

# Spatio-Temporal Slices for Frame Cut Detection in Video

N A Sorokina<sup>1</sup> and V A Fedoseev<sup>1,2</sup>

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34, Samara, Russia, 443086

<sup>2</sup>Image Processing Systems Institute - Branch of the Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Molodogvardeyskaya str. 151, Samara, Russia, 443001

**Abstract.** The paper proposes an approach for unauthorized inter-frame video change detection using spatio-temporal slices. This approach can significantly reduce the amount of data processed and replace video processing with image processing that could be performed much faster. To test the efficiency of this approach, we consider a simple algorithm in analyzing adjacent rows of a slice and then classifying the rows based on its result. Experimental studies have revealed that this algorithm shows moderate results in terms of quality, but it has a great potential for improvement, which confirms the prospects of spatio-temporal slices for the given problem.

## 1. Introduction

### 1.1. Problem statement

Today digital video plays an increasingly important role in the society. In 2015, according to the Sandvine report [1], the percentage of video and audio in North American traffic exceeded 70%. In addition, according to the Ericsson report [2], by 2019 the percentage of video in mobile traffic should exceed 50% (and already now it is above 40%). The reasons for that is not only entertainment industry development, but also the growing market for video surveillance systems (up to 20% per annum, according to the Markets and Markets analysts' report [3]), and their widespread introduction to both large business structures and small companies. As a consequence, the data received by surveillance systems are increasingly used in investigative activities or as evidence in legal proceedings. For this reason, such data must be reliably protected from unauthorized alteration by intruders.

One of the most common methods of unauthorized video alteration is inter-frame modification, which includes the removal of video fragments or their replacement with copies of other ones. Such changes can remove crime evidence or data about movements of persons or vehicles that are important in a particular context. In video signals obtained from a stationary camera, such changes can be practically invisible. In the case of a moving camera, such changes can also be hard to detect when an intruder cuts or replaces short-term fragments.

In this paper, we consider the problem of detecting inter-frame artificial changes in video signals taken from a moving or stationary camera. The detection method should work well with video signals stored in various formats, and also combine high detection accuracy with high speed.

### 1.2. Review of related studies

In practice, the detection of artificial changes in video can be carried out using digital forensics methods developed since the second half of the 2000s [4-6]. The main achievements in this direction are associated with H. Farid, A.C. Popescu, S. Prasad, J. Fridrich, A. Piva, M. Barni. So,

the latest two in the review paper [7] classified methods of digital video forensics onto the following groups:

- 1) camera-based methods that analyze various video artifacts to determine the optical system of the camera;
- 2) coding-based methods that identify artifacts resulting from encoding video using certain codecs;
- 3) geometry, or physics-based methods that detect violations in the physical or geometric parameters of the objects observed;
- 4) pixel-based methods based on detecting changes at the pixel level of the video.

Examples of algorithms aimed to detect inter-frame modifications can be found, in particular, in [8-14]. Most of them are coding-based methods [8-11] and geometry / physics-based [12-14].

The coding-based methods [8-11] are based on the properties of certain video formats (usually different versions of MPEG) and assume the separation of frames into different types (P-frames, I-frames, etc.). Therefore, these methods do not satisfy the claimed universality requirement regarding the data format. In addition, a significant number of such methods (in particular, [8-10]) may be used to determine the fact of video change, but do not allow to determine the exact location of the changes (in the time domain).

As for the geometry / physics-based methods, many of them (in particular, [12-13]) are based on the use of an optical flow to track changes frame by frame. This technology works well for the case of a stationary camera, but for a moving camera, it needs to take into account camera movements which can be unknown. Moreover, the methods based on optical flow do not provide a high computational performance. Another method from this group [14] utilizes another approach. As a vivid example, this paper considers a simple scene with moving balls. To detect the removal of a video fragment, the method first tracks ball positions, and then detects physically unjustified deviations of the found trajectories, which is possible evidence of unauthorized changes. The main drawback of this method is the need to solve a complex problem of tracking several moving objects. However, the basic idea of this method is rather attractive: to detect unauthorized alterations, we can try to analyze video data over a long time interval.

In this paper, we test a method based on the same idea of detecting deviations in time. However, as the analyzed data, we propose to use the so-called spatio-temporal slices of video images that are slices of the video data cube along the time axis and one of the spatial axes (for example, the horizontal one). If you build several such horizontal slices of video at a certain vertical interval, the resulting set of images will give quite enough information about object movements, although the data will have much less volume comparing with the original video. Moreover, for the processing of these data, we can use computationally effective image processing methods, in particular, parallel-recursive FIR filters [15, 16].

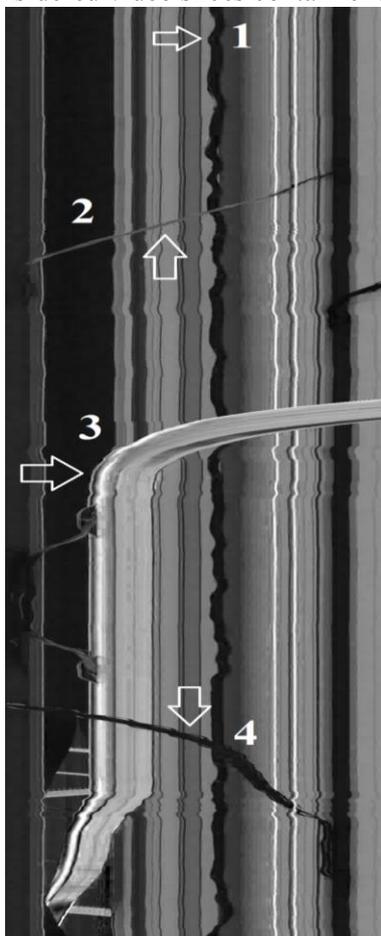
In papers [17, 18] a similar approach is used for road object detection in the problem of autonomous navigation., The algorithms [17, 18] also process not all the pixels of each frame, but only horizontal lines spaced from each other at equal intervals. This allows the authors to ensure the solution of the problem with satisfactory accuracy in real time.

The paper is organized as follows. Section 2 illustrates the traces of natural events in spatio-temporal slices and outlines the principles for detecting inter-frame video changes with their help. Section 3 describes a simple cut detection method based on these principles. Finally, Section 4 describes the experimental studies.

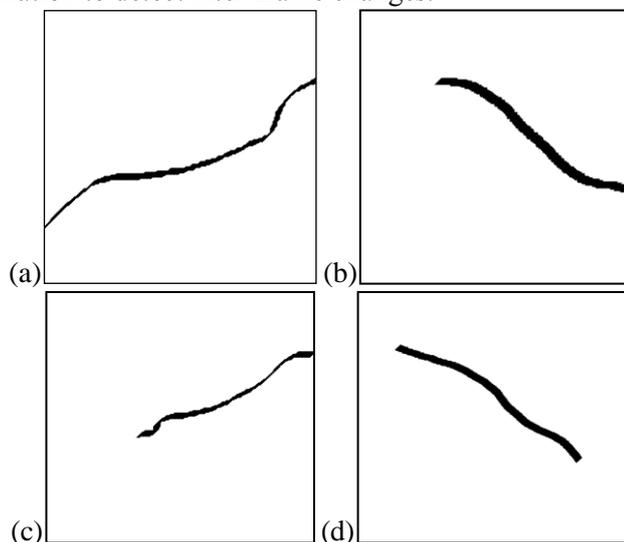
## 2. Object movements at spatio-temporal slices

Figure 1 shows an example of a spatio-temporal slice of a video obtained by a stationary camera. In this figure, the following events occurred are marked with numbers: (1) – movements by the hands of a standing person, (2) – movements of a person who emerged from one door and entered another, (3) – appearance and stopping of a car, (4) - the appearance of a person from the left border of the frame and their movement. As the figure shows, these events lead to the appearance of smooth curves on the cuts characterizing object movements. You can also note the local background shifts due to camera deviations or fluctuations of observable objects (trees, advertising signs, etc.). Figure 2 illustrates four typical types of object movements observed in spatio-temporal slices.

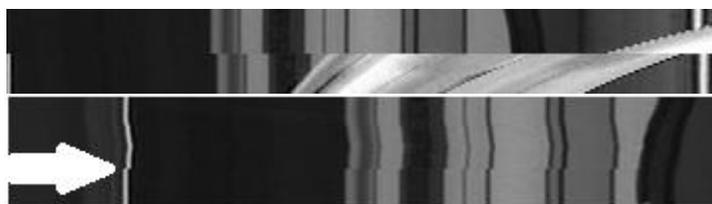
In case of unauthorized video alteration in the time domain, video slices can clearly visualize them by horizontal offsets, as shown in Figure 3. These offsets can appear over the entire frame width. Furthermore, if we consider a set of slices made at different vertical positions, such offsets will be observed in the same lines of the slices corresponding to the same time shift. Thus, we assume that the considered video slices contain enough information to detect inter-frame changes.



**Figure 1.** A spatio-temporal slice of a video containing several events.



**Figure 2.** Four types of object movement in spatio-temporal slices: (a) an object enters the camera view and goes beyond it, (b) an object appears in the camera view and goes beyond it, (c) an object enters the camera view and disappears, (d) an object appears and disappears within the camera view.



**Figure 3.** Examples of video fragment deletion shown in spatio-temporal slices.

### 3. Cut detection method

As noted in Section 2, the evidence for unnatural inter-frame changes is sharp horizontal shifts in the slice image (see Figure 3). To detect them, we can estimate the displacement of each next row relative to the previous one. Formally speaking, we are aimed to find the shift  $\Delta_k$  of the  $(k+1)$ -th row relative to the  $k$ -th row, which would provide the smallest error between neighboring samples:

$$\Delta_k = \arg \min_{\delta} \varepsilon_k(\delta), \quad (1)$$

where  $\delta$  is an integer argument characterizing the line shift, and  $\varepsilon_k(\delta)$  is the error function estimated from the equation:

$$\varepsilon_k(\delta) = \sum_{j=\max(0, -\delta)}^{\min(W-1-\delta, W-1)} (I(k, j) - I(k+1, j+\delta))^2. \quad (2)$$

In (2),  $I(k, j)$  is the grayscale value of the slice image at  $k$ -th row and  $j$ -th column,  $W$  is the frame width. We can numerically solve problem (1)-(2) using the correlation approach.

Next, we can use the obtained  $\Delta_k$  and  $\varepsilon_k(\Delta_k)$  (for simplicity, we will sign the latest one as  $\varepsilon_k$ , i.e. without the argument) to detect artificial changes based on the following assumptions. Low values of  $\Delta_k$  and  $\varepsilon_k$  indicate unaltered videos, while sharp leaps in either  $\Delta_k$  or  $\varepsilon_k$  may give evidence of artificial changes (see Fig. 4-5).

To detect artificial changes using  $\Delta_k$  and  $\varepsilon_k$ , we used the algorithm based on supervised learning. It classified slice image rows into two classes: “Cut” and “Non-cut”. To train the algorithm, we used the following features of the rows:

$$p_1 = |\Delta_k|, \quad (3)$$

$$p_2 = \varepsilon_k, \quad (4)$$

$$p_3 = \min(|\Delta_k - \Delta_{k-1}|, |\Delta_{k+1} - \Delta_k|), \quad (5)$$

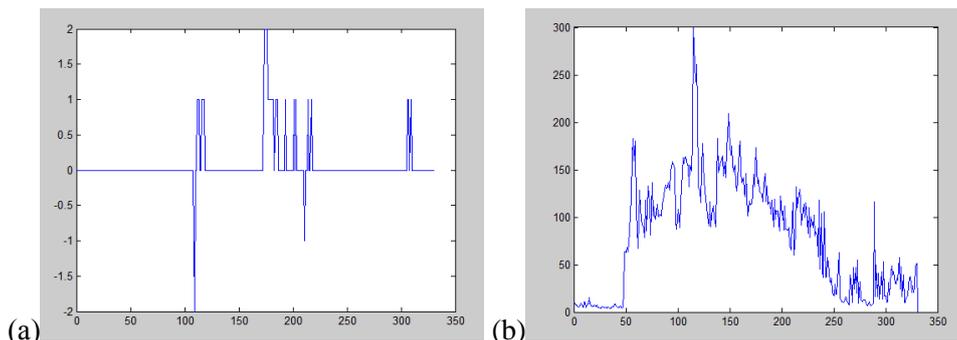
$$p_4 = \min(|\varepsilon_k - \varepsilon_{k-1}|, |\varepsilon_{k+1} - \varepsilon_k|), \quad (6)$$

$$p_5 = |\Delta_k - \text{med}(\Delta_k)|, \quad (7)$$

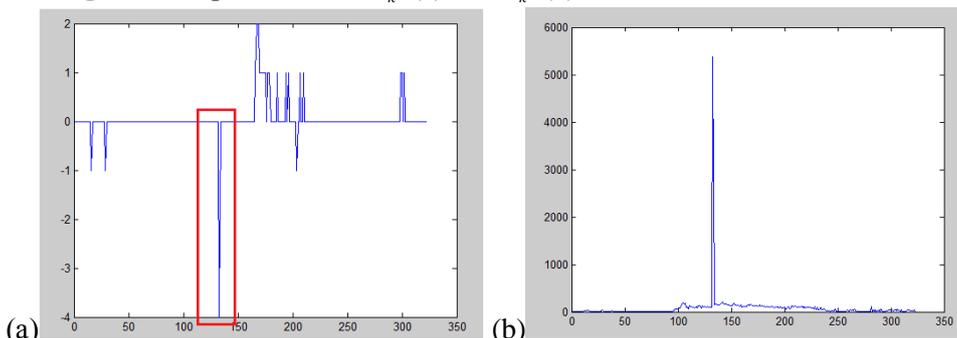
$$p_6 = |\varepsilon_k - \text{med}(\varepsilon_k)|. \quad (8)$$

We should note especially that we calculated the features (3)-(8) in the local maxima only. Function  $\text{med}(x)$  in equations (7)-(8) means the median of the 4-point neighborhood of pixel  $x$  excluding  $x$  itself.

To speed up the algorithm, we classified the only rows corresponding to the local maxima of  $p_3$  and  $p_4$ .



**Figure 4.** Dependence of  $\Delta_k$  (a) and  $\varepsilon_k$  (b) on  $k$  for an unaltered video.



**Figure 5.** Dependence of  $\Delta_k$  (a) and  $\varepsilon_k$  (b) on  $k$  for a video with a cut at frame 130.

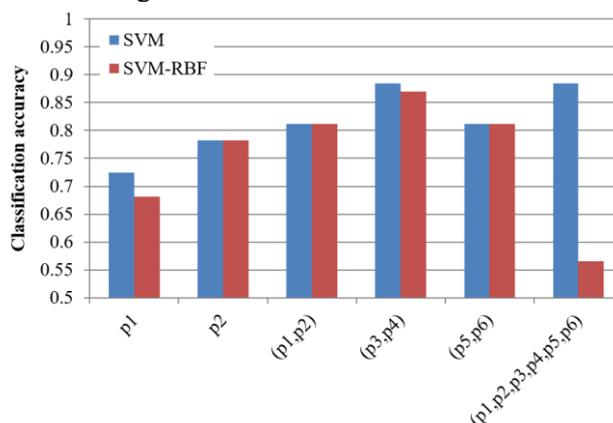
#### 4. Experimental investigations

To test the proposed method, we used two types of video: DVR recordings (made from a moving car), and model records (made mainly by a stationary camera and containing typical pedestrian movements).

To conduct the experiments, we made spatio-temporal slices of the source videos, and then divide them onto 100-line fragments. A half of these fragments were obtained from 100 consecutive frames, while the other half was composite and contained a frame cut in the 50th line. The length of the gap in

the frames was a parameter of the experiment. Then for each image, we calculated feature sets. 70% of the data obtained was used as a training set, whereas the remaining 30% was a test sample. When testing the method, it did not use the information about the line of the gap. The experimental studies were carried out in two stages. The first one was aimed to select the most appropriate feature set and classifier model. We considered two models: linear SVM and non-linear SVM with radial basis function. At the second stage, we investigated the algorithm performance for various gap lengths and for different types of video. In addition, at the second stage, we analyzed the efficiency of combining data from different lines of several slices, which was made by summarizing lines.

The first stage of the experiments was carried out on DVR recordings with the gap length of 30 and without slice summation. The results of this stage in terms of classification accuracy (equal to the fraction of the correct classifications) are given in Figure 6. The diagram in Figure 6 shows that the best accuracy values are resulted from the use of the  $(p_3, p_4)$  feature set and linear SVM. Therefore, these options were further used in the second stage (see results in Table 1). The obtained results show that slice summation improves the classification quality sufficiently. We may also notice that the algorithm works better on DVR videos which contain a rapidly changing background. In general, the final results allow us to conclude that the proposed method is able to solve the considered problem, even in the simplified version of the algorithm described in Section 2.



**Figure 6.** Results of the second stage of the experiment: classification model and feature set selection.

**Table 1.** Results of the second stage of the experiment: accuracy estimation for different video types.

| Video Type | Cut Length | Slice Summation | Cut Detection Accuracy |
|------------|------------|-----------------|------------------------|
| DVR        | 10         | +               | 0.8406                 |
|            |            | -               | 0.8400                 |
|            | 30         | +               | 0.9063                 |
|            |            | -               | 0.8696                 |
|            | 60         | +               | 0.9375                 |
|            |            | -               | 0.8732                 |
| Model      | 10         | +               | 0.8732                 |
|            |            | -               | 0.7654                 |
|            | 30         | +               | 0.7971                 |
|            |            | -               | 0.7200                 |
|            | 60         | +               | 0.8485                 |
|            |            | -               | 0.7308                 |

## 5. Conclusion

In this paper, we have tested the approach based on spatio-temporal video slices to solve the problem of detecting unauthorized inter-frame changes in video. This method is theoretically capable to solve this problem with a high speed due to the processing of only a part of the video signal, as well as using

fast image processing techniques. To test the efficiency of this approach, we proposed a simple algorithm for detecting inter-frame changes and performed some numerical experiments. Our studies showed that the algorithm provides an accuracy of not less than 0.8 and works better for video captured with a moving camera. The results of the studies allow us to conclude that the method of spatio-temporal image slices looks promising, but the algorithm should be significantly improved in terms of increasing the accuracy at short gap lengths and when using video from a stationary camera.

## 6. References

- [1] Anon *CNW Sandvine: Over 70% of North American traffic is now streaming video and audio* (Access mode: <http://www.newswire.ca/news-releases/sandvine-over-70-of-north-american-traffic-is-now-streaming-video-and-audio-560769981.html>)
- [2] Anon *Internet Video Streaming to Dominate Mobile Data Traffic by 2019 - ISPreview UK* (Access mode: <http://www.ispreview.co.uk/index.php/2014/06/internet-video-streaming-dominate-mobile-data-traffic-2019.html>)
- [3] Anon *Video Surveillance Market by Applications & Management Services 2015* MarketsandMarkets (Access mode: <http://www.marketsandmarkets.com/Market-Reports/surveillance-277.html>)
- [4] Poisel R and Tjoa S 2011 Forensics Investigations of Multimedia Data: A Review of the State-of-the-Art *Sixth International Conference on IT Security Incident Management and IT Forensics* 48-61
- [5] Rocha A, Scheirer W, Boulton T and Goldenstein S 2011 Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics *ACM Comput. Surv.* **43** 26:1-26:42
- [6] Gashnikov M V, Glumov N I, Kuznetsov A V, Mitekin V A, Myasnikov V V and Sergeev V V 2016 Hyperspectral remote sensing data compression and protection *Computer Optics* **40(5)** 689-712 DOI: 10.18287/2412-6179-2016-40-5-689-712
- [7] Bestagini P, Fontani K M, Milani S, Barni M, Piva A, Tagliasacchi M and Tubaro K S 2012 An overview on video forensics *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)* 1229-1233
- [8] Shanableh T 2013 Detection of frame deletion for digital video forensics *Digital Investigation* **10** 350-360
- [9] Stamm M C, Lin W S and Liu K J R 2012 Temporal Forensics and Anti-Forensics for Motion Compensated Video *IEEE Transactions on Information Forensics and Security* **7** 1315-1329
- [10] Gironi A, Fontani M, Bianchi T, Piva A and Barni M 2014 A video forensic technique for detecting frame deletion and insertion *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 6226-6230
- [11] Wu Y, Jiang X, Sun T and Wang W 2014 Exposing video inter-frame forgery based on velocity field consistency *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2674-2678
- [12] Chao J, Jiang X and Sun T 2013 A Novel Video Inter-frame Forgery Model Detection Scheme Based on Optical Flow Consistency *The International Workshop on Digital Forensics and Watermarking* 267-281
- [13] Wang W and Farid H 2006 Exposing Digital Forgeries in Video by Detecting Double MPEG Compression *Proceedings of the 8th Workshop on Multimedia and Security* (New York, NY, USA: ACM) 37-47
- [14] Zhang J, Su Y and Zhang M 2009 Exposing Digital Video Forgery by Ghost Shadow Artifact *Proceedings of the First ACM Workshop on Multimedia in Forensics* 49-54
- [15] Glumov N I, Myasnikov V V and Sergeev V V 1996 Parallel-recursive local image processing and polynomial bases *Proceedings of Third International Conference on Electronics, Circuits, and Systems* **2** 696-699

- [16] Myasnikov V V 2007 Fast algorithm for recursive computation of the convolution of an image with a two-dimensional inseparable polynomial FIR filter *Pattern Recognit. Image Anal.* **17** 421-427
- [17] Kiy K I and Dickmanns E D 2004 A color vision system for real-time analysis of road scenes *IEEE Intelligent Vehicles Symposium* 54-59
- [18] Kiy K I 2015 A New Real-Time Method of Contextual Image Description and Its Application in Robot Navigation and Intelligent Control *Computer Vision in Control Systems* **2** 109-133

### **Acknowledgments**

This work was supported by the Russian Foundation for Basic Research (grants 16-29-09494, 16-41-630676), by the Ministry of Education and Science (grant MK-1907.2017.9), and by the Federal Agency for Scientific Organizations (Agreement 007-GZ/43363/26).