

# Распознавание действий человека с использованием снижения размерности и машины опорных векторов

Л.В. Ширипова<sup>1</sup>, Е.В. Мясников<sup>1,2</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, 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 представлен метод распознавания действий человека с использованием видеосъемки в оптическом диапазоне. Подход, предложенный в работе, состоит в выделении движущегося человека на видеоряде с последующей нормализацией размера, выделением подпоследовательностей и снижением размерности с использованием метода главных компонент. Классификация действий человека выполняется с использованием машины опорных векторов. Экспериментальные исследования, выполненные на базе данных Weizmann, позволили определить наилучшие значения параметров метода. Полученные результаты показали, что при небольшом количестве классов может быть достигнута высокая точность классификации действий человека.

## 1. Введение

Распознавание действий человека активно применяется в различных сферах: при создании интерфейсов человек-машина, в сфере развлечений, в обеспечении общественной безопасности и т.д.

Распознавание действий человека предполагает решение двух задач [1]:

1. Извлечение признаков информации, т.е. преобразование видеопотока или последовательности изображений в вид, пригодный для последующей классификации;

2. Собственно классификация признаков информации, полученной на первом этапе.

Для решения данных задач было предложено множество подходов, подробно описанных в работе [1]. Рассмотрим некоторые из них.

Для выделения признаков информации авторы работы [6] предложили выделять силуэт из каждого кадра, получать изображения разницы между соседними кадрами и строить окончательное изображение, накладывая полученные изображения друг на друга. Полученное изображение получило название Motion Energy Image (MEI). Помимо этого, вводится понятие Motion History Image (MHI) - изображение, интенсивность каждого пикселя которого зависит от времени возникновения действия в данной точке. Предложенный подход показал хорошие результаты, однако он имеет недостатки при изменении угла наблюдения.

Для устранения указанной проблемы в работе [7] предложен обобщающий подход, связанный с использованием 3D motion history volume (MHV). MHV основывается на 3D вокселях, полученных для различных углов наблюдения. Далее используется преобразование Фурье для получения признаков, инвариантных к положению и вращению.

Другой подход к получению признаков связан с выделением пространственно-временных точек интереса (space-time interest points, STIPs). Так, авторы работы [2,8] расширили детектор углов Харриса на пространственно-временную область. Гауссова функция затем используется для определения изменений в движении в пространственной и временной областях. В работах [9, 10, 11] для получения признаков используется гистограмма направленных градиентов (HOG) и гистограмма оптического потока (HOF). Однако точки интереса помогают получить информацию только за короткий промежуток времени. Для отслеживания изменений точек интереса в работе [12] было предложено использовать KLT трекер (Kanade–Lucas–Tomasi feature tracker). В работе [13] в качестве признаков используются простые параметры выпуклых фигур.

Для классификации полученных признаков используются различные подходы, например, машина опорных векторов (SVM) [14, 15, 9], классификатор по ближайшему соседу (k-NN) [16, 17, 18], а также скрытые Марковские модели (HMM) [19, 20, 21].

В настоящей работе для решения задачи распознавания действий человека применяется подход, основанный на снижении размерности с использованием метода главных компонент и последующей классификации с использованием машины опорных векторов. Подобный подход успешно использовался нами ранее [22, 23] при решении задачи идентификации человека по походке.

Работа организована следующим образом. В разделе 2 приводятся описание разработанного метода распознавания действий человека. В разделе 3 описываются результаты экспериментальных исследований, выполненных на базе данных Weizmann. В конце работы приводится заключение и список литературы.

## 2. Методы

Предложенный ранее метод [22, 23] включает следующие этапы:

- выделение движущегося человека на видеопоследовательности,
- нормализацию размера выделенного фрагмента видеопоследовательности,
- выделение подпоследовательностей,
- снижение размерности выделенного фрагмента видеопоследовательности,
- классификацию видеопоследовательности.

### 2.1. Выделение движущегося человека на видеопоследовательности

На первом этапе разработанного метода осуществляется выделение движущегося человека на видеопоследовательности. Наиболее часто для такого выделения в случае, когда источником видеопоследовательности является камера видеонаблюдения, используются методы вычитания фона. Основная идея методов этого класса заключается в использовании некоей модели фона и принятии решения о принадлежности того или иного пикселя к фону или движущемуся объекту, исходя из его соответствия модели фона. Сама модель фона с течением времени постепенно уточняется. Хотя в простейших приложениях в качестве модели фона может быть использовано усредненное по времени наблюдения изображение, в рассматриваемой задаче лучшие результаты дают более сложные модели, например [24-26].

В настоящей работе для выделения на видеопоследовательности движущегося человека использовался метод вычитания фона на основе смеси гауссовских распределений (Gaussian mixture model, GMM) [25]. При использовании указанного метода каждый отсчет фона моделируется взвешенной суммой (смесью) гауссиан. При этом вес той или иной гауссианы определяется временем, в течение которого соответствующей гауссиане цвет присутствует на видеопоследовательности.

Отметим, что при выборе метода на основе смеси гауссовских распределений учитывались как собственные предварительные эксперименты, так и опыт использования указанного метода другими исследователями при решении рассматриваемой задачи [27, 28].

По завершении первого этапа метода формируется набор масок, соответствующих отдельным кадрам видеопоследовательности. При этом каждая маска отражает результат сегментации кадра на область переднего плана, соответствующую движущемуся человеку, и фона.

### *2.2. Нормализация размера выделенного фрагмента видеопоследовательности*

На втором этапе метода полученные покадрово маски обрабатываются следующим образом. Вначале производится поиск центра масс каждой области переднего плана, затем определяются линейные размеры области, и выполняется кадрирование (усечение изображения маски). После этого кадрированное изображение подвергается масштабированию (сжатию) до заданных размеров.

Очевидно, что с учетом временной координаты размерность описывающих движение человека данных остается высокой даже после их нормализации (кадрирования и сжатия) изображений. В связи с этим производится снижение размерности данных, описывающих движение человека.

### *2.3. Выделение подпоследовательностей*

Для каждой последовательности кадров, содержащей движение, выделяется набор подпоследовательностей заданной длины. Выделение подпоследовательностей осуществляется с некоторым заданным шагом, начиная с начала исходной последовательности. Подробное описание выделения подпоследовательностей приведено в предыдущих работах [22, 23].

Для каждой выделенной подпоследовательности вектор признаков формируется следующим образом: каждый нормализованный кадр подпоследовательности разворачивается в строку, и полученные для отдельных кадров строки сцепляются между собой. Вектора признаков подпоследовательностей всех последовательностей формируют входную матрицу для метода главных компонент.

### *2.4. Снижение размерности с использованием метода главных компонент*

Для снижения размерности многомерных данных используются как линейные, так и нелинейные методы. Наиболее часто используются линейные методы, такие как метод главных компонент (principal component analysis, PCA) [29] и анализ независимых компонент (ICA). Нелинейные методы снижения размерности (например, нелинейное отображение [30], ISOMAP [31], LLE [32]) используются реже в связи с высокой вычислительной сложностью таких методов. Следует отметить, что в последнее время предпринимаются попытки ускорения таких методов.

В настоящей работе применяется метод главных компонент, как наиболее часто используемый в таких случаях. Этот метод осуществляет поиск линейной проекции в подпространство меньшей размерности, максимизирующей разброс данных. Метод PCA часто рассматривается, как линейный метод снижения размерности данных, минимизирующий потерю информации.

После нахождения главных компонент в качестве признакового описания берется проекция данных исходных подпоследовательностей на первые  $N$  главных компонент.

### *2.5. Классификация видеопоследовательностей*

Полученные в результате метода главных компонент признаки используются для обучения классификатора, в качестве которого в настоящей работе выступает машина опорных векторов (support vector machines, SVM) [33]. При этом классы соответствуют отдельным действиям, а вектора признаков, полученные для всех соответствующих этим действиям подпоследовательностей, соответствуют отдельным реализациям.

Отметим, что приведенное в текущем разделе описание справедливо для режима обучения, в котором производится настройка, как процедуры снижения размерности, так и классификатора. В режиме тестирования данные проходят обработку аналогичным образом, за исключением того, что при снижении размерности используются параметры полученного в режиме обучения линейного преобразования с последующей классификацией обученным классификатором SVM.

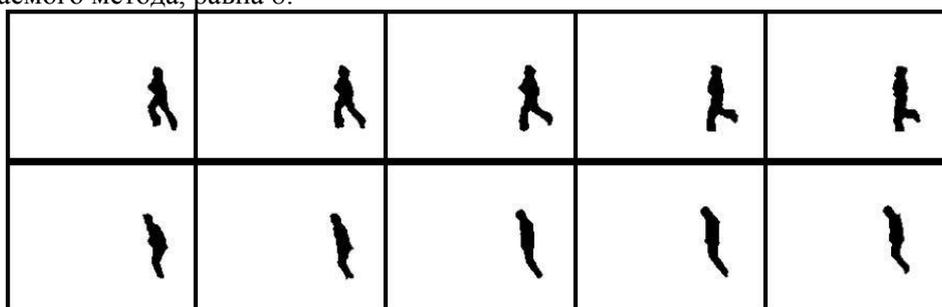
## **3. Эксперименты**

Предложенный в работе метод был реализован на языке C++ с использованием библиотеки OpenCV. При выполнении экспериментальных исследований был использован ПК на базе Intel Core i5-3470 CPU 3.2 ГГц.

Для экспериментального исследования предложенного метода были использованы раскадрированные видеопоследовательности из открытой базы данных Weizmann (рисунок 1). Указанная база данных содержит последовательности бинарных изображений, соответствующих отдельным кадрам видеопоследовательности, на которых уже выполнено выделение движущихся объектов (сегментация переднего плана и фона). В базе данных представлены видеопоследовательности для 9 человек, выполняющих 10 различных действий. Всего база данных содержит 90 последовательностей. Таким образом, в каждом классе было 9 последовательностей. Минимальная длина последовательности составляла 28 кадров.

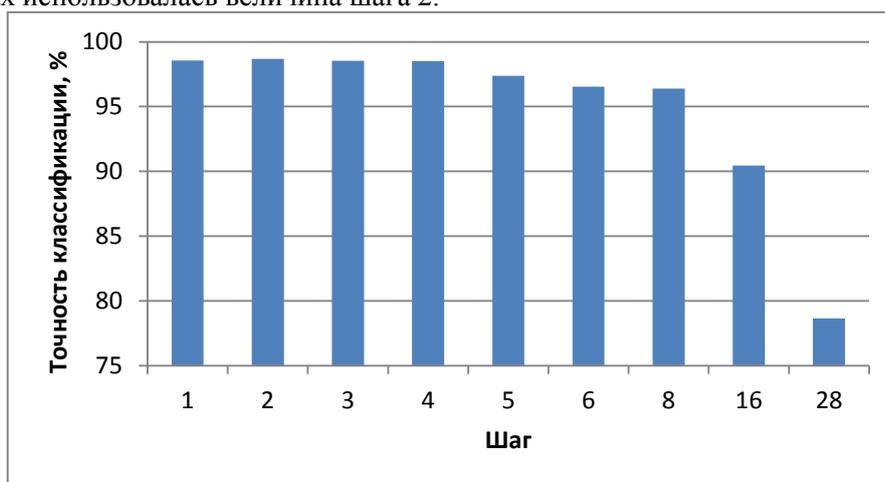
Классы были разделены на обучающую и тестовую выборки. Обучающая выборка содержала 4 последовательности, тестовая – 5. Последовательности каждого класса были предварительно обработаны с использованием алгоритма, описанного в п.2.2. Далее производилось выделение подпоследовательностей согласно п.2.3. Затем выполнялось снижение размерности данных с использованием метода, описанного в п.2.4 и классификация с использованием алгоритма, описанного в п.2.5. Для определения качества работы метода оценивалась точность классификации, определяемая как доля правильно классифицированных объектов.

В первом эксперименте исследовалась зависимость точности классификации от величины шага, используемого при выделении подпоследовательностей. Длина подпоследовательности составляла 28 кадров. Выходная размерность векторов признаков, формируемых на шаге 2.4 рассматриваемого метода, равна 8.



**Рисунок 1.** Примеры последовательностей из базы данных Weizmann: бег и прыжки в длину.

Результаты эксперимента представлены на рисунке 2. Экспериментально было определено, что наилучшая точность классификации достигается при малых значениях шага. В дальнейших экспериментах использовалась величина шага 2.



**Рисунок 2.** Зависимость точности классификации от величины шага.

Во втором эксперименте исследовалась зависимость точности классификации от выходной размерности векторов признаков, формируемых на шаге 2.4 рассматриваемого метода. При выполнении эксперимента выходная размерность варьировалась от 4 до 512, а другие

параметры оставались фиксированными. В частности, шаг, используемый при выделении подпоследовательностей, составлял 2 кадра.

Результаты эксперимента представлены на рисунке 3. Как видно из приведенных результатов, наилучшие значения показателя точности классификации достигаются для размерности 8, 16 и 64.

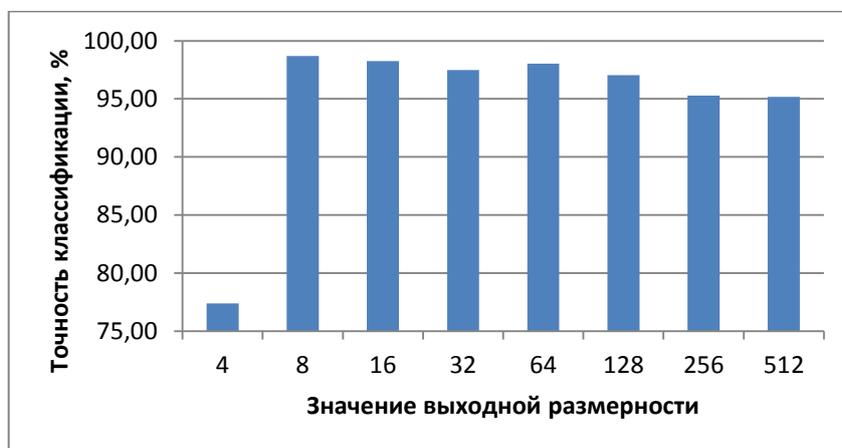


Рисунок 3. Зависимость точности классификации от размерности вектора признаков.

В третьем эксперименте исследовалась зависимость точности классификации от длины подпоследовательностей. При выполнении эксперимента длина подпоследовательности варьировалась от 4 до 28 кадров, а другие параметры оставались фиксированными. Так, шаг, используемый при выделении подпоследовательностей, составлял 2 кадра, выходная размерность равнялась 64.

Результаты эксперимента представлены на рисунке 4. Как видно из приведенных результатов, наилучшие значения показателя точности классификации достигаются при длине подпоследовательностей 28 кадров.

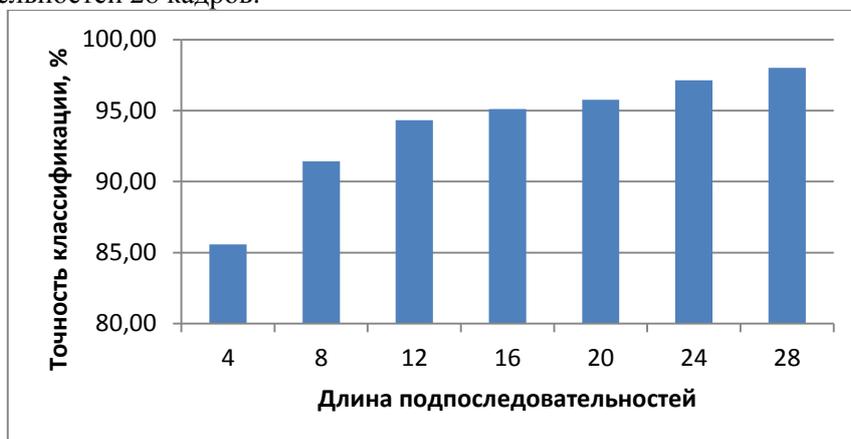


Рисунок 4. Зависимость точности классификации от длины подпоследовательностей.

Как видно из результатов приведенных экспериментов, при относительно небольшом количестве классов (10 классов) обеспечивается высокая (не менее 95%) точность классификации.

#### 4. Заключение

Предложенный нами метод распознавания действий человека состоит в выделении движущегося человека на видеоряде, нормализации размера, выделении подпоследовательностей, снижении размерности с использованием метода главных компонент и классификации с использованием машины опорных векторов.

Эксперименты, выполненные на базе данных Weizmann, позволили определить наилучшие значения параметров разработанного метода. Было показано, что при небольшом количестве классов (10 классов) предложенный метод обеспечивает на указанной базе данных высокую (при широком диапазоне параметров - не менее 95%, а при использовании лучших значений - 98%) точность классификации личности.

В будущем планируется расширить список алгоритмов, используемых для формирования признакового описания и список используемых методов классификации.

## 5. Литература

- [1] Kong, Yu. Human action recognition and prediction: a survey / Yu Kong, Yun Fu // *Journal of Latex class files*. – 2018. – Vol. 19.
- [2] Laptev, I. On space-time interest points // *IJCV*. – 2005. – Vol. 64(2). – P. 107-123.
- [3] Raptis, M. Poselet key-framing: A model for human activity recognition / M. Raptis, L. Sigal // *CVPR*, 2013. – P. 2650-2657
- [4] Ji, S. 3d convolutional neural networks for human action recognition / S. Ji, W. Xu, M. Yang, K. Yu // *IEEE Trans. Pattern Analysis and Machine Intelligence*. – 2013. – Vol. 35(1). – P. 221-231.
- [5] Carreira, J. Quo vadis, action recognition? a new model and the kinetics dataset / J. Carreira, A. Zisserman // *CVPR*, 2017.
- [6] Bobick, A. The Recognition of Human Movement Using Temporal Templates / A.F. Bobick, J.W. Davis // *IEEE Trans Pattern Analysis and Machine Intelligence*. – 2001. – Vol. 23. – P.257-267
- [7] Weinland, D. Free viewpoint action recognition using motion history volumes / D. Weinland, R. Ronfard, E. Boyer // *Computer Vision and Image Understanding*. – 2006. – Vol. 104. – P. 249-257.
- [8] Laptev, I. Space-time interest points / I. Laptev, T. Lindeberg // *ICCV*. – 2003. – P. 432-439.
- [9] Laptev, I. Learning realistic human actions from movies / I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld // *CVPR*, 2008. DOI: 10.1109/CVPR.2008.4587756.
- [10] Klaser, A. A spatio-temporal descriptor based on 3d-gradients / A. Klaser, M. Marszalek, C. Schmid // *BMVC*, 2008. – P. 275-284.
- [11] Dalal, N. Histograms of oriented gradients for human detection / N. Dalal, B. Triggs // *CVPR*, 2005. DOI: 10.1109/CVPR.2005.177.
- [12] Messing, R. Activity recognition using the velocity histories of tracked keypoints / R. Messing, C. Pal, H. Kautz // *ICCV*, 2009. – P. 104-111.
- [13] Gosciewska, K. Silhouette-based action recognition using simple shape descriptors / K. Gosciewska, D. Frejlichowski // Springer, 2018.
- [14] Laptev, I. Recognizing Human Actions: A Local SVM Approach / C. Schuldt, I. Laptev, B. Caputo // *Proc. ICPR*, Cambridge, UK, 2004.
- [15] Marszalek, M. Actions in context / M. Marszalek, I. Laptev, C. Schmid // *Computer Vision and Pattern Recognition*, 2009. – P. 2929-2936.
- [16] Blank, M. Actions as space-time shapes / M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri // *Proc. ICCV*, 2005. – P. 1-8.
- [17] Laptev, I. Retrieving actions in movies / I. Laptev, P. Perez // *ICCV*, 2007. DOI: 10.1109/ICCV.2007.4409105.
- [18] Tran, D. Human activity recognition with metric learning / D. Tran, A. Sorokin // *ECCV*, 2008. – P. 548-561.
- [19] Duong, T.V. Activity recognition and abnormality detection with the switching hidden semi-markov model / T.V. Duong, H.H. Bui, D.Q. Phung, S. Venkatesh // *CVPR*, 2005. DOI: 10.1109/CVPR.2005.61.
- [20] Rajko, S. Real-time gesture recognition with minimal training requirements and on-line learning / S. Rajko, G. Qian, T. Ingalls, J. James // *CVPR*, 2007. DOI: 10.1109/CVPR.2007.383330.
- [21] Ikizler, N. Searching video for complex activities with finite state models / N. Ikizler, D. Forsyth // *CVPR*, 2007. DOI: 10.1109/CVPR.2007.383168.

- [22] Shiripova, L. Gait analysis for person recognition using principal component analysis and support vector machines / L. Shiripova, O. Strukova, E. Myasnikov // CEUR Workshop Proceedings. – 2018. – Vol. 2210. – P. 170-176.
- [23] Shiripova, L. Comparative analysis of classification methods for human identification by gait / L. Shiripova, E. Myasnikov // AIST-2018. CEUR Workshop Proceedings. (in Print).
- [24] KadewTraKuPong, P. An improved adaptive background mixture model for real-time tracking with shadow detection / P. KadewTraKuPong, R. Bowden // Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems: Computer Vision and Distributed Processing. – 2001. – P. 1-5.
- [25] Improved adaptive Gaussian mixture model for background subtraction / Z. Zivkovic // ICPR. – 2004. – P. 1-4.
- [26] Godbehere, A.B. Visual Tracking of Human Visitors under Variable-Lighting Conditions for a Responsive Audio Art Installation / A.B. Godbehere, A. Matsukawa, K. Goldberg // American Control Conference (ACC). – 2012. – P. 1-8.
- [27] Murukesh, C. Secured Authentication through Integration of Gait and Footprint for Human Identification / C. Murukesh, K. Thanushkodi, P. Padmanabhan, M.D. Feroze Naina // Journal of Electrical Engineering and Technology. – 2014. – Vol. 9(6). – P. 2118-2125.
- [28] Wang, L. Automatic Gait Recognition Based on Statistical Shape Analysis / L Wang, T. Tan, W. Hu, H. Ning // Transactions on Image Processing. – 2003. – Vol. 12. – P. 1-13.
- [29] Fukunaga K. Introduction to Statistical Pattern Recognition / K. Fukunaga. – London: Academic Press, 2003.
- [30] Sammon, J.W. A nonlinear mapping for data structure analysis // Transactions on Computers. – 1969. – Vol. 18. – P. 401-409.
- [31] Tenenbaum, J.B. A Global Geometric Framework for Nonlinear Dimensionality Reduction / J.B. Tenenbaum, V. de Silva, J.C. Langford // Science. – 2000. – P. 2319-2323.
- [32] Roweis, S.T. Nonlinear Dimensionality Reduction by Locally Linear Embedding / S.T. Roweis, L.K. Saul // Science. – 2000. – Vol. 290. – P. 2323-2326.
- [33] Cortes, C. Support-vector networks / C. Cortes, V. Vapnik // Machine Learning. – 1995. – Vol. 20(3). – P. 273-297.

### **Благодарности**

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-29-03190-офи-м.

# Human action recognition using dimensionality reduction and support vector machine

L.V. Shiripova<sup>1</sup>, E.V. Myasnikov<sup>1,2</sup>

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34A, 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 presents human action recognition using a video recorded in the optical range. The method proposed in this paper consists in the detection of a moving person on a video sequence with size normalization, formation a set of subsequences and dimensionality reduction using the principal component analysis technique. The classification of the human action is carried out using the support vector machine classifier. The experimental studies performed using the Weizmann dataset allowed us to determine the best values of the method parameters. The obtained results show that with a small number of classes, high classification accuracy can be achieved.