



Silhouette-Based Action Recognition Using Simple Shape Descriptors

Katarzyna Gościewska^(✉) and Dariusz Frejlichowski

Faculty of Computer Science and Information Technology, West Pomeranian
University of Technology, Szczecin, Żołnierska 52, 71-210 Szczecin, Poland
{kgosciewska,dfrejlichowski}@wi.zut.edu.pl

Abstract. This paper presents human action recognition method based on silhouette sequences and simple shape descriptors. The proposed solution uses single scalar shape measures to represent each silhouette from an action sequence. Scalars are then combined into a vector that represents the entire sequence. In the following step, vectors are transformed into sequence representations and matched with the use of leave-one-out cross-validation technique and selected similarity or dissimilarity measure. Additionally, action sequences are pre-classified using the information about centroid trajectory into two subgroups—actions that are performed in place and actions during which a person moves in the frame. The average percentage accuracy is 80%—the result is very satisfactory taking into consideration the very small amount of data used. The paper provides information on the approach, some key definitions as well as experimental results.

Keywords: Action recognition · Simple shape descriptors
Binary silhouette sequences

1 Introduction

Vision-based human action recognition is aimed at labelling image sequences with action labels. It is applicable in several domains such as video surveillance, human-computer interaction, video retrieval, scene analysis and automatic video annotation for efficient searching. The topic has been attracting many researchers for recent years—it is especially seen in constantly updated surveys and reviews (e.g. [1–6]) as well as growing number of methods and algorithms. According to [3] the main task in action recognition is feature extraction, and features can vary in complexity. Action recognition can include low-level features and interests points or higher level representations such as long-term trajectories and semantics, as well as silhouettes seen as a progression of body posture [4]. When a frame or sequence representation is available then action recognition is considered as a classification problem [2].

Motion recognition can be related to elementary gestures, primitive actions, single activities, interactions or complex behaviours and crowd activities. This

paper focuses only on actions and does not consider any other information except full body shape of object’s binary silhouette. This excludes the recognition of single gestures. Moreover, action recognition should not be confused with gait recognition which identifies personal styles of movement. Action recognition also excludes the use of context information about background and interactions with objects or other people. Action recognition in which we are interested in can be seen as a non-hierarchical approach that is aimed to recognize primitive, short and repetitive actions [7], among which are jumping, walking or running. These actions are also classified as heterogeneous actions [8]. The most popular benchmark datasets consisting of these type of actions are KTH [9] and Weizmann [10].

Some silhouette-based state-of-the-art solutions are discussed in Sect. 2. Section 3 contains a description of the proposed method—consecutive processing steps together with the type of input data and applied algorithms are explained in detail. Section 4 includes information on the experimental conditions and results. Section 5 concludes the paper.

2 Related Works on Silhouette-Based Human Action Recognition

This section gives a short overview on some methods and algorithms which belong to the same category as our method, i.e. relate to action recognition based on sequences of binary silhouettes corresponding to simple actions such as running, bending or waving.

Bobick and Davis [11] proposed a technique that uses silhouettes to generate motion energy images (MEI) depicting where the movement is, and motion history images (MHI) to show how the object is moving. Hu moments are then extracted from MEI and MHI, and the resultant action descriptors are matched using Mahalanobis distance. Hu moments are also applied in [12] and calculated using modified MHI which stands as a feature for action classification. The modification involves a change in decay factor—instead of constant linear decay factor an exponential decay factor is used which emphasizes the recent motion and increases recognition accuracy. Hidden Markov Model (HMM) is applied for classification. The combination of modified MHI and HMM can achieve 99% accuracy which exceeds the results when the HMM is used for silhouettes only.

Silhouette features can also be identified by model fitting, e.g. using a star figure which helps to localize head and limbs [13]. In turn, Gorelick et al. [10] proposed to accumulate silhouettes into three-dimensional representations and to employ Poisson equation to extract features of human actions. The opposite to this is to use only selected frames, so called key poses (e.g. [14–16]). For example, the authors of [14] proposed simple action representation method based only on key poses without any temporal information. Single silhouette—a pose—is represented as a collection of line-pairs. A matching scheme between two frames is proposed to obtain similarity values. Then authors use a method for extracting candidate key poses based on k-medoids clustering algorithm and learning algorithm to rank the potentiality of each candidate—candidate key frames with

the highest potentiality scores are selected as final key poses. Action sequence classification step requires the comparison of each frame with key poses to assign a label. A single sequence is then classified based on the majority voting.

The authors of [17] proposed two new feature extraction methods based on well-known Trace transform, which is employed for binary silhouettes that represent single action period. The first method extracts Trace transforms for each silhouette and then a final history template is created from these transforms. The second method uses Trace transform to construct a set of invariant features for action sequence representation. The classification process carried out with the use of Weizmann database and Radial Basis Function Kernel SVM gave percentage results exceeding 90%.

In [18] various feature extraction methods based on binary shape are investigated. The first group contains approaches using contour points only, such as Cartesian Coordinate Features, Fourier Descriptor Features, Centroid-Distance Features and Chord-Length Features. Other approaches use information about all points of a silhouette, and these are Histogram of Oriented Gradients, Histogram of Oriented Optical Flow and Structural Similarity Index Measure. All experiments in cited paper are based on space-time approach and above-mentioned features are extracted from Aligned Silhouettes Image, which is an accumulation of all frames in one video. That gives one image per sequence capturing all spatial and temporal features. Action sequences are classified using K-Nearest-Neighbour and Support Vector Machine. The highest correct recognition rate was obtained for Histogram of Oriented Gradients feature and K-Nearest-Neighbour classifier with Leave-One-Video-Out technique.

There are also some recent papers published in 2018 that still rely on KTH and Weizmann datasets in the experiments. The authors of [19] proposed the integration of Histogram of Oriented Gradients and Principal Component Analysis to obtain feature descriptor which is then used to train K-Nearest-Neighbour classifier. Such combination of methods enabled to achieve average classification accuracy exceeding 90%. In [20] a new approach for action recognition is presented. Human silhouettes are extracted from the video using texture segmentation and average energy images are formed. The representation of these images is composed of shape-based spatial distribution of gradients and view independent features. Moreover, Gabor wavelets are used to compute additional features. All features are fused to create robust descriptor which together with SVM classifier and Leave-One-Out technique resulted in recognition accuracy equal to 97.8% for Weizmann dataset.

3 Proposed Method

In the paper we suggest a processing procedure which uses binary silhouette sequences as input data for human action recognition. Our solution is based on simple shape parameters and well-known algorithms which are composed in several consecutive steps enabling to obtain satisfactory classification accuracy while preserving low-dimensionality of representation at the same time. To obtain that

goal we employ various simple shape descriptors—shape measurements or shape factors—that characterize shape by a single value. Simple geometric features are often used to discriminate shapes with large differences, not as a standalone shape descriptors. However, in our method we combine shape representations of many frames in one vector and then transform it into sequence representation. This paper is a continuation of the research presented in [21,22].

In our experiments we use the Weizmann database and therefore the explanation of the proposed method is adapted to it. The Weizmann database for action classification purposes contains foreground binary masks (in Matlab format), extracted from 90 video sequences (180×144 px, 50 fps) using background subtraction [10]. Therefore, the original images are 180×144 pixels in size and one image contains one human silhouette (see Fig. 1 for example).



Fig. 1. Exemplary aligned silhouettes of running action taken from the Weizmann database for action classification.

3.1 Processing Steps

The proposed method includes following steps:

1. Each silhouette is replaced with its convex hull which is aligned with respect to its centroid in the image centre. A convex hull is the smallest convex region containing all points of the original shape. Single input image has background with black pixels and the foreground object consisting of white pixels. Image size equals to 144×144 pixels.
2. Each shape is represented using a single value—a shape measurement or shape ratio, so called simple shape descriptor. The reason for using different shape descriptors is the necessity to experimentally select such a shape feature that will result in the highest classification accuracy.
3. For each sequence, all simple shape descriptors are combined into one vector and values are normalized to interval $[0,1]$.
4. Each vector has different length due to various number of frames. In order to equalize them all vectors are transformed into frequency domain using periodogram or Fast Fourier Transform (FFT) for vectors which were zero-padded to the maximum sequence length. Both approaches are used separately or combined. Each transformed vector becomes a sequence representation.
5. A pre-classification step is introduced which divides the database into two subgroups—actions that are performed in place and actions during which a person changes location. To do so, the information about centroid trajectory is used. Then, the following step is performed in both subgroups separately.

6. Final sequence representations are matched with the use of leave-one-out cross-validation technique to obtain classification accuracy. For comparison, we use three matching measures—standard correlation coefficient, C1 correlation [23] and Euclidean distance. The percentage of correctly classified action sequences indicates the effectiveness of the proposed approach.

3.2 Simple Shape Descriptors

The simplest measurements such as area A and perimeter P may be calculated as the number of pixels of shape's region or contour respectively [24]. Another basic shape feature is its diameter, which can be expressed using Feret diameters (Feret measures): X Feret and Y Feret which are the distances between the minimal and maximal coordinates of a contour, horizontal and vertical respectively. The X/Y Feret is the ratio of the X Feret to Y Feret, and Max Feret is the maximum distance between any two points of a contour.

Another six shape factors are compactness, roundness, circularity ratio, circle variance, ellipse variance and the ratio of width to length. Compactness can be computed as the ratio of the square of the shape's perimeter to its area. The most compact shape is a circle. The roundness is a measure of a shape's sharpness and is based on two basic shape features, i.e. the area and perimeter. The circularity ratio defines the degree of similarity between an original shape and a perfect circle. It can be calculated as the ratio of the shape's area to the square of the shape's perimeter. The ellipse variance is defined as the mapping error of a shape fitting an ellipse that has an equal covariance matrix as the shape [25]. The ratio of width to length is based on the distance between the shape's centroid and its contour points. It is computed as a ratio of the maximal distance between the centroid and the contour points to the minimal distance.

A minimum bounding rectangle (MBR) is the smallest rectangular region containing all points of the original shape. The area, perimeter, length and width of MBR can be used as measurements. Moreover, they can be combined either with the area of an original shape or with each other to create three different shape factors—rectangularity, eccentricity and elongation. Rectangularity is the ratio of the area of a shape to the area of its MBR. Eccentricity is calculated as the ratio of width to length of the MBR, whereby length represents the longer side of the MBR and width the shorter one. Elongation is then the value of eccentricity subtracted from 1 [25]. The calculation of width and height of a single bounding rectangle is quite obvious. For width we detect the shorter rectangle side—the distance between two certain rectangle corners. During the analysis of several bounding rectangles in a sequence we cannot constantly use the same points. As silhouette is changing, a shorter MBR side, calculated constantly for the same two points, can become a longer one. It affects experimental results therefore in our experiments we always consider the shorter side as MBR width.

Principal axes are two unique line segments that cross each other orthogonally within the centroid of a shape. The computation of eccentricity and elongation comprises several steps. Firstly, the covariance matrix C of the contour shape is calculated. Secondly, the lengths of the principal axes are equivalent to the

eigenvalues λ_1 and λ_2 of the matrix C . Furthermore, the eccentricity is the ratio of λ_2 to λ_1 , and the elongation is the eccentricity value multiplied by 2. We use the eccentricity value only.

4 Experimental Conditions and Results

4.1 Database

Binary silhouettes are usually obtained from the background subtraction process. Due to possible segmentation problems there is a risk of artefacts occurrence—obtained silhouettes may be disturbed by excess pixels or the lack of some (noisy data). Most common background subtraction drawbacks come from variable scene conditions: moving background, moving shadows, different lighting, foreground/background texture and colour similarity, etc. Another problematic issue is associated with the action itself—different people can perform a specific action in a different way and with different speed. We should also emphasize the importance of the manner how the video sequences were captured—lighting, camera point of view, resolution, video duration, etc. Despite several difficulties there are still many characteristics which enable the differentiation of actions. It is crucial to find shape features, or a combination of some, which are distinctive for a particular action.

In our research we use the Weizmann dataset [10] for action classification that contains binary masks in Matlab format. The original binary masks were obtained by background subtraction process from 90 video sequences (180×144 px, 50 fps). In our approach we change an original silhouette to its convex hull, centred in 144×144 pixel size image. Such change resulted in classification improvement by several percent. Exemplary convex hulls of silhouettes from a running action are depicted in Fig. 2.



Fig. 2. Exemplary convex hulls for running action—the person ran from the right to the left.

The actions under recognition belong to two subgroups. Actions performed in place are: bend, jump in place on two legs, jumping jack, wave one hand and wave two hands. In turn jump forward on two legs, run, gallop sideways, skip and walk are actions during which a person changes location.

4.2 Experimental Conditions

The aim of the experiments was to indicate the best combination of algorithms to be employed in the proposed method that results in highest classification accuracy. In each experiment the approach was tested using a different combination of algorithms for sequence representation and matching measures for classification. For shape representation we use 20 various simple shape descriptors, namely: area, perimeter, X Feret, Y Feret, X/Y Feret, Max Feret, compactness, roundness, circularity ratio, circle variance, ellipse variance, ratio of width to length, rectangularity, MBR elongation, MBR eccentricity, MBR area, MBR width, MBR length, MBR perimeter and PAM eccentricity. All shape descriptors were tested in each combination.

During the experiments each silhouette was represented using one simple shape descriptor and all descriptors for one sequence were combined into one vector. Such vector was transformed into final sequence representation using one of the following methods:

1. FFT's magnitude and periodogram.
2. FFT and periodogram.
3. Periodogram only.
4. FFT's phase angle and periodogram.
5. FFT for zero-padded vectors.
6. FFT for zero-padded vectors, but only half of the spectrum is taken.

All sequences were pre-classified into two subgroups—actions that are performed in place (static) and actions during which a person changes location (moving). The division is made using centroid location on the following frames. For example, for the second subgroup, the centroid would change its location on every frame while in the first subgroup the centroid trajectory would be very short. These trajectories are obtained using original, unaligned silhouette data.

The next step is performed in each subgroup separately. We take one sequence representation and match it with the rest of representations to indicate the most similar one by calculating similarity using correlation coefficient (or C1 correlation) or dissimilarity using Euclidean distance. The correct classification is when a tested representation is considered the most similar to other representation belonging to the same class. The percentage of correct classifications (in each subgroup and averaged) is a measure of the effectiveness of the experiment. The experiments were carried out using Matlab R2017b.

4.3 Results

The experimental results are presented in the following tables. Table 1 gives information on the highest percentage accuracy in each experiment that was obtained using a specific simple shape descriptor. The results are averaged for both subgroups (static and moving actions). It turned out that the best results are achieved when a combination of MBR width for shape representation is used, sequence representations are zero-padded to the maximum sequence length and

Fast Fourier Transform is applied, and subsequently transformed representations are classified using Euclidean distance. In this case the highest average classification accuracy equals 80.2%, whereof for actions performed in place the accuracy is 83.3% and for actions with silhouettes that change locations—77.1%.

Table 1. Highest average percentage accuracy for each experiment with the indication of the applied simple shape descriptor.

	Correlation coefficient	C1 correlation	Euclidean distance
FFT’s magnitude and periodogram	60.1% perimeter	60.4% X/Y Feret	62.4% X/Y Feret
FFT and periodogram	45.1% MBR width	55.5% MBR width	52.4% MBR width
Periodogram only	69.4% MBR width	72.0% MBR width	73.2% MBR width
FFT’s phase angle and periodogram	29.2% compactness	26.0% compactness	29.3% compactness
FFT for zero-padded vectors	70.5% PAM eccentricity	74.6% width/length ratio	80.2% MBR width
As above, but half the spectrum	69.6% MBR width	73.8% PAM eccentricity	77.2% PAM eccentricity

Figure 3 contains a graph illustrating percentage classification accuracy for all simple shape descriptors used in the experiment that brought the best results. Bars refer to the average classification accuracy, the dashed line shows the accuracy for action sequences on which a person performs an action in place, and the continuous line refers to the accuracy obtained for action sequences in which a silhouette is changing location within consecutive frames.

Due to the fact, that all sequences were pre-classified into two subgroups we are able to analyse the results separately, as if we had two datasets—with static and moving actions. Tables 2 and 3 contain percentage classification results obtained in each experiment together with the indication which shape feature was used. It turned out that the best results are obtained when we use different simple shape descriptors—MBR width should be used for actions performed in place (83.3%) and PAM eccentricity for actions during which a person changes location (85.4%). However, in both cases, it may be accepted to use the FFT for zero-padded vectors to obtain sequence representations, and Euclidean distance in classification step.

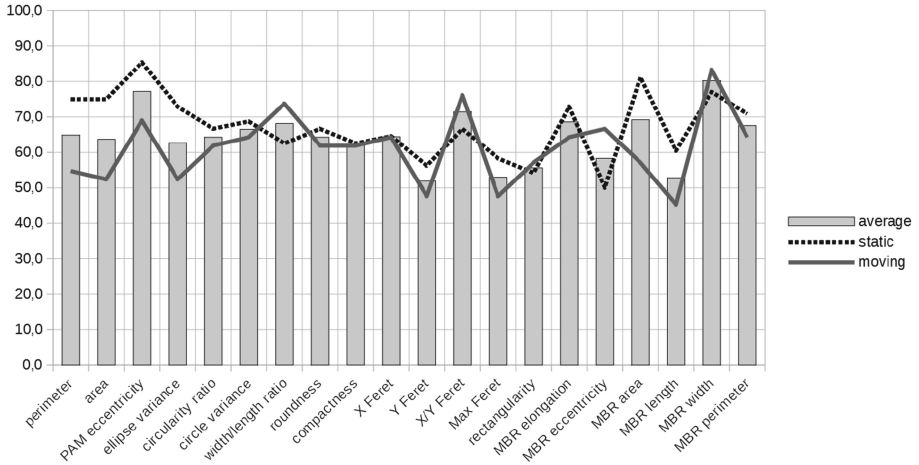


Fig. 3. Percentage classification accuracy for the best experiment using zero-padded vectors, Fast Fourier Transform and Euclidean distance. The highest accuracy is obtained when MBR width is used as a shape descriptor.

Table 2. Highest percentage accuracy for static actions with the indication of the applied simple shape descriptor.

	Correlation coefficient	C1 correlation	Euclidean distance
FFT's magnitude and periodogram	66.7% X/Y Feret	66.7% X/Y Feret	71.4% perimeter
FFT and periodogram	54.8% MBR width	57.1% PAM eccentricity	61.9% MBR width
Periodogram only	69.1% circle variance	76.2% width/length ratio	71.4% width/length ratio
FFT's phase angle and periodogram	35.7% MBR width	26.2% PAM eccentricity	35.7% compactness
FFT for zero-padded vectors	71.4% MBR width	76.2% width/length ratio	83.3% MBR width
As above, but half the spectrum	69.1% MBR width	76.2% width/length ratio	78.6% PAM eccentricity

Table 3. Highest percentage accuracy for moving actions with the indication of the applied simple shape descriptor.

	Correlation coefficient	C1 correlation	Euclidean distance
FFT’s magnitude and periodogram	62.5% MBR width	62.5% MBR area	62.5% Max Feret
FFT and periodogram	45.8% perimeter	56.2% X Feret	52.1% MBR area
Periodogram only	79.2% MBR width	81.3% area	81.3% PAM eccentricity
FFT’s phase angle and periodogram	33.4% Y Feret	37.5% circularity ratio	29.2% circularity ratio
FFT for zero-padded vectors	83.4% perimeter	83.4% PAM eccentricity	85.4% PAM eccentricity
As above, but half the spectrum	85.4% perimeter	83.4% PAM eccentricity	85.4% PAM eccentricity

5 Summary and Major Conclusions

The paper covered the topic of action recognition based on simple shape features. The proposed approach represents binary silhouettes using simple shape parameters or ratios to create sequence representations, and then classify these representations into action classes in two steps. Firstly, action sequences are pre-classified using the information about centroid trajectory into two subgroups—actions that are performed in place and actions during which a person moves in the frame. Secondly, action representations in each subgroup are classified using leave-one-out cross-validation technique. The highest percentage of correct classifications indicates the best combination of algorithms that should be applied in the proposed approach.

During the experiments it turned out that the best results are obtained when vectors are composed of the MBR width values, transformed using Fast Fourier Transform and then matched using Euclidean distance. The average percentage accuracy is 80.2%, whereof for actions performed in place the accuracy equals 83.3% and for actions with silhouette displacement—77.1%. However, in the second subgroup with moving actions it is more effective to use PAM eccentricity instead of the MBR width because the classification accuracy increases to 85.4%. Ultimately, due to the very small amount of information used for shape representation the results are very satisfactory and promising.

Future plans include the investigation of other sequence representation techniques which would eliminate the influence of unusual silhouettes disturbing the

typical action characteristics and would help to ease classification process by making representations equal in size. It is also worth checking some combinations of simple shape measures that would help to make shape representations more discriminative and invariant. Moreover, further research will focus on other classification methods as well.

References

1. Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **104**(2), 90–126 (2006)
2. Poppe, R.: A survey on vision-based human action recognition. *Image Vis. Comput.* **28**(6), 976–990 (2010)
3. Weinland, D., Ronfard, R., Boyer, E.: A survey of vision-based methods for action representation, segmentation and recognition. *Comput. Vis. Image Underst.* **115**(2), 224–241 (2011)
4. Borges, P.V.K., Conci, N., Cavallaro, A.: Video-based human behavior understanding: a survey. *IEEE Trans. Circ. Syst. Video Technol.* **23**(11), 1993–2008 (2013)
5. Cheng, G., Wan, Y., Saudagar, A.N., Namuduri, K., Buckles, B.P.: Advances in human action recognition: a survey. *CoRR* (2015)
6. Herath, S., Harandi, M., Porikli, F.: Going deeper into action recognition: a survey. *Image Vis. Comput.* **60**, 4–21 (2017)
7. Vishwakarma, S., Agrawal, A.: A survey on activity recognition and behavior understanding in video surveillance. *Vis. Comput.* **29**(10), 983–1009 (2013)
8. Chaquet, J.M., Carmona, E.J., Fernández-Caballero, A.: A survey of video datasets for human action and activity recognition. *Comput. Vis. Image Underst.* **117**(6), 633–659 (2013)
9. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. In: *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 3, pp. 32–36 (2004)
10. Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(12), 2247–2253 (2007)
11. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(3), 257–267 (2001)
12. Alp, E.C., Keles, H.Y.: Action recognition using MHI based Hu moments with HMMs. In: *IEEE EUROCON 2017–17th International Conference on Smart Technologies*, pp. 212–216 (2017)
13. Chen, D.Y., Shih, S.W., Liao, H.Y.M.: Human action recognition using 2-D spatio-temporal templates. In: *2007 IEEE International Conference on Multimedia and Expo*, pp. 667–670 (2007)
14. Baysal, S., Kurt, M.C., Duygulu, P.: Recognizing human actions using key poses. In: *2010 20th International Conference on Pattern Recognition*, pp. 1727–1730 (2010)
15. Chaaoui, A.A., Climent-Pérez, P., Flórez-Revuelta, F.: Silhouette-based human action recognition using sequences of key poses. *Pattern Recogn. Lett.* **34**(15), 1799–1807 (2013)
16. Islam, S., Qasim, T., Yasir, M., Bhatti, N., Mahmood, H., Zia, M.: Single- and two-person action recognition based on Silhouette shape and optical point descriptors. *Sig. Image Video Process.* **12**(5), 853–860 (2018)

17. Goudelis, G., Karpouzis, K., Kollias, S.: Exploring trace transform for robust human action recognition. *Pattern Recogn.* **46**(12), 3238–3248 (2013)
18. Al-Ali, S., Milanova, M., Al-Rizzo, H., Fox, V.L.: Human action recognition: contour-based and Silhouette-based approaches. In: Favorskaya, M.N., Jain, L.C. (eds.) *Computer Vision in Control Systems-2*. ISRL, vol. 75, pp. 11–47. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-11430-9_2
19. Jahagirdar, A.S., Nagmode, M.S.: Silhouette-based human action recognition by embedding HOG and PCA features. In: Bhalla, S., Bhateja, V., Chandavale, A.A., Hiwale, A.S., Satapathy, S.C. (eds.) *Intelligent Computing and Information and Communication*. AISC, vol. 673, pp. 363–371. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7245-1_36
20. Vishwakarma, D.K., Gautam, J., Singh, K.: A robust framework for the recognition of human action and activity using spatial distribution gradients and Gabor wavelet. In: Reddy, M.S., Viswanath, K., K.M., S.P. (eds.) *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications*. AISC, vol. 628, pp. 103–113. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-5272-9_10
21. Gościewska, K., Frejlichowski, D.: Action recognition using silhouette sequences and shape descriptors. In: Choraś, R.S. (ed.) *Image Processing and Communications Challenges 8*, pp. 179–186. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47274-4_21
22. Gościewska, K., Frejlichowski, D.: Moment shape descriptors applied for action recognition in video sequences. In: Nguyen, N.T., Tojo, S., Nguyen, L.M., Trawiński, B. (eds.) *Intelligent Information and Database Systems*, pp. 197–206. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54430-4_19
23. Brunelli, R., Messelodi, S.: Robust estimation of correlation with applications to computer vision. *Pattern Recogn.* **28**(6), 833–841 (1995)
24. Yang, L., Albrechtsen, F., Lønnestad, T., Grøttum, P.: Methods to estimate areas and perimeters of blob-like objects: a comparison. In: *Proceedings of IAPR Workshop on Machine Vision Applications*, pp. 272–276 (1994)
25. Yang, M., Kpalma, K., Ronsin, J.: A survey of shape feature extraction techniques. In: Yin, P.-Y. (ed.) *Pattern Recognition*, pp. 43–90. INTECH Open Access Publisher (2008)