Method for frame removal detection in static camera surveillance video

Bavrina A.Y.

IPSI RAS - branch of the FSRC "Crystallography andPhotonics" RAS,
Samara university

Objective of the study:

- New method for passive protection of a surveillance camera video from a video fragment deletion attack

The essence of the method:

- The method is based on the construction of local features for the samples of two consecutive frames, followed by a multilayer neural network classification. Post-processing and calculation of statistics based on the results of the classification helps to decide whether a given pair of frames is sequential or a number of frames were cut between them

Benefits:

- Experiments show the efficiency of detection the fact of a fragment removal even from stationary scenes, when such a deletion is visually imperceptible

Frame removal detection based on interframe local feature analysis

- feature calculation

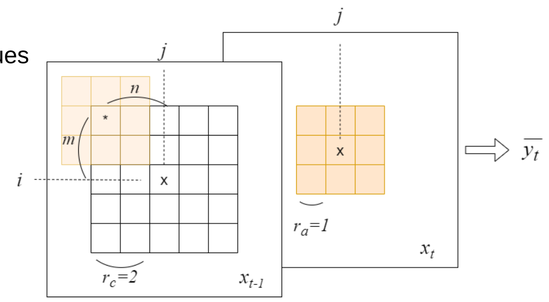
for every sample of the current frame $x_t(i, j)$

$$\overline{y}_t(i,j) = \{ f(x_t(i-k, j-l), x_{t-1}(i-m-k, j-n-l)) \}$$

$$f(x_t, x_{t-1}) = \sum_{c=1..3} (x_t^c - x_{t-1}^c)^2$$ - the distance between pixel values

$$m, n \in [-r_c, r_c] \qquad k, l \in [-r_a, r_a]$$

feature vector length – $n_f = (2r_c+1)^2 \cdot (2r_a+1)^2$



- MNN classification

pixel-wise classification
2 inner layers (of size $2n_f+1$ and n_f), *softmax* AF
classes: “cut” and “sequential”
output is the confidence value for “cut” class
for each pair of sequential frames “**confidence image**” is formed

- post-processing

median filtering of the confidence image

- final decision

“**confidence value**” calculation – the average of the confidence image samples
if confidence value is grater then a threshold, then frame deletion in the current position exists

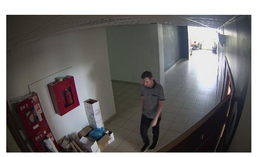
Test video set

Test video parameters

	Road	**SU_corridor**
frame size	1080x1920	1080x1920
fps	100	25
codec	H264-MPEG AVC (part 10)	H264-MPEG AVC (part 10)
GOP	GOP=24 (I,P,B-frames)	GOP=50 (I,P-frames)
video length	1 video more than 1 hour long	5 videos 1-2 minutes long
video particularities	Outdoor camera. Small camera jitter and illumination changes.	Indoor camera. Includes natural and artificial illumination scenes.
source	J. Sochor, R. Juranek, J. Spanhel, L. Marsik, A. Siroky, A. Herout, and P. Zemcik, “Comprehensive Data Set for Automatic Single Camera Visual Speed Measurement,” IEEE Transactions on Intelligent Transportation Systems, vol. 20, pp. 1633-1643, 2019.	Falcon Eye FE-IPC-BV2-50pa

Examples of test video frames

«Road» video



«SU_corridor» video

Acknowledgements: The work was funded by the RFBR grant #19-29-09045

Experiments

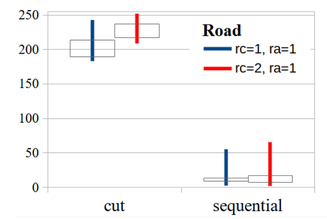
sequential frames

Confidence image for sequential frames

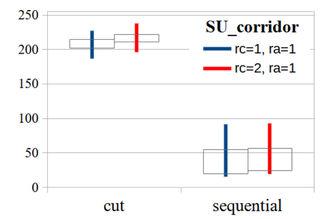
nonsequential frames

Confidence image for nonsequential frames

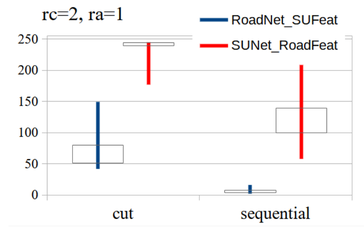
Statistics for confidence value



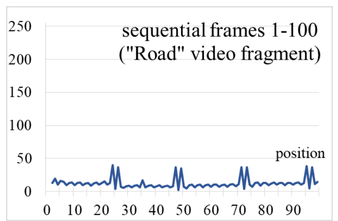
Statistics of the confidence value for the «Road» test video. Good separabiliyt of classes «cut» and «sequential».



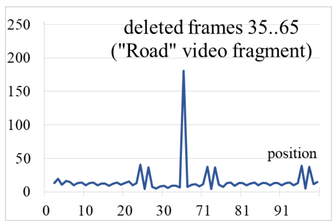
Statistics of the confidence value for the «SU_corridor» test video set. Good separabiliyt of classes «cut» and «sequential».



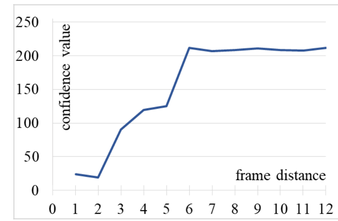
The results of applying the net, tuned for one video, to another video. Poor separabiliyt of classes «cut» and «sequential».



Variation of the confidence value for the «Road» test video fragment. Sequential frames 1..100.

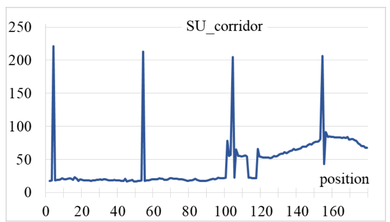


Variation of the confidence value for the «Road» test video fragment. Frames 35-65 were cut from the video fragment.



The dependence of the confidence value on the distanse between frames («Road» video fragment). The removal even a few frames will be detected.

Restrictions of the method



Variation of the confidence value for the «SU_corridor» test video fragment. Sequential frames 1..179. False frame deletion detection for every I-frame.

Conclusions:

- The proposed method contributes to the construction of a universal data protection system in the area where many algorithms configured to track the movement of objects do not work correctly (skipping the frame removal in video with the little or no movement).
- The proposed way of local feature construction makes it possible to take into account both the movement of objects and "small structures", specific to the particular video type (caused by the combination of camera parameters, compression method, and environment).
- Limitations of the method were also revealed, caused by both the unsatisfactory video quality (camera and compression parameters) and the behavior of objects in the frame (overlapping one of the light sources).