



# Интеллектуальный видеоанализ опасных ситуаций



Леся АНИЩЕНКО  
Lesya N. ANISHCHENKO

Сергей ИВАШОВ  
Sergey I. IVASHOV



Алексей СКРЕБКОВ  
Alexey V. SKREBKOV

*Анищенко Леся Николаевна – кандидат технических наук, старший научный сотрудник МГТУ им. Н. Э. Баумана, Москва, Россия.*  
*Ивашов Сергей Иванович – кандидат технических наук, начальник лаборатории МГТУ им. Н. Э. Баумана, Москва, Россия.*  
*Скребков Алексей Валентинович – кандидат технических наук, доцент Российского университета транспорта (МИИТ), Москва, Россия.*

## Intelligent Video Analysis of Dangerous Situations

(текст статьи на англ. яз. –  
English text of the article – p. 189)

**Статья посвящена разработке системы интеллектуального анализа видеозаписей камер наружного наблюдения, позволяющей выявлять опасные ситуации на объектах железных дорог на примере детекции падений в зоне пути. Предложен метод предобработки видеоряда с целью формирования пространства признаков, основанный на использовании вычитания фона по методу гауссовой смеси, последующем отслеживании перемещения человека при помощи фильтра Калмана и деформации формы подвижного объекта в результате применения прокрустового анализа. Обоснован подбор оптимального состава пространства признаков и дополнительных эвристик, обеспечивающих выделение эпизодов падений по видеозаписи со средним качеством каппы Козна 0,62 по сравнению с визуальным анализом оператором.**

*Ключевые слова:* железная дорога, безопасность, видеонаблюдение, интеллектуальный видеоанализ, распознавание движений, машинное обучение, анализ формы.

**З**адача выявления потенциально опасных лиц на объектах транспорта осуществляется по данным, полученным при помощи камер наружного наблюдения. Разработчики систем видеонаблюдения, в частности, предлагают комплексы с использованием биометрической идентификации (например, система NeoFace, очки R7) [1, 2], а также распознавания эмоций по мимике (например, DeepFace) [3–6]. Основным недостатком такого типа систем является то, что информация о нарушителе может не содержаться в существующей базе данных.

В распоряжении аналитиков есть алгоритмы, позволяющие как отслеживать забытые вещи на перронах [7], так и выявлять объекты на пути следования поезда [8], а также обнаруживать возникновение задымлений [9], однако в подавляющем большинстве случаев данные с камер видеонаблюдения применяются лишь для формирования архивов видеозаписей, а возможности зафиксированной информации с удалённых объектов практически

не используются. Это обусловлено прежде всего техническими трудностями, связанными с особенностями интеллектуальных систем видеонаблюдения, а именно, чувствительностью анализатора к условиям освещённости, наличию вибрации, которая неизбежна для подвижного состава на железнодорожном транспорте, и тому подобным фоновым зависимостям.

Отдельно следует выделить задачу обеспечения безопасности на перронах малозагруженных станций пригородного сообщения, особенно в ночное время. При этом в случае возникновения внештатной ситуации оказавшемуся в беде человеку может оказаться некому помочь по причине отсутствия других пассажиров. В связи с чем возрастает актуальность использования технических средств, способных в автоматизированном режиме подать оповещающий сигнал дежурному по станции. В рамках этой статьи рассмотрен вопрос о такого рода системе для распознавания движений человека по записям камер видеонаблюдения и выявления эпизодов его падений в зоне риска.

## ЭКСПЕРИМЕНТАЛЬНЫЕ ЗАПИСИ

В большинстве работ по выявлению эпизодов падений при помощи анализа видеоизображений оценка эффективности предложенных алгоритмов классификации оказывается искусственно завышена вследствие ограниченности используемой в качестве тестирующей и обучающей выборок базы данных видеозаписей. При этом записи получены для одних и тех же состояний обстановки (чаще всего в условиях лаборатории, а не приближенных к реальности) при наличии равномерной освещённости анализируемой области пространства. В роли испытуемых, выполняющих в том числе и артефакты движения типа «падение», выступает один и тот же человек, причём падения однообразны как по виду самого движения и предшествующим ему действиям, так и по тому, под каким углом расположен «исполнитель» по отношению к камере в момент падения. Кроме того, надо отметить, что почти всегда падения выполняются на амортизирующий мат, имеющий контрастный цвет с одеждой испытуемого.

В нашем анализе использована открытая база данных видеозаписей лаборатории электроники и обработки изображений Национального центра научных исследований в г. Шалон-сюр-Сон (Франция) [12]. К её достоинствам стоит отнести следующие факторы:

1. Видеоизображения получены для различных состояний окружающей обстановки.

2. Присутствует неравномерность освещённости сцены проведения эксперимента, включая ситуацию, когда вследствие ограниченности динамического диапазона камеры и наличия в кадре области высокой яркости (область окна) видеоизображение человека имело малый контраст по сравнению с обстановкой.

3. В экспериментах принимали участие четыре испытуемых (3 мужчины и 1 женщина).

4. Падения тел происходили под разными углами обзора, причём как из положения стоя, так и из положения сидя.

5. Падения выполнялись на специально подготовленное амортизирующее основание и непосредственно на пол.

Авторами были проанализированы 108 записей из указанной базы данных, 84 из которых содержали единичный эпизод падения. Для каждой из записей оператором визуально были определены номера кадров начала и окончания эпизода. Продолжительность артефактов типа «падение» составила  $22 \pm 9$  кадров. Учитывая, что видеозаписи сделаны для частоты дискретизации 25 или 30 кадров в секунду, получаем продолжительность артефакта, равную  $0,7 \pm 0,3$  с.

## АЛГОРИТМ ОБРАБОТКИ ВИДЕОРЕЯДА

Задача классификации в компьютерном зрении разделяется на две подзадачи: предобработку изображения и классификацию. Этап предобработки необходим для преобразования визуальных данных в форму, приемлемую при использовании алгоритмами классификации. В результате формируется пространство признаков. При этом требуется не только сформировать само пространство признаков изображения, но и исключить из рассмотрения неинформативные, одновременно сохра-



нив все признаки, существенные для решения поставленной задачи. На этапе классификации происходит обучение классификатора, причём в качестве обучающей и тестирующей выборки используются оценки признаков, полученные в ходе предобработки экспериментального видеоряда.

## ПРЕДОБРАБОТКА ДАННЫХ

1) Выделение объекта на фоне обстановки. Для каждого кадра на изображении проводилось последовательное вычитание фона с использованием встроенного функционала Matlab, реализующего метод гауссовой смеси (обучение проводилось по 20 кадрам). Затем осуществлялась сегментация изображения, в результате которой точки изображения, не являющиеся фоном, объединялись в блоки, если их суммарное количество в едином блоке было не меньше порогового значения. При анализе экспериментальных данных использовано эмпирически подобранное значение порога, равное 50 пикселям для размера кадра 320 на 240. В случае дальнейшего использования предложенного алгоритма для анализа данных, имеющих иное пространственное разрешение, этот параметр может быть изменён пропорционально.

2) Отслеживание положения в кадре движущегося объекта. Как только по результатам сегментации был получен вывод о наличии подвижного объекта в кадре, создавался фильтр Калмана, который использовался для отслеживания перемещений цели. В нашей работе был применён фильтр, реализованный при помощи функции Matlab `Configure Kalman Filter`. Непостоянство скорости движения объекта слежения учтено с помощью дополнительного параметра фильтра (`Motion Noise`) [13].

При анализе видеоизображения возможны два случая использования фильтра Калмана:

– подвижный объект обнаружен на кадре: тогда фильтр предсказывает положение объекта в кадре видеопоследовательности и использует данные о новом положении объекта для корректировки результатов выделенного при помощи фильтрации положения объекта;

– подвижный объект не обнаружен: тогда полученное при помощи фильтра положение детектируемого объекта в кадре формируется исключительно на основе данных анализа предыдущих кадров.

3) Анализ деформации формы подвижного объекта. В качестве дескриптора контекста формы в отличие от [14], где анализировались все краевые точки объекта, взяты только подвижные характерные краевые точки. Для них выполнялся прокрустов анализ формы [15] по следующему алгоритму.

На каждом из кадров видеоряда для  $k$  характерных краевых точек силуэта человека определялся комплексный вектор:

$Z = [z_1, z_2, \dots, z_k], z_j = x_j + iy_j,$   
где  $i$  – мнимая единица;  $x_j$  и  $y_j$  – координаты  $j$ -й точки изображения.

Центрированная оценка вектора  $Z_C$  формировалась путём умножения вектора координат  $Z$  на центрирующую матрицу  $C$ :

$$Z_C = Z \cdot C,$$

$$C = I_k - \frac{1}{k} \cdot \mathbf{1}_k \cdot \mathbf{1}_k^T,$$

где  $I_k$  – единичная матрица размерностью  $k \times k$ ;  $\mathbf{1}_k$  – единичный вектор размерностью  $k$ .

Для двух центрированных вышеописанным образом последовательностей  $v = (v_1, v_2, \dots, v_k)$  и  $w = (w_1, w_2, \dots, w_k)$  вычислялось расстояние  $D(v, w)$  между ними:

$$D(v, w) = \sqrt{1 - \frac{|v \cdot w|^2}{\|v\|^2 \cdot \|w\|^2}}.$$

Вычисленное таким образом расстояние (прокрустова метрика) для последовательностей характерных точек двух следующих друг за другом кадров является оценкой, чувствительной к значительным деформациям формы объекта, например, вследствие падения.

## ВЫДЕЛЕНИЕ ПРИЗНАКОВ

После завершения предобработки для каждого из кадров видеозаписи оценивался следующий набор параметров, формирующий пространство признаков ( $\Pi$ ):

- 1) оценка расстояния  $D(v, w)$ ;
- 2) координаты центра тяжести области изображения, соответствующей подвижному объекту;

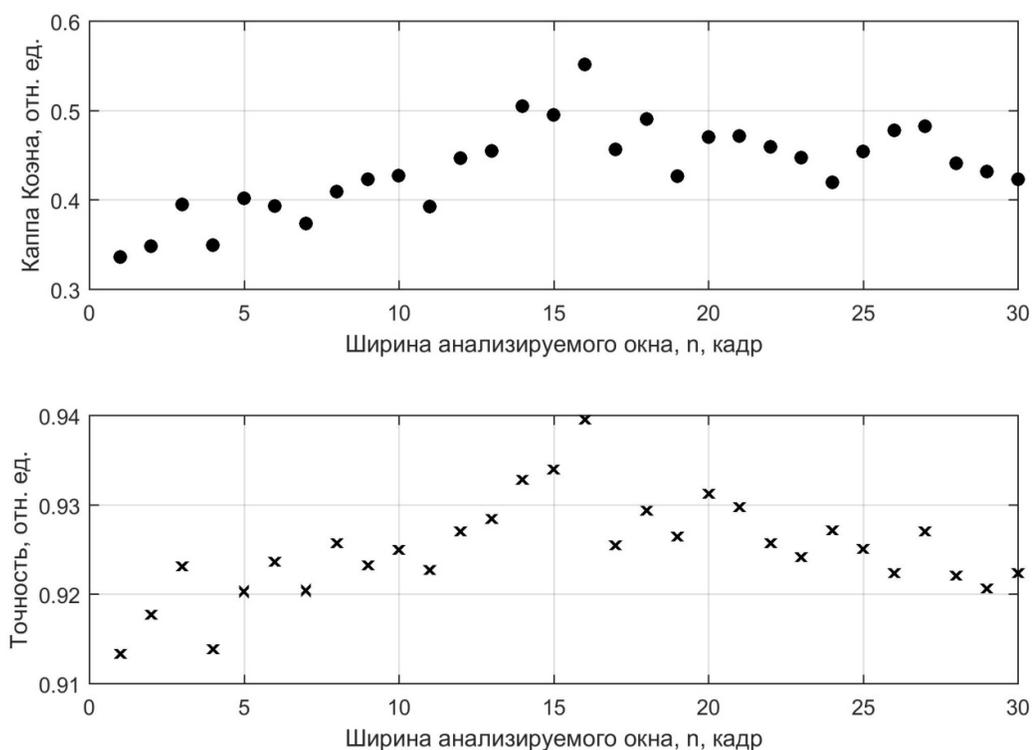


Рис. 1. Зависимости каппы Коэна и точности от ширины анализируемого окна.

3) площадь области изображения, соответствующей подвижному объекту;

4) скорость перемещения центра тяжести области изображения, соответствующей подвижному объекту, вдоль оси  $x$  и  $y$  для двух последовательных кадров.

Выбор данных признаков обусловлен анализом научной литературы по данной тематике [11, 14, 15].

Пространство признаков П1 в ходе выполнения работы было дополнено набором признаков, характеризующих скорость перемещения вдоль оси  $x$  и  $y$  центра тяжести, соответствующей подвижному объекту области изображения в окне длительностью  $n$  кадров. В результате чего формировалось дополненное пространство признаков (П2).

### ВЫБОР СОСТАВА ВЕКТОРА

Анализируемый массив данных, как было указано, содержал 108 видеозаписей (продолжительностью от 20 секунд до 1 минуты), в 84 из которых имелся единичный эпизод падения. Каждый кадр видеозаписи на основе визуального анализа оператором

был отнесён к классу «падение» или «не падение». Выборка экспериментальных данных разделена на обучающую (использовалась для обучения классификатора и подбора оптимального состава пространства признаков) и тестирующую (использовалась при итоговой оценке эффективности классификатора). В обучающую выделено 75 % кадров видеозаписей (65610), оставшиеся 25 % (21870 кадров) составили тестирующую выборку.

Нами использовался классификатор на основе дерева решений из библиотеки машинного обучения Matlab. Для оценки ошибки классификации применён метод перекрёстной проверки (cross-validation) на  $k = 10$  блоках.

В качестве меры оценки качества классификации предложено использовать точность, а также каппу Коэна (меру межэкспертного согласия), поскольку выборка является несбалансированной. Отметим, что выборка называется несбалансированной, если имеет место несбалансированность классов, а именно: они представлены неравномерно. Например,



Матрица ошибок для пространства признаков П1

Не падение		Результат классификации	
		Падение	
Истинный класс	Не падение	19468	871
	Падение	975	556
Каппа Коэна		0,33	
Точность		0,91	

Таблица 2

Матрица ошибок для пространства признаков П2

Не падение		Результат классификации	
		Падение	
Истинный класс	Не падение	19811	528
	Падение	620	911
Каппа Коэна		0,58	
Точность		0,95	

Таблица 3

Матрица ошибок для пространства признаков П2 с использованием дополнительных эвристик

Не падение		Результат классификации	
		Падение	
Истинный класс	Не падение	19845	494
	Падение	559	972
Каппа Коэна		0,62	
Точность		0,95	

в нашем случае менее 10 % выборки соответствует артефактам падений.

Для пространства признаков П1 результаты применения классификатора на тестирующей выборке приведены в таблице 1. Значения в ячейках матрицы ошибок соответствуют количеству кадров, верно либо ошибочно классифицированных предложенным алгоритмом.

Рассмотрим, как влияет на эффективность классификации использование дополненного пространства признаков П2. На рис. 1 приведены данные о том, как изменяется величина каппы Коэна и точность классификации в случае использования дополненного пространства признаков П2 при ширине окна  $n$ , варьирующейся от 2 до 30 кадров. Верхняя граница диапазона величины ширины окна  $n$  выбрана исходя из максимальной продолжительности по времени эпизода падения, равной 1 с, что соответствует 30 кадрам при частоте дискретизации 30 кадров в секунду.

Как следует из рис. 1, при дополнении пространства признаков П1 до пространства признаков П2 каппа Коэна возрастает приблизительно в два раза: с 0,31 до 0,58. При этом наилучшие оценки каппы Коэна и точности соответствуют ширине анализируемого окна, равной 16 кадрам. Матрица ошибок и оценки качества классификации для этого случая приведены в таблице 2.

Визуальное сопоставление результатов применения классификатора и разметки видеозаписи оператором, приведённое на рис. 2, показывает, что относительно низкое значение каппы Коэна обусловлено «фрагментированностью» результатов классификации.

С целью повышения точности было предложено использовать дополнительные правила (эвристики), направленные на объединение кадров, классифицированных как падения, в единый фрагмент записи в случае, если они расположены не более чем в  $m$  кадрах друг от друга. Также

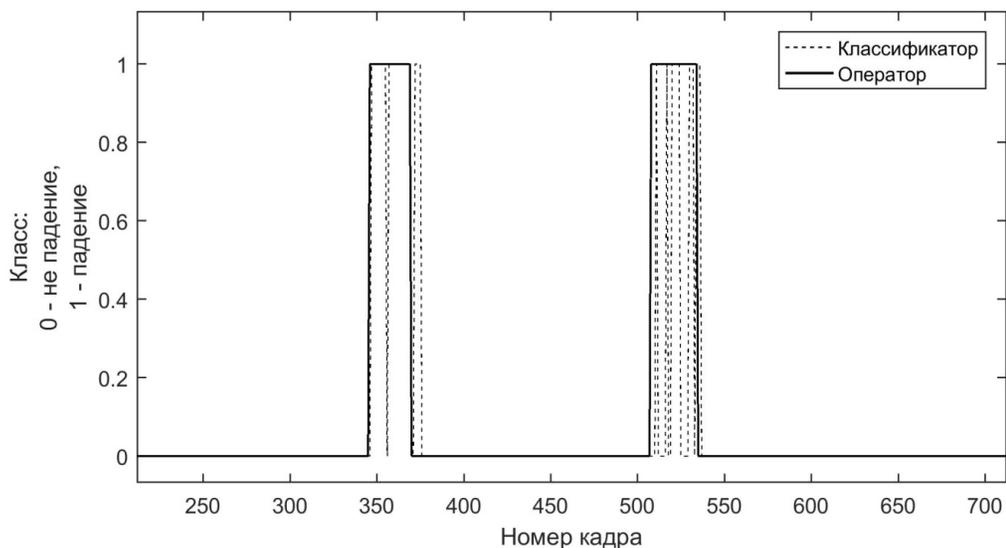


Рис. 2. Сопоставление результатов обработки видеозаписи оператором и классификатором.

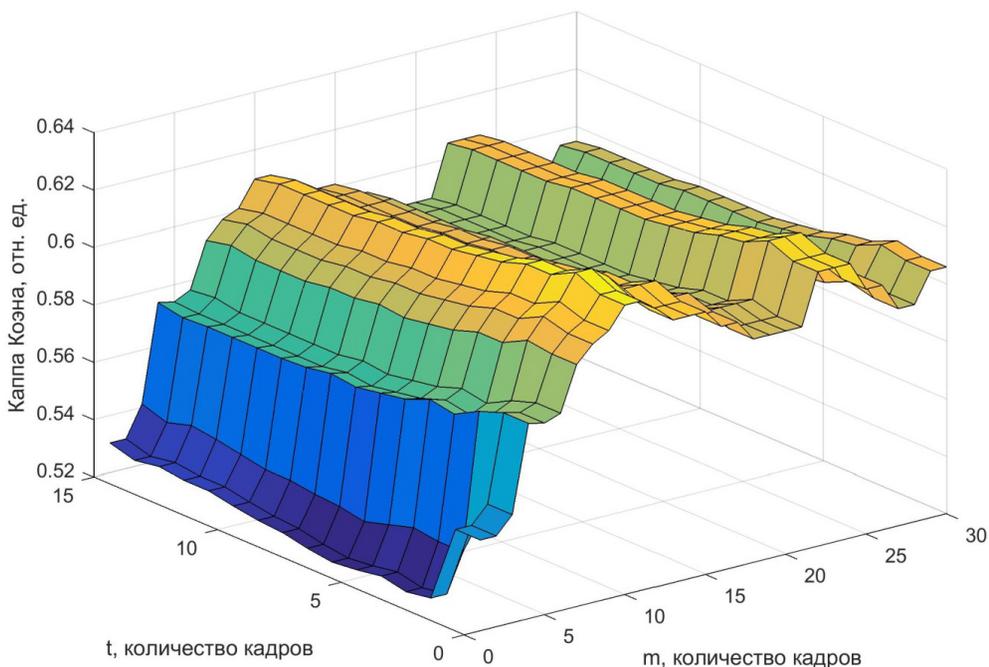


Рис. 3. Зависимость величины каппы Коэна от эвристических параметров.

был введён параметр  $t$ , соответствующий минимальной продолжительности эпизода падения.

Варьируя значениями  $m$  в пределах от 1 до 30 кадров и  $t$  в диапазоне от 0 до 15 кадров, проведена оценка зависимости величины каппы Коэна для обучающей выборки от этих параметров.

Как следует из рис. 3, максимальная величина каппы Коэна достигается при  $m = 10$  кадрам и  $t = 2$  кадрам. В таблице 3 приведена матрица ошибок результатов классификации с использованием предложенных дополнительных эвристик для полученных оптимальных значений параметров.



Таким образом, использование в качестве пространства признаков П2 для ширины окна  $n = 16$  и дополнительных эвристик позволяет добиться точности покадровой классификации эпизодов падений по видеозаписи с величиной каппы Коэна 0,62.

## ЗАКЛЮЧЕНИЕ

Предложен метод интеллектуального анализа видеозаписей камер наружного наблюдения, позволяющий выявлять ситуации, представляющие опасность для жизни и здоровья людей на железнодорожном транспорте на примере выявления эпизодов падений. Разработан алгоритм предобработки видеоряда с целью формирования пространства признаков, основанный на использовании вычитания фона по методу гауссовой смеси, последующем отслеживании перемещения человека при помощи фильтра Калмана и деформации формы подвижного объекта в результате применения прокрустового анализа формы. Предложенный метод классификации на основе древа решений протестирован на базе данных из 108 видеозаписей, 84 из которых содержали единичный эпизод падения.

Проведённое сравнительное исследование нескольких наборов признаков позволило обосновать выбор оптимального набора признаков и дополнительных эвристик, обеспечивающих выделение эпизодов падений по видеозаписи со средним качеством каппы Коэна 0,62.

В дальнейшем предполагается расширение экспериментальной выборки и дополнение классификатора возможностью распознавания по видеозаписи людей, находящихся в состоянии алкогольного или наркотического опьянения, по особенностям их походки.

## ЛИТЕРАТУРА

1. Advances Recognition Systems: Rapid-Access Biometric and Credential Solution, NeoFace Express [Электронный ресурс] / Офф. сайт корпорации NEC, 2017. URL: <https://www.necam.com/docs/?id=6e812b4d-2a12-40ed-9fea-fae81550c7aa> (Доступ 05.11.2017).

2. SMARTGLASSES7 [Электронный ресурс] / Офф. сайт компании ODG, 2017. URL: [https://www.osterhoutgroup.com/pub/static/version1515417478/frontend/Infortis/ultimo/en\\_US/pdf/R-7-TechSheet.pdf](https://www.osterhoutgroup.com/pub/static/version1515417478/frontend/Infortis/ultimo/en_US/pdf/R-7-TechSheet.pdf) (Доступ 05.11.2017).

3. Yaniv Taigman, Ming Yang, Marc Aurelio Ranzato, Lior Wolf. «DeepFace: Closing the Gap to Human-Level Performance in Face Verification», Conference on Computer Vision and Pattern Recognition (CVPR), June 24, 2014.

4. Parkhi [et al]. Deep Face Recognition [Электронный ресурс]: URL: <https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf> (Доступ 05.11.2017).

5. Mohammadian A., Aghaeinia H., Towhidkhal F. Video-based facial expression recognition by removing the style variations in Image Processing, IET, 2015, vol. 9, No. 7, pp. 596–603.

6. Iosifidis A., Tefas A., Pitas I. Class-specific Reference Discriminant Analysis with application in Human Behavior Analysis, IEEE Transactions on Human-Machine Systems, 2015, vol. 45, no. 3, pp. 315–326.

7. Maddalena L., Petrosino A. Stopped Object Detection by Learning Foreground Model in Videos, in IEEE Transactions on Neural Networks and Learning Systems, May 2013, vol. 24, no. 5, pp. 723–735.

8. Amrutha M. P., Vince Paul. Study of Different Obstacle Detection Methods in Railway Track, International Journal of Innovative Research in Computer and Communication Engineering, Jan 2017, vol. 5, No. 1, pp. 1204–1208.

9. Filonenko A., Hernández D. C., Jo K. H. Real-time smoke detection for surveillance, 2015, IEEE 13<sup>th</sup> International Conference on Industrial Informatics (INDIN), Cambridge, 2015, pp. 568–571.

10. Rougier C., Meunier J., St-Arnaud A., Rousseau J. Fall detection from human shape and motion history using video surveillance, Proc. 21<sup>st</sup> Int. Conf. AINAW, 2007, vol. 2, pp. 875–880.

11. Lee T., Mihailidis A. An intelligent emergency response system: Preliminary development and testing of automated fall detection, J. Telemed. Telecare, 2005, vol. 11, no. 4, pp. 194–198.

12. Charfi I., Miteran J., Dubois J., Atri M., Tourki R. Optimised spatio-temporal descriptors for real-time fall detection: comparison of SVM and Adaboost based classification, Journal of Electronic Imaging (JEI), Vol. 22, Iss. 4, pp. 17, October 2013.

13. MathWork Documentation: Create Kalman filter for object tracking [Электронный ресурс] / Офф. сайт MathWorks, 1994–2017. URL: <https://www.mathworks.com/help/vision/ref/configurekalmanfilter.html> (Доступ 05.11.2017).

14. Mori G., Malik J. Estimating human body configurations using shape context matching, in Proc. Eur. Conf. Comput. Vision, 2002, vol. 2352, pp. 150–180.

15. Rougier C., Meunier J., St-Arnaud A., Rousseau J. Robust Video Surveillance for Fall Detection Based on Human Shape Deformation, in IEEE Transactions on Circuits and Systems for Video Technology, May 2011, vol. 21, no. 5, pp. 611–622.

Координаты авторов: **Анищенко Л. Н.** – anishchenko@rslab.ru, **Ивашов С. И.** – sivashiv@rslab.ru, **Скребков А. В.** – skrebkov\_av@mail.ru.

Статья поступила в редакцию 04.10.2017, принята к публикации 05.11.2017.

**Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант № 17–20–03034).**