.openvino.ai/latest/index.html> (7.03.2022).
- [10] Satellite images processing [Electronic resource]. — Access mode: https://github.com/itlab-vision/satellite_images_processing/ (7.03.2022).

Гиперспектральные и тепловые изображения в ранней диагностике засухи растений пшеницы с помощью ХАИ

М.А. Лысов

Нижегородский Государственный Университет им. Н.И.
Лобачевского
Нижегород, Россия
internal_horizon@mail.ru

В.Е. Турлапов

Нижегородский Государственный Университет им. Н.И.
Лобачевского
Нижегород, Россия
vadim.turlapov@itmm.unn.ru

Аннотация—Исследованы возможности вегетационных индексов (ВИ) гиперспектрального изображения (HSI) и теплового инфракрасного (TIR) в ранней диагностике засухи растений пшеницы. Рассмотрены 5 ВИ и TIR, для характеристики которых выбраны 5 статистических признаков их гистограмм: max, min, max-min, mean, std. На основе однослойного перцептрона (SLP) построен классификатор состояния растения, с помощью которого установлено, что признаки индекса NDVI достаточны для ранней диагностики засухи пшеницы. Построенный SLP классификатор обладает свойствами объяснимого ИИ.

Ключевые слова— обработка гиперспектральных изображений, ранняя диагностика, стресс засухи растений, объяснимый искусственный интеллект.

1. ВВЕДЕНИЕ

В интересах точного земледелия исследована задача ранней диагностики засухи растений по изображениям с гиперспектральных (HSI) и Thermal IR (TIR) сенсоров, которыми сегодня оснащаются квадрокоптеры. TIR-сенсоры выбраны с целью фиксации температуры листа, повышением которой является самым ранним признаком условия стресса. HSI использовались как источник множества вегетационных индексов, контролирующих наличие и состояние зеленой массы, влажности.

Наиболее эффективными методами решения задач точного земледелия являются сегодня методы искусственного интеллекта (ИИ) [1], [2], [3]. Однако, самыми актуальными свойствами применяемых сегодня моделей ИИ, становятся надежность и объяснимость решений ИИ, являющиеся основными признаками объяснимого ИИ (ХАИ). Объяснимость и надежность часто являются следствием обработки больших объемов данных [4],[5]. Но чаще, достаточно правильно преобразовать данные, чтобы построить объяснимый вектор признаков, а затем нейронную сеть, ориентированную на классификацию или регрессию по этим признакам. Исследовать построенную сеть на надежность решения. Что и является задачей данной работы.

2. ИСХОДНЫЕ ДАННЫЕ И НАТУРНЫЙ ЭКСПЕРИМЕНТ

Проведен эксперимент по наблюдению за стрессом засухи растений пшеницы в течение 25 дней с фиксацией состояния растений через 2-3 дня [6]. Наблюдались растения в 3 ящиках по 30 горшков с растениями в каждом: 15 слева поливались; 15 справа не поливались. Состояние растений регистрировалось с направления 90°

к поверхности с помощью трех камер (сенсоров): гиперспектральной (HSI) - Specim IQ (range: 400-1000 nm, spectral resolution: 7nm, channels: 204; 512x512 pix); тепловой ИК (TIR) - Thermal imager Testo 885-2; RGB высокого разрешения. Общий объем - 72ГБ. В ходе эксперимента зафиксированы разности неполиваемых и поливаемых растений в температуре (по изображениям TIR) и в потере воды (%), через взвешивание растений), в 1, 3, 6, 8, 10, 12, 14, 16, 19, 22, 25 дни эксперимента. Зафиксированы следующие пороговые изменения состояния растений по сравнению с контрольными: 1) через 5 дней - повышение средней температуры растений на 0,2 градуса; 2) через 11 - начало потери воды растением (около 8% объема воды). Первый признак, подъем температуры растения, является самым ранним свидетельством стресса засухи, проходящего еще без потери воды и видимых изменений зеленой массы. Детектирование стресса растения до наступления потери воды соответствует критерию «раннее» детектирование.

3. ОБРАБОТКА ИЗОБРАЖЕНИЙ ДЛЯ РАННЕЙ ДИАГНОСТИКИ ЗАСУХИ РАСТЕНИЙ

Для решения задачи ранней диагностики стресса в полевых условиях исследована эффективность следующих вегетационных индексов: 1) NDVI (normalized difference vegetation index); 2) GNDVI (green NDVI); 3) GCI (green chlorophyll vegetation index); 4) SIPI (structure insensitive pigment vegetation index); 5)BLUE. В качестве шестого индекса присоединено также TIR-изображение. Следуя [6], при обучении модели на вход сети вместо изображений, решено подавать их статистические признаки: {max, min, max-min, mean, std}. Вычисления выполнялись по маске растений, на основе NDVI. На Рис.1 показаны интересные примеры поведения признаков:

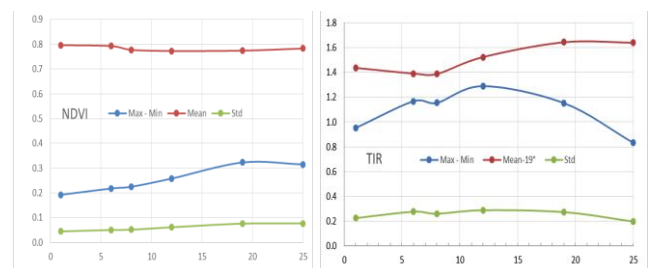


Рис. 1. Поведение признаков по дням эксперимента: а)NDVI и б)TIR

Для выполнения исследования эффективности классификации стресса засухи, по указанным выше индексам, использован классификатор на основе

полносвязной сети однослойного перцептрона (SLP_c). Число входов классификатора в итоге равно 6х5, число нейронов на скрытом слое выбрано равным числу исследуемых индексов, т.е. 6. В качестве классов на выходе установлены 6 дней, ключевых для раннего детектирования признаков засухи и полноты охвата срока эксперимента {1, 6, 8, 12, 19, 25}.

На классификаторе SLP_c исследована эффективность классификации указанных выше 6 дней для каждого признака. Обучение выполнялось на 30% пикселей под маской растений (маски построены автоматически по NDVI изображениям), а тестирование качества обучения, на оставшихся 70%. Для обучения использованы TIR изображений под маской растений в дни регистрации.

Продолжительность обучения на Intel Core i3-8130U, 2.2ГГц, 4 ядра, 4 Гб для каждому признаку порядка 1 секунды. Точность (Accuracy) оценена отдельно: по первым четырем дням {1,6,8,12}; по всем шести дням. Интересные лидерские величины показаны в Табл.1.

Таблица 1. ЭФФЕКТИВНОСТЬ ОБЕСПЕЧИВАЕМАЯ ОТДЕЛЬНЫМ ПРИЗНАКОМ

Признак	Точность (Accuracy)	
	Дни: 1, 6, 8, 12	Дни: 1, 6, 8, 12 + 19, 25
NDVI Min	0,75	0,75
NDVI Max – Min	1,0	0,83
NDVI Mean	0,625	0,58
NDVI Std	0,875	0,75
TIR Max – Min	0,75	0,66
BLUE Min	0,625	0,75
SIPI Max	0,75	0,66
SIPI Min	0,75	0,66
SIPI Max – Min	0,75	0,66

Обращает на себя внимание высокая эффективность признаков NDVI даже каждого в отдельности. Особенно NDVI Max – Min. Затем также на SLP_c, реализующей использование смеси индексов, исследована эффективность пар из наиболее эффективных признаков. При совместном использовании NDVI Max–Min и Std мы получаем точность 1.0 на всех днях классификации, благодаря монотонному поведению NDVI Std. Тот же эффект достигается при сочетании признаков TIR Max–Min и NDVI Std. Обе пары признаков имеют ясный биологический смысл. В итоге, в качестве классификатора ключевых дней ранней диагностики засухи пшеницы в полевых условиях удалось построить простую полносвязную нейронную сеть на принципах объяснимого искусственного интеллекта. Такую сеть можно построить на данных, как только одного HSI сенсора, так и на данных двух сенсоров HSI+TIR или TIR+Blue(RGB).

Наиболее близким к нашему state-of-the-art исследованием является [7]. В нем также для HSI, но без TIR, проведено исследование 9 индексов на основе HSI, а также средних сигнатур HSI-пикселей растений в качестве образов для каждого из дней испытаний {0,3,6,9,12,15,18,21,24} и производных от сигнатур. Отслежен также трек засухи в условиях обводненного орошения (T25). В качестве методов классификации использовались методы: RF, SVM, MLP (Multilayer Perceptron, с числом слоев от 1 до 5). Исследованием установлено, что часть индексов показывает себя на

ранних днях (до 9), другая - на поздних, но ни один из индексов не может быть использован для качественной диагностики во все дни и лучший, но не идеальный, результат дает смесь индексов. Результат близкий к идеальному дает использование MLP с кривыми средних HSI-сигнатур и их производных входов, обозначенные в [7] как DNN-Full, DNN-Deriv. Утверждается однако, что из-за использования гиперспектров как признаков необходимо обучение классификатора для каждого типа растений и, возможно, условий полива.

4. ЗАКЛЮЧЕНИЕ

Исследованы возможности решения задачи ранней диагностики засухи растений пшеницы по данным гиперспектральных (HSI) и тепловых (TIR) изображений в интересах точного земледелия. В качестве инструмента исследования на основе Single Layer Perceptron построен исследовательский XAI классификатор SLP_c ключевых состояний: норма; начало повышения температуры листа без потери воды (6 день); начало потери воды (12 день); визуально выраженное высыхание (19, 25 дни). На входе классификатор принимает вместо изображения вектор признаков из 30 статистических параметров, по 5 для каждого индексного изображения, включая (max-min). Благодаря (max-min) индекс NDVI оказался наилучшим из индексов по эффективности в полевых условиях и достаточным для детектирования всех 6 ключевых дней с точностью 100% независимо от типа растений и конкретных условий засухи. Время обучения на Intel Core i3-8130U, 2.2ГГц, 4 ядра, 4 Гб порядка 1 сек. Архитектура XAI классификатора, может быть оптимизирована под тип используемого сенсора и индекса.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации, соглашение № 075-15-2020-808.

ЛИТЕРАТУРА

- [1] Jha, K. A comprehensive review on automation in agriculture using artificial intelligence / K. Jha, A. Doshi, P. Patel, M. Shah // Artificial Intelligence in Agriculture. – 2019. – Vol. 2. - P. 1-12. DOI 10.1016/j.aiaa.2019.05.004.
- [2] Talaviya, T. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides / T. Talaviya / Artificial Intelligence in Agriculture. – 2020. – Vol. 4. – P. 58-73. DOI 10.1016/j.aiaa.2020.04.002.
- [3] Pathan, M. Artificial cognition for applications in smart agriculture: A comprehensive review / M. Pathan // Artificial Intelligence in Agriculture. – 2020. – Vol. 4. – P. 81-95. DOI 10.1016/j.aiaa.2020.06.001.
- [4] Gorban, A.N. Blessing of dimensionality: mathematical foundations of the statistical physics of data / A.N. Gorban, I.Y. Tyukin // Philos. Trans. Royal Soc. A. – 2118. – Vol. 376. – P. 20170237.
- [5] Gorban, A.N. One-trial correction of legacy AI systems and stochastic separation theorems/ A.N. Gorban, R. Burton, I. Romanenko, I.Y. Tyukin // Information Sciences. – 2019. – Vol. 484. – P. 237-254.
- [6] Maximova, I. Study of xai-capabilities for early diagnosis of plant drought / I. Maximova, E. Vasiliev, A. Getmanskaya, D. Kior, V. Sukhov, V. Vodeneev, V. Turlapov / IJCNN: International Joint Conference on Neural Networks, INNS, IEEE. – Shenzhen, China, 2021.
- [7] Dao, Ph.D. Plant drought impact detection using ultra-high spatial resolution hyperspectral images and machine learning / Ph.D. Dao // International Journal of Applied Earth Observation and Geoinformation. – 2021. – Vol. 102(8). – P. 102364. DOI: 10.1016/j.jag.2021.102364.



Министерство образования
и науки Самарской области



САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

Самарский университет

ИСОИ  **ИПСИ**

Институт систем обработки изображений РАН



Institute of Electrical
and Electronics Engineers



Photonics



ФОТОНИКА



Компьютерная оптика