

# Обнаружение межкадровых изменений видео с использованием пространственно-временных срезов

Н.А. Сорокина<sup>1</sup>, В.А. Федосеев<sup>1,2</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. Академика С.П. Королева, Московское шоссе 34, Самара, Россия, 443086

<sup>2</sup>Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

**Аннотация.** В работе рассмотрен подход к обнаружению несанкционированных межкадровых изменений видео на основе построения пространственно-временных срезов. Данный подход позволяет существенно снизить объём обрабатываемых данных и свести обработку видео к обработке одного или нескольких изображений. Для проверки работоспособности данного подхода рассмотрен простой алгоритм обнаружения изменений, сводящийся к анализу соседних строк среза и последующей классификации строк на основе полученных в результате этого анализа признаков. В ходе экспериментальных исследований данный алгоритм не показал выдающихся результатов в смысле качества обнаружения (точность обнаружения колеблется в зависимости от типа видео и длины разрыва кадров и составляет не ниже 0.8), однако он имеет большой потенциал для улучшений, что подтверждает перспективность использования пространственно-временных срезов для решения данной задачи.

## 1. Введение

### 1.1. Постановка задачи

В последнее время цифровые видеоданные играют всё возрастающую роль в жизни общества. Так, уже в 2015 году согласно отчёту сетевой компании Sandvine [1] совокупная доля видео и аудио в общем объёме интернет-трафика США превысила 70%. Кроме того, согласно отчёту компании Ericsson [2], к 2019 году доля видео во всём мобильном трафике должна превысить 50% (и уже сейчас она выше 40%). Связано это не только с развитием индустрии развлечений, но и со стабильно растущим рынком систем видеонаблюдения (до 20% в год, согласно отчёту аналитиков компании Markets and Markets [3]) и их повсеместным внедрением не только крупными бизнес-структурами и органами власти (в целях контроля офисов, улиц крупных городов, важных промышленных и культурных объектов), но и небольшими частными компаниями (для мониторинга внутренних помещений и прилегающих территорий). Как следствие, данные, полученные системами наблюдения, всё чаще используются при проведении следственных действий или в качестве доказательства при судебных разбирательствах. По этой причине такие данные должны надёжно защищаться от несанкционированного изменения злоумышленниками.

Одними из наиболее распространённых методов несанкционированной модификации видеoinформации являются временные (межкадровые) изменения, к которым относятся удаление фрагментов видео или их замена копиями других фрагментов (с возможными изменениями). Эти методы могут применяться для удаления свидетельств преступления или данных о перемещении каких-либо лиц или транспортных средств, которые представляют важность в определённом контексте. На видеосигналах, полученных с неподвижной камеры, подобные изменения могут быть практически незаметны. Однако также они могут применяться и при съёмке движущейся камерой в случае вырезания (или замены) кратковременных фрагментов.

В настоящей работе решается задача обнаружения межкадровых искусственных изменений видео, снятого при помощи подвижной или неподвижной камеры, в условиях отсутствия оригинального (неискажённого) видеосигнала. При этом метод обнаружения должен работать с видеосигналами, представленными в различных форматах, а также сочетать высокую точность обнаружения с высоким быстродействием.

### *1.2. Обзор существующих решений*

На практике обнаружение искусственных изменений видео может осуществляться при помощи методов пассивной защиты, активно разрабатывающихся начиная со второй половины 2000-х гг. Основные достижения в данной области связаны с именами Н. Farid, А.С. Popescu, S. Prasad, J. Fridrich, А. Piva, М. Barni. Так, двое последних в обзорной работе [4] выделяют следующие основные группы методов обнаружения искусственных искажений видео:

- 1) camera-based: методы, использующие различные артефакты видео, определяющие оптическую систему устройства, которым производилась съёмка;
- 2) coding-based: методы, идентифицирующие артефакты, возникающие в результате кодирования или перекодирования видео с использованием определённых кодеков;
- 3) geometry/physics-based: методы, обнаруживающие нарушения в физических или геометрических параметрах объектов на сцене;
- 4) pixel-based: методы, основанные на обнаружении изменений на уровне пикселей видео.

Примеры алгоритмов, предназначенных для обнаружения изменений во временной области, можно найти, в частности, в работах [5-11]. Преимущественно они относятся к группам coding-based [5-8] и geometry/physics-based [9-11].

Методы группы coding-based [5-8] основаны на использовании определённых форматов видео (как правило, MPEG различных версий) и предполагают разделение кадров на различные типы (Р-кадры, I-кадры и пр.). Поэтому эти методы не удовлетворяют заявленному требованию универсальности относительно формата данных. Кроме того, значительная часть подобных методов (в частности, [5-7]) могут использоваться для определения факта изменения видео, но не позволяют установить точное местоположение правок (во времени).

Что касается методов группы geometry/physics-based, то многие из них (в частности, [9-10]) основаны на использовании оптического потока для отслеживания изменений по кадрам. Эта технология хорошо работает для случая неподвижной камеры, но при подвижной камере возникают трудности, связанные с необходимостью учёта движения камеры, характер которого неизвестен. Кроме того, использование подобных методов не обеспечивает высокую скорость работы. В работе [11] предложен иной метод. Для примера в ней рассматривается простая сцена с движущимися шариками. Для обнаружения удаления фрагмента видео строятся траектории движения всех шариков, являющиеся результатом отслеживания их позиций, и в случае обнаружения физически необоснованных отклонений траекторий принимается решение о наличии несанкционированных изменений. Главный недостаток данного метода – необходимость решения сложной задачи технического зрения по отслеживанию траекторий нескольких движущихся объектов. Однако идеологически он базируется на довольно привлекательной идее – обнаружении неестественных отклонений при анализе данных на продолжительном временном интервале.

В настоящей работе тестируется метод, основанный на той же идее обнаружения отклонений во времени, однако в качестве анализируемых данных принято решение

использовать так называемые пространственно-временные срезы видео – изображения, являющиеся срезом куба данных видеосигнала вдоль временной оси и одной из пространственных осей (для определённости – горизонтальной). Если построить несколько подобных срезов видео с определённым интервалом по вертикальной оси, то полученный набор изображений даст достаточно полную информацию о смещении объектов, но при этом он будет иметь куда меньший объём в сравнении с исходным видео, а для его обработки можно будет использовать вычислительно эффективные методы обработки изображений, в частности, параллельно-рекурсивными КИХ-фильтрами [12]. В работах [13, 14] для обнаружения дорожных объектов в задаче автономной навигации также осуществляется обработка не всех пикселей каждого кадра, а лишь горизонтальных линий, отстоящих друг от друга на равных интервалах. Это позволяет обеспечить решение поставленной задачи с удовлетворительной точностью в режиме реального времени.

Работа организована следующим образом. В разделе 2 иллюстрируется представление естественных событий на пространственно-временных срезах и излагаются принципы обнаружения межкадровых изменений видео с их помощью. В разделе 3 описывается простой метод анализа срезов, а раздел 4 посвящён описанию экспериментальных исследований. Завершают работу заключение и список литературы.

## 2. Движение объектов на срезах видео

На рисунке 1 представлен пример пространственно-временного среза видео, снятого при помощи малоподвижной камеры (она установлена на штативе, но может дрожать от порывов ветра). На данном рисунке цифрами отмечены следующие события: 1 – движения руками стоящего человека, 2 – перемещения человека, вышедшего из одной двери и зашедшего в другую, 3 – появление и остановка в кадре автомобиля, 4 – появление человека из-за левой границы кадра и перемещения. Как видно на рисунке, перечисленные события приводят к появлению на срезе плавных линий, характеризующих смещение объектов. Также можно отметить локальные смещения фона, возникающие из-за дрожания камеры или наблюдаемых объектов (деревьев, рекламных вывесок и пр.). На рисунке 2 отмечены четыре типа перемещений объектов, наблюдаемых на срезах видео.

В случае несанкционированной модификации видео во временной области на срезах появляются чётко различимые горизонтальные смещения, как показано на рисунке 3. Причём эти смещения могут проявляться на всей ширине кадра, а если рассматривать набор срезов, сделанных на разной высоте, то такие смещения будут наблюдаться в одной и той же строке среза, соответствующей одному и тому же моменту времени. Таким образом, можно сделать предположение, что рассмотренные срезы видео содержат достаточно информации для обнаружения межкадровых изменений.

## 3. Метод обнаружения межкадровых изменений

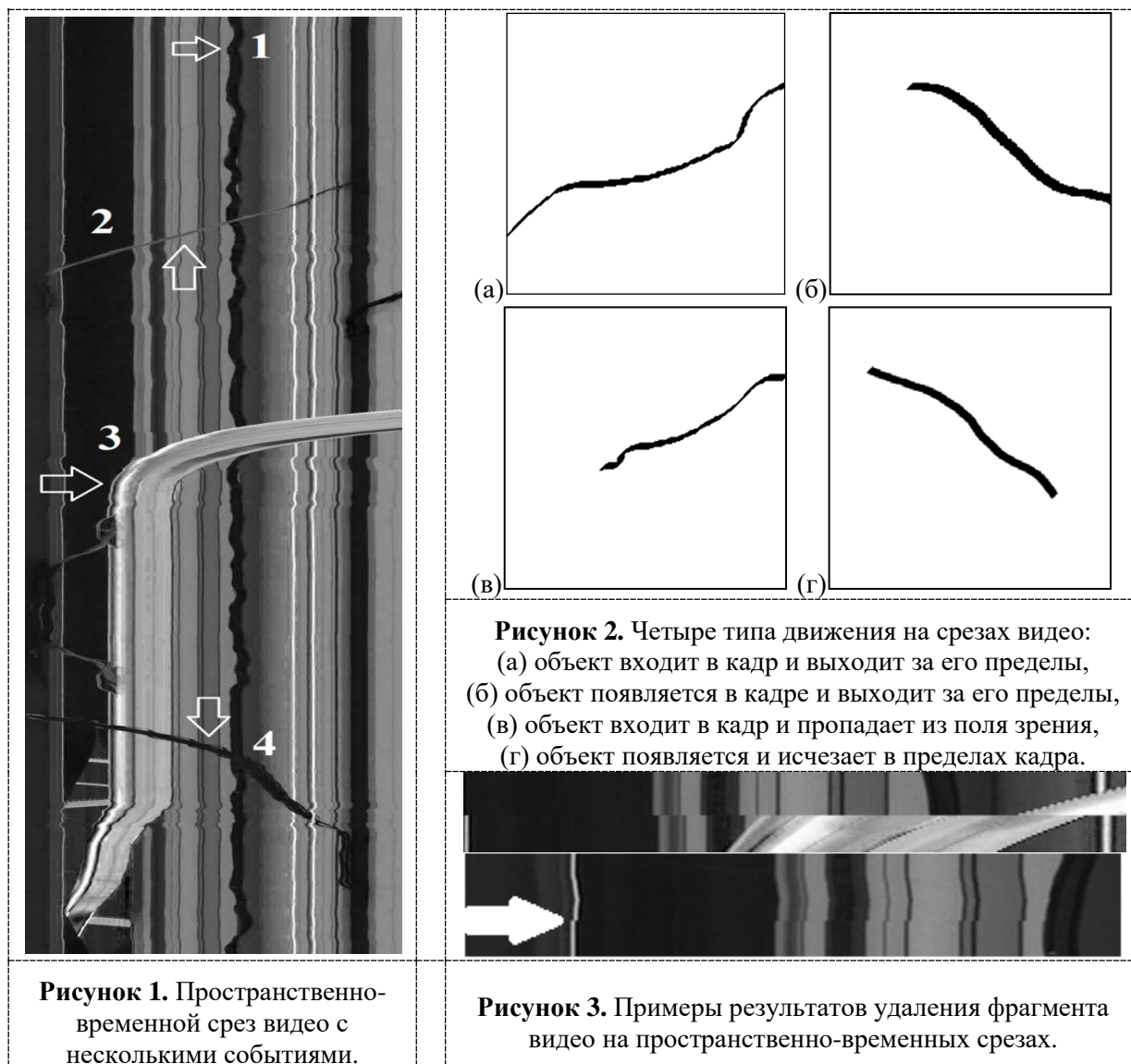
Как отмечалось в разделе 2, свидетельством неестественных изменений во временной области являются резкие горизонтальные сдвиги на срезах, не имеющие временной продолжительности (см. рисунок 3). Для обнаружения подобных сдвигов можно оценить смещение каждой следующей строки среза относительно предыдущей, то есть найти такой сдвиг  $\Delta_k$  ( $k+1$ )-й строки относительно  $k$ -й, который бы привел к наименьшей ошибке между соседними отсчетами:

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

где  $\delta$  – целочисленный аргумент, характеризующий сдвиг строк, а  $\varepsilon_k(\delta)$  – функция ошибки, которую можно оценить по формуле

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

В выражении (2)  $I(k, j)$  – это яркость изображения среза в  $k$ -й строке и  $j$ -м столбце,  $W$  – ширина кадра. Искать решение задачи (1)-(2) можно при помощи коррелятора.



Далее полученные величины  $\Delta_k$  и  $\varepsilon_k(\Delta_k)$  (последнюю для простоты будем обозначать просто  $\varepsilon_k$  – без аргумента) могут использоваться для обнаружения искусственных изменений. При их отсутствии  $\Delta_k$  и  $\varepsilon_k$  малы, а при наличии на графиках зависимости одной или обеих из этих величин от номера строки  $k$  присутствуют скачки, не коррелирующие с окрестными значениями (см. рисунки 4-5).

Для обнаружения несанкционированных изменений на основе данных  $\Delta_k$  и  $\varepsilon_k$  использовался алгоритм, осуществляющий классификацию строк среза на два класса «монтаж» и «не монтаж» на основе обучения. При этом в качестве признаков использовались значения

$$p_1 = |\Delta_k|, \tag{3}$$

$$p_2 = \varepsilon_k, \tag{4}$$

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

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

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

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

рассчитываемые в точках локальных максимумов. Функция  $med(x)$ , используемая в формулах (7)-(8) означает взятие медианы по окрестности из 4-х соседних с  $x$  точек, исключая саму точку  $x$ .

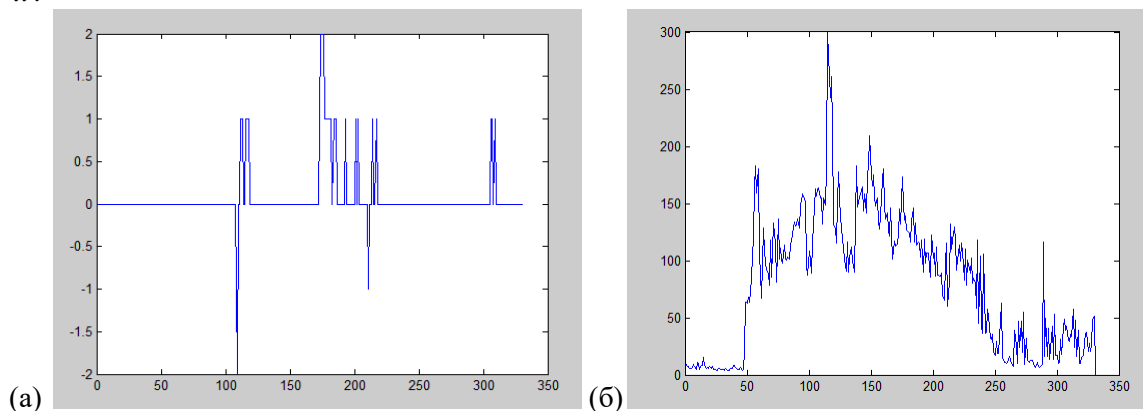


Рисунок 4. Графики  $\Delta_k$  (а) и  $\varepsilon_k$  (б) в зависимости от  $k$  для видео без изменений.

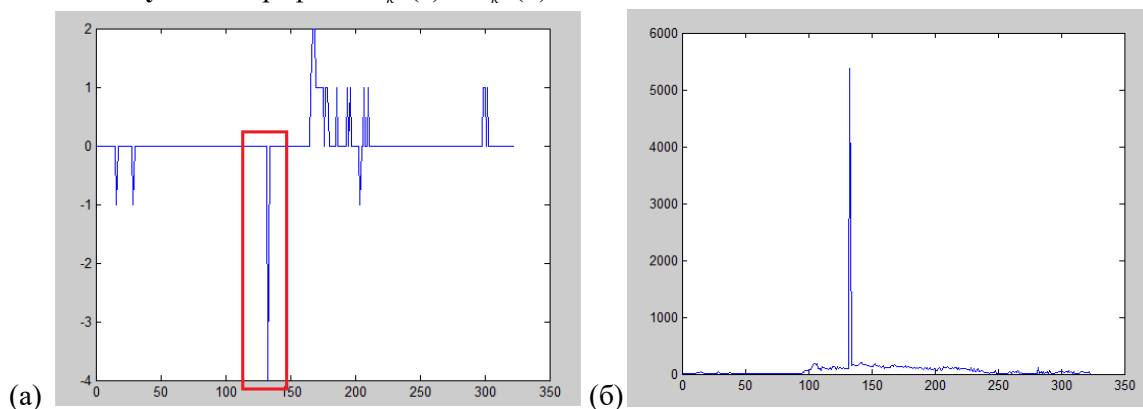


Рисунок 5. Графики  $\Delta_k$  (а) и  $\varepsilon_k$  (б) в зависимости от  $k$  для видео с монтажом в кадре 130.

Для повышения быстродействия классификация осуществлялась не для всех строк, а только для строк  $k$ , соответствующих локальным максимумам  $p_3$  и  $p_4$ .

#### 4. Экспериментальные исследования

Для апробации алгоритма были использованы видеосигналы двух различных типов: записи видеорежиссера и модельные записи, снятые в помещении главным образом неподвижной камерой и содержащие типичные примеры движения объектов (перечисленные на рисунке 2). Далее мы будем обозначать эти типы видео соответственно «регистратор» и «моделирование».

Для дальнейших экспериментов по исходным сигналам строились пространственно-временные срезы, из которых нарезались тестовые фрагменты длиной 100 строк (кадров). Половина таких фрагментов была получена из 100 последовательных кадров, другая половина была составная и содержала разрыв в 50-й строке. Длина разрыва в кадрах являлась параметром эксперимента. При апробации алгоритма информация о предполагаемой строке разрыва не использовалась. На каждом из полученных изображений осуществлялся расчёт признаков. Далее по 70% полученного набора данных происходило обучение, а 30% использовалось в качестве тестовой выборки.

Экспериментальные исследования проводились в два этапа. Целью первого являлся выбор наилучшего набора признаков и модели классификатора (рассматривались две модели – линейный классификатор SVM и нелинейный классификатор SVM с радиальной базисной функцией). На втором этапе исследовалась работа алгоритма при различных длинах разрыва и на различных типах видео. Кроме того, на втором этапе исследовалась эффективность

объединения данных от нескольких срезов посредством их суммирования при решении рассматриваемой задачи.

Первый этап экспериментов проводился на видео типа «регистратор» при длине разрыва 30 и без суммирования срезов. Результаты данного этапа в виде диаграммы значений точности классификации (равной доле правильных классификаций) отражены на рисунке 6. Как видно из диаграммы, к наилучшим результатам привело использование набора признаков ( $p_3, p_4$ ) и линейного классификатора SVM. Этот вариант использовался на втором этапе экспериментов, результаты которых, отражённые в таблице 1, показывают целесообразность процедуры суммирования срезов, а также лучшую работу алгоритма на видео «регистратор», содержащем быстро меняющийся фон. В целом результаты экспериментов позволяют сделать вывод о работоспособности метода обнаружения изменений во временной области на основе пространственно-временных срезов даже в упрощённом варианте алгоритма, описанном в разделе 2.

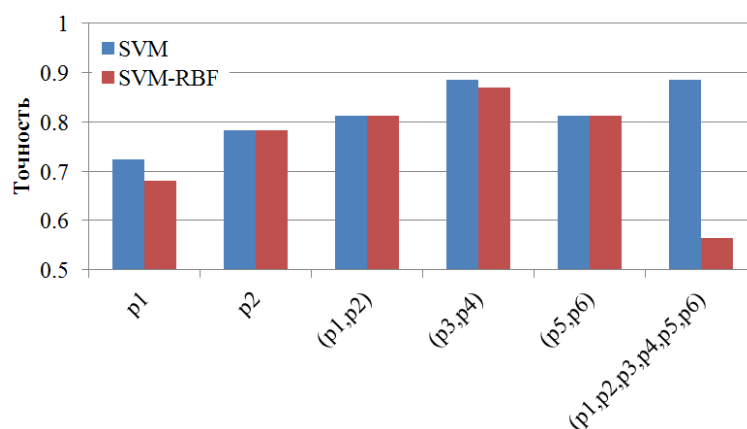


Рисунок 6. Результаты первого этапа экспериментов.

Таблица 1. Результаты второго этапа экспериментов.

| Тип видео     | Длина разрыва | Суммирование срезов | Точность обнаружения разрывов |
|---------------|---------------|---------------------|-------------------------------|
| Регистратор   | 10            | +                   | 0.8406                        |
|               |               | -                   | 0.84                          |
|               | 30            | +                   | 0.9063                        |
|               |               | -                   | 0.8696                        |
|               | 60            | +                   | 0.9375                        |
|               |               | -                   | 0.8732                        |
| Моделирование | 10            | +                   | 0.8732                        |
|               |               | -                   | 0.7654                        |
|               | 30            | +                   | 0.7971                        |
|               |               | -                   | 0.72                          |
|               | 60            | +                   | 0.8485                        |
|               |               | -                   | 0.7308                        |

## 5. Заключение

В работе рассмотрен подход к решению задачи обнаружения несанкционированных межкадровых изменений видео на основе построения пространственно-временных срезов, который теоретически способен позволить решать данную задачу с высоким быстродействием за счёт обработки лишь части видеосигнала, а также за счёт использования быстрых процедур

обработки изображений. Для проверки работоспособности данного подхода апробирован алгоритм обнаружения межкадровых изменений, сводящийся к попарной корреляции строк среза и последующей классификации на основе полученных признаков. Экспериментальные исследования показали, что данный алгоритм обеспечивает точность не ниже 0.8 и лучше проявляет себя на видео, снятом при помощи движущейся камеры. Результаты исследований позволяют сделать вывод о целесообразности использования пространственно-временных срезов для обнаружения изменений в видео, однако алгоритм должен быть существенно улучшен в части повышения точности при малых длинах разрывов и при использовании видео со стоячей камеры.

## 6. Благодарности

Работа выполнена при поддержке Федерального агентства научных организаций (соглашение № 007-ГЗ/ЧЗ363/26), Минобрнауки РФ в рамках гранта президента РФ МК-1907.2017.9, а также РФФИ (гранты 16-29-09494, 16-41-630676).

## 7. Литература

- [1] Sandvine: Over 70% of North American traffic is now streaming video and audio [Electronic resource]. – Access mode: <http://www.newswire.ca/news-releases/sandvine-over-70-of-north-american-traffic-is-now-streaming-video-and-audio-560769981.html>.
- [2] Internet Video Streaming to Dominate Mobile Data Traffic by 2019 [Electronic resource]. – Access mode: <http://www.ispreview.co.uk/index.php/2014/06/internet-video-streaming-dominate-mobile-data-traffic-2019.html>.
- [3] Global Video Surveillance Market, Applications and Management Services Forecasts (2010-2015) [Electronic resource]. – Access mode: <http://www.marketsandmarkets.com/Market-Reports/surveillance-277.html>.
- [4] Milani, S. An overview on video forensics / S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva // *APSIPA Transactions on Signal and Information Processing*. – 2012. – Vol. 1.
- [5] Shanableh, T. Detection of frame deletion for digital video forensics / T. Shanableh // *Digital Investigation*. – 2013. – Vol. 10(4). – P. 350-360.
- [6] Wang, W. Exposing Digital Forgeries in Video by Detecting Double MPEG Compression / W. Wang, H. Farid // *Proceedings of the 8th Workshop on Multimedia and Security*. – 2006. – P. 37-47.
- [7] Stamm, M.C Temporal Forensics and Anti-Forensics for Motion Compensated Video / M.C. Stamm, W.S. Lin, K.J.R. Liu // *IEEE Transactions on Information Forensics and Security*. – 2012. – Vol. 7(4). – P. 1315-1329.
- [8] Gironi, A. A video forensic technique for detecting frame deletion and insertion / A. Gironi, M. Fontani, T. Bianchi, A. Piva, M. Barni // *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. – 2014. – P. 6226-6230.
- [9] Wu, Y. Exposing video inter-frame forgery based on velocity field consistency / Y. Wu, X. Jiang, T. Sun, W. Wang // *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. – 2014. – P. 2674-2678.
- [10] Chao, J. A Novel Video Inter-frame Forgery Model Detection Scheme Based on Optical Flow Consistency / J. Chao, X. Jiang, T. Sun // *The International Workshop on Digital Forensics and Watermarking 2012*. – 2013. – P. 267-281.
- [11] Zhang, J. Exposing digital video forgery by ghost shadow artefact / J. Zhang, Y. Su, M. Zhang // *Proc. First ACM Workshop on Multimedia in Forensics, 2009*.
- [12] Методы компьютерной обработки изображений / под ред. В.А. Сойфера. – Изд. 2-е, испр. – М.: Физматлит, 2003.
- [13] Kiy, K.I. A color vision system for real-time analysis of road scenes / K.I. Kiy, E.D. Dickmanns // *2004 IEEE Intelligent Vehicles Symposium*. – 2004. – P. 54-59.
- [14] Kiy, K.I. A New Real-Time Method of Contextual Image Description and Its Application in Robot Navigation and Intelligent Control / K.I. Kiy // *Computer Vision in Control Systems-2. Intelligent Systems Reference Library*. – 2015. – Vol. 75.

# Spatio-Temporal Image Slices for Frame Cut Detection in Video

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

<sup>1</sup>Samara National Research University, Moskovskoye shosse, 34, Samara, Russia, 443086

<sup>2</sup>Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 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 by 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.

**Keywords:** video analysis, forgery detection, spatio-temporal analysis.