applied sciences

MDPI

Article

The Analysis of Shape Features for the Purpose of Exercise Types Classification Using Silhouette Sequences

Katarzyna Gościewska * **and Dariusz Frejlichowski**

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

* Correspondence: kgosciewska@wi.zut.edu.pl

Received: 04 August 2020; Accepted: 23 September 2020; Published: 25 September 2020

check for updates

Abstract: This paper presents the idea of using simple shape features for action recognition based on binary silhouettes. Shape features are analysed as they change over time within an action sequence. It is shown that basic shape characteristics can discriminate between short, primitive actions performed by a single person. The proposed approach is tested on the Weizmann database using a various number of classes. Binary foreground masks (silhouettes) are replaced with convex hulls, which highlights some shape characteristics. Centroid locations are combined with some other simple shape descriptors. Each action sequence is represented using a vector with shape features and Discrete Fourier Transform. Classification is based on leave-one-sequence-out approach and employs Euclidean distance, correlation coefficient or C1 correlation. A list of processing steps for action recognition is explained and followed by some experiments that yielded accuracy exceeding 90%. The idea behind the presented approach is to develop a solution for action recognition that could be applied in a kind of human activity recognition system associated with the Ambient Assisted Living concept, helping adults increasing their activity levels by monitoring them during exercises.

Keywords: action recognition; silhouette sequences; shape features; ambient assisted living; active ageing

1. Introduction

Human Activity Recognition (HAR) based on the video content analysis approaches is gaining more and more interest thanks to the wide variety of possible applications, such as video surveillance, human-computer interfaces or monitoring of patients and elderly people in their living environments. An exemplary structure of the HAR system may consist of the following general modules: motion segmentation, object classification, human tracking, action recognition and semantic description [1]. If a focus is put to action recognition (exercise classification), it can be assumed that input data type, localised objects and their positions are known. Based on the taxonomy presented in [2] the techniques for action recognition are divided into holistic and local representations. Holistic solutions use global representations of human shape and movement, accumulating several features. The most popular solutions include Motion History Image and Motion Energy Image templates proposed by Bobick and Davis [3] or Space-Time Volume representation introduced by Yilmaz and Shah [4]. Local representations usually are based on interest points which are used to extract a set of local descriptors, e.g., Space-Time Interest Points by Laptev [5]. Instead of aggregating features from all frames, some researchers propose to extract only several foreground silhouettes, so called key poses (e.g., [6,7]). If binary silhouettes are used as input data, various shape features can be extracted and combined, such as shape and contour [8], orientation [9] or skeleton [10]. Apart from traditional approaches, more challenging tasks can benefit from the application of deep learning techniques, such as Convolutional Neural Networks [11]. Ultimately, the choice of methods is dependant, among others, on the application scenario and data complexity.

A task of exercise classification can be related to the Ambient Assisted Living (AAL), which refers to concepts, products and services introducing new technologies for people in all phases of life, allowing them to stay healthy, independent, safe and well-functioning at their living environment. In the era of an ageing society and a significant proportion of older people living alone or unattended, expanding the range of care support options is becoming more and more important. Another major focus of AAL is prolonging the time people can live on their own, being in good health and in good physical shape. This is related to the increasing life expectancy and successful ageing. The World Health Organisation policy in Active Ageing applies to physical, mental and social well-being, and is defined in [12] as "the process of optimizing opportunities for health, participation and security in order to enhance quality of life as people age". Among people aged 45 and over, non-communicable diseases (NCDs) are the most frequent causes of mortality and disability all over the world. The risk of NCDs morbidity is higher in this age group; however, risk factors may originate in younger years. NCDs include, among others, cardiovascular diseases, hypertension, diabetes, chronic obstructive pulmonary disease, musculoskeletal conditions and mental health conditions. One of the risk factors is a sedentary lifestyle [12]. Low level of physical activity and lack of exercises can directly lead to obesity which increases the risk of NCDs as well. The study presented in [13] shows that greater physical fitness is associated with reduced risk of developing many NCDs. The authors of [14] advise promoting positive health behaviour rather than reducing negative ones, such as above-mentioned sedentary lifestyle. It is recommended to focus on the benefits of physical activity, provide motivation and promote self-care. Models based on social-cognitive behavioural theory are indicated as self-regulatory strategies that can contribute to increasing physical activity based on skills such as goal setting and self-monitoring of progress.

This paper follows the idea of active ageing and the use of activity monitoring solutions. However, the approach which is here proposed aims only at recognizing primitive actions that resemble some recommended exercise types [15], such as resistance, aerobic, stretching, balance and flexibility exercises. A specific scenario is assumed in which a person wants to do a workout in front of a laptop where a video with exercises is displayed. The laptop camera captures people's activities and the algorithm analyses them. The classification is performed in order to determine the amount, frequency and duration of a specific exercise. This, in some way, may encourage a person to engage in more physical activity. Due to presented reasoning, Section 2. Related works is focused on the methods and techniques used in action recognition approaches based on video content analysis. In our approach, we use foreground masks extracted from video sequences, each representing single person performing an action. Foreground masks carry information about an object's pose, shape and localisation. Therefore, various features can be retrieved and combined in order to create an action representation—here it is proposed to combine trajectory, simple shape descriptors and Discrete Fourier Transform (DFT).

The rest of the paper is organised as follows: Section 2 presents selected related works, that concern the recognition and classification of similar actions. Section 3 explains consecutive steps of the proposed approach together with applied methods and algorithms. Section 4 describes experimental conditions and presents the results. Section 5 discusses the results and concludes the paper.

2. Related Works

An action can be defined as a single person short activity composed of multiple gestures organised in time that lasts up to several seconds or minutes [1,16], e.g., running, walking or bending. Many actions can be performed for a longer time than several minutes, however due to their periodic characteristic only a short action span is used for recognition. The recognition process is here understood as assigning action labels to sequences of images [17]. Then, action classification can be based on various features, such as colour, grey levels, texture, shape or characteristic points like centroid or contour. Selected features are numerically represented using specific description algorithms in a form of so-called representation or descriptor. According to [2], good representation for action recognition has to be easy to calculate, provide a description for as many classes as possible and reflect similarities between look-alike actions. There is a large body of literature on video-based action recognition and related topics investigating wide variety of methods and algorithms using diverse features. An interest is reflected in the still emerging surveys and reviews (e.g., [2,18–22]). Due to many techniques on action classification reported in the literature, here we refer to several works that correspond to our interests in terms of the methods and data used.

The authors of [23] propose a novel pose descriptor based on the human silhouette, called Histogram of Oriented Rectangles. The human silhouette is represented by a set of oriented rectangular patches, and the distribution of these patches is represented as oriented histograms. Histograms are classified by different techniques, such as nearest neighbour classification, Support Vector Machine (SVM) and Dynamic Time Warping (DTW), among which the last one turned out to be the most accurate. Another silhouette based feature was proposed in [24] which uses Trace transform for a set of silhouettes representing single period of action. The authors introduce two feature extraction methods: History Trace Templates (a sequence representation with spatio-temporal information) and History Triple Features (a set of invariant features calculated for every frame). The classification is performed using Radial Basis Function Kernel SVM and Linear Discriminant Analysis is applied for dimensionality reduction. Action recognition based on silhouettes is presented in [25] as well. All silhouettes in a sequence are represented as time series (using a rotation invariant distance function) and each of them is transformed into so-called Symbolic Aggregate approXimation (SAX). An action is then represented by a set of SAX vectors. The model is trained using the random forest method and various classification methods are tested. The authors of [26] propose a novel feature for action recognition based on silhouette contours only. A contour is divided into radial bins of the same angle using centroid coordinates and a summary value is obtained for each bin. A summary value (variance, max value or range) depends on Euclidean distances from centroid to contour points in every radial bin. The proposed feature is used together with a bag of key poses approach and tested in single- and multi-view scenarios using DTW and leave-one-out procedure.

The authors of [8,9] use accumulated silhouettes (all binary masks of an action sequence are compressed into one image) instead of every silhouette separately. In [8] various contour- and region-based features are combined, such as Cartesian Coordinate Features and Histogram of Oriented Gradients (HOG). SVM and K-nearest neighbour (KNN) classifiers are used (the latter one in two scenarios). In total, seven different features and three classifiers are experimentally tested. The highest accuracy is reported for a combination of HOG and KNN in leave-one-sequence-out scenario. In [9] an average energy silhouette image is calculated for each sequence. Then region of interest is detected and several features are calculated: edge distribution of gradients, directional pixels and rotational information. These feature vectors are combined in action representations which are then classified using SVM classifier. In [10], instead of accumulated silhouettes, the authors propose the cumulative skeletonised image—all foreground objects' skeletons of each action sequence are aligned to the centroid and accumulated into one image. Action features are extracted from these cumulative skeletonised images. In an off-line phase the most discriminant human body regions are selected and classified in an online phase using SVM. The authors of [27] propose a motion descriptor which describes patterns of neighbouring trajectories. Two-level occurrence analysis is performed to discover motion patterns of trajectory points. Actions are classified using SVM with different kernels or random forest algorithm. The approach proposed in [28] employs spectral domain features for action classification, however silhouette features are not involved. Instead, the two-dimensional Discrete Fourier Transform is applied to each video frame and a part of high amplitude coefficients is taken. For a given sequence, selected coefficients of all frames are concatenated into action representation. Larger representations can be reduced using Principal Component Analysis. Action classification is performed using SVM or a simple classifier based on Euclidean distance.

3. Materials and Methods

An action recognition procedure is proposed. It is based on simple features extracted from the entire silhouette and its characteristic points, which are combined into action representation. The proposed approach applies our previous findings and recent research, aiming at improving the results presented in [29]. The dataset and the general processing steps are the same and will be explained in the following subsections. Several changes at the data preprocessing step are introduced, and a new parameter is added for the action representation method that previously yielded the highest accuracy.

3.1. Data Preprocessing

Due to the use of the Weizmann database [30] it is assumed that for each video sequence there is a set of binary images, and these images are foreground masks extracted from video frames. One sequence represents one action type and one image contains one silhouette. Frames in which an object is occluded or too close to the edge of the video frame are removed. The direction of the action is checked and, if necessary, the video frames are flipped so that all objects in the sequence move from left to right. Then, each silhouette is replaced with its convex hull, which reduces the impact of some artefacts (e.g., additional pixels) introduced during background subtraction (see Figure 1 for examples). It is indicated in [29] that the use of convex hulls improves classification accuracy.

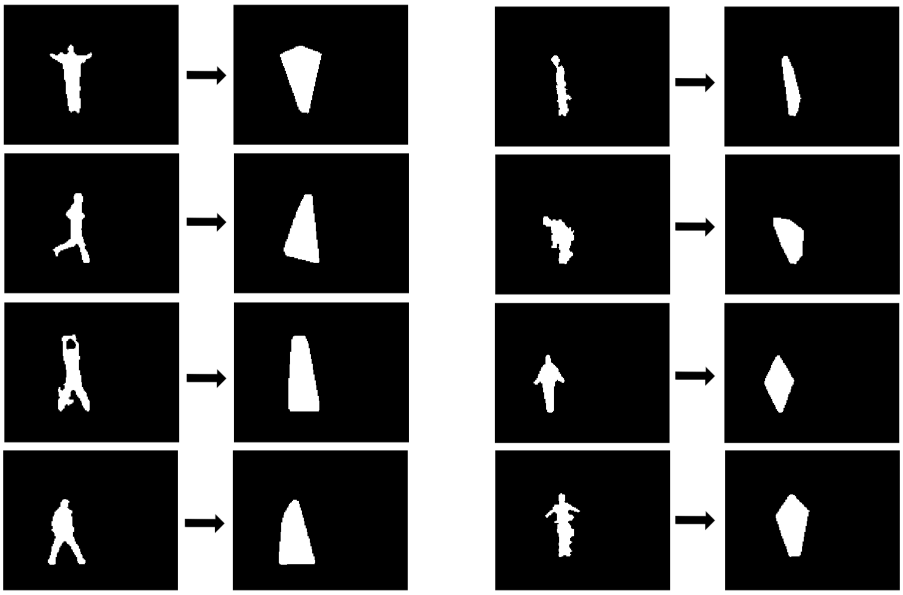

Figure 1. Sample silhouettes from the Weizmann database [30] and the corresponding convex hulls.

Before the actual classification, the dataset is divided into two subsets based on the centroid locations on the consecutive frames. This is related to action characteristics—some of actions are performed by a person standing in place (short trajectory) and the rest contain a person who changes location in every frame (long trajectory). Examples are given in Figure 2. This procedure can be called a coarse classification. It influences subsequent steps of the approach which are performed separately in each subgroup. Therefore, there is a possibility of selecting different features and parameters better suited to the specific action types.

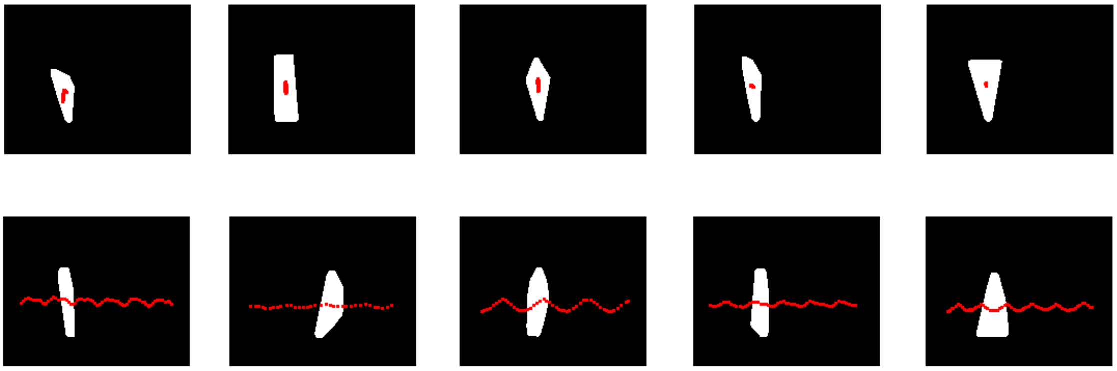

Figure 2. Exemplary trajectories for ten different actions of one actor: actions performed in place are in the top row (bending, jumping-jack, jumping in place, one-hand waving and two-hand waving) and actions with changing location of a silhouette are depicted in the bottom row (jumping forward, running, galloping sideways, skipping and walking). Centroid trajectory is displayed over sample frame from a corresponding video sequence.

3.2. Shape Representation

In this step, each image from the dataset is represented as a single number using a selected shape description algorithm—each number is a simple shape descriptor. The descriptors of all frames from a sequence are combined into one vector and values are normalized to [0, 1] range. This makes it easy to observe how the individual shape features change over time and how they differ between actions. Figure 3 depicts example vectors as line graphs using very simple feature which is an area of a convex hull. Each input action sequence can be denoted as a set of binary masks $BM_i = \{bm_1, bm_2, ..., bm_n\}$, which is represented by a set of normalized descriptors $SD_i = \{sd_1, sd_2, ..., sd_n\}$, and n is the number of frames in a particular sequence.

Simple shape descriptors are basic shape measurements and shape ratios, often used to describe general shape characteristics. A shape measurement is a relative value dependent on the scale of the object. Shape ratio is an absolute value that can be calculated using some shape measurements. Selected simple shape descriptors are listed below (based on [31–34]):

- Area and perimeter, as the number of pixels belonging to the shape's region or contour respectively.
- Feret measures (Feret diameters):
  - X Feret and Y Feret, the distances between the minimal and maximal horizontal and vertical coordinates of a contour respectively;
  - X/Y Feret, the ratio of the X Feret to Y Feret;
  - Max Feret, the maximum distance between any two points of a contour.
- Shape factors:
  - Compactness, the ratio of the square of the shape's perimeter to its area;
  - Roundness, measures shape's sharpness based on area and perimeter;
  - Circularity ratio, defines how a shape is similar to a circle. It can be estimated as the ratio of the shape's area to the shape's perimeter square. It is also called a circle variance and calculated based on the mean and standard deviation obtained using distances from centroid to the contour points;

  - Ellipse variance, defines how a shape is similar to an ellipse and can be estimated as a mapping error of a shape fitting an ellipse where both have the same covariance matrix. Similarly to circle variance, mean and standard deviation are used;
  - Width/length ratio, the ratio of the maximal to the minimal distance based on distances between centroid and contour points.
- Minimum bounding rectangle (MBR)—defines a smallest rectangular region that contains all points of a shape. A MBR can be measured in different ways and some ratios can be calculated:
  - MBR measurements, which include area, perimeter, length and width. Length and width can be calculated based on specific pairs of MBR corner points, however in our experiments we always consider the shorter MBR side as its width;
  - Rectangularity, the ratio of the area of a shape to the area of its MBR;
  - Eccentricity, the ratio of width to length of the MBR (length is the longer side of the MBR and width is the shorter one);
  - Elongation, a value of eccentricity subtracted from 1.
- Principal axes method (PAM), which defines two unique line segments that cross each other orthogonally within a shape's centroid. The lengths of the principal axes are used to calculate eccentricity which is the measure of aspect ratio.

3.3. Action Representation

In the next step, all SD vectors are transformed into action representations (AR) using the Discrete Fourier Transform. A SD, in its form, is similar to shape signature and the one-dimensional version of the Discrete Fourier Transform can be applied. The number of elements in each SD is different due to various number of frames in video sequences. Therefore, to prepare action representations equal in size, the N-point Discrete Fourier Transform is calculated, where N is the predefined number of resultant Fourier coefficients. If N is larger than n, then SD vectors are appended with zeros in the time domain (zero-padding) which corresponds to the interpolation in the frequency domain. Otherwise, SD vectors are truncated and then Fourier coefficients are calculated. As a result, each AR contains N absolute values of Fourier coefficients. Usually, it was recommended that the vectors under transformation should have a length equal to a power of 2, due to the computational complexity. However, current implementations of the Discrete Fourier Transform can handle arbitrary size transforms, e.g., Fast Fourier Transform algorithm available in the FFTW library [35].

3.4. Final Classification

For action classification a standard leave-one-out cross-validation procedure is adopted. In each iteration, one sequence is left out and matched with the rest of sequences based on AR vectors. An AR which resulted to be the most similar (or less dissimilar) to the one under processing indicates its class. Indications from all iterations are verified with the original action labels and the percentage of correctly classified objects is taken (classification accuracy). For matching, three different measures are applied, namely Euclidean distance [36], correlation coefficient based on Pearson's correlation [37] and C1 correlation based on L1-norm [38].

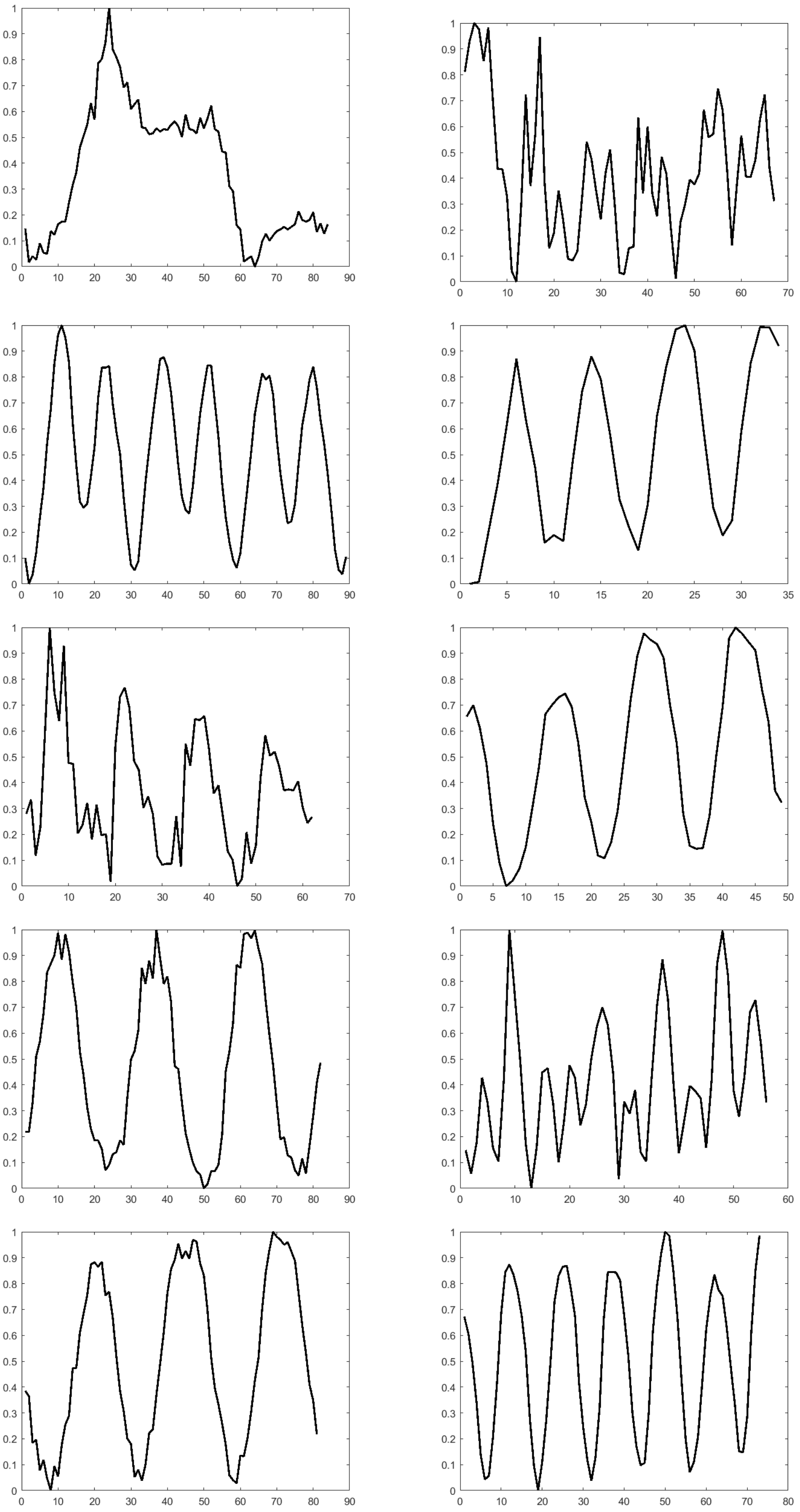

Figure 3. Line graphs showing normalized area values for actions presented in Figure 2. The X axis corresponds to the consecutive frame numbers, while the Y axis corresponds to the normalized area values of the foreground object in each frame. Line graphs in the left column correspond to the actions performed in place (bending, jumping-jack, jumping in place, one-hand waving and two-hand waving), and line graphs for actions with changing location of a silhouette are depicted in the right column (jumping forward, running, galloping sideways, skipping and walking).

4. Experimental Conditions and Results

The experiments were performed with the use of the Weizmann dataset [30], which consists of short video sequences that last up to several seconds (144 × 180 px, 50 fps). Foreground binary masks extracted from the database were made available by its authors and are here used as input data. There are masks for 93 sequences, however three of them are removed—one actor doubled three actions by moving with two different directions. In result, the database has 10 action classes and each action type is performed by nine different actors. During the experiments we follow the data processing steps presented in the previous section. Firstly, each frame is preprocessed individually and then all sequences are divided into two subsets—actions performed in place and actions with changing location of a silhouette. After preprocessing the number of images in a sequence varies from 28 to 146. The group of actions performed in place consists of five action classes: 'bend', 'jump in place', 'jumping jack', 'wave one hand' and 'wave two hands', whereas actions with changing location of a silhouette are: 'jump forward on two legs', 'run', 'skip', 'walk' and 'gallop sideways'. Figure 4 depicts some selected masks after preprocessing step (for two different actions). The next steps, including shape description, action representation and action classification, are performed separately in each subgroup. Ultimately, the classification accuracy values of both subgroups are averaged, which gives the final effectiveness of the approach.

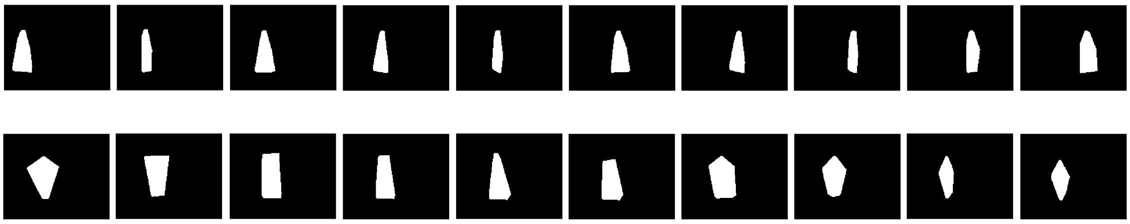

Figure 4. Exemplary preprocessed masks for 'walk' (**top row**) and 'jumping jack' (**bottom row**) actions.

The main part of the experiments refers to the assumed application scenario, which is related to the Ambient Assisted Living and the concept of active ageing. In this scenario, a human activity analysis is performed to identify types of exercises. Physical activity is indicated as one of the methods of preventing the risk of developing non-communicable diseases. Based on the social-cognitive behavioural theory it is advised to promote self-care and incorporate self-monitoring of progress. Nowadays, video content analysis techniques and the popularity of cameras (in laptops, smartphones) facilitate the implementation of exercise monitoring solutions. In order to carry out the experiment concerning the recognition of types of exercises, we composed a database using selected classes from the Weizmann database. Action classes were compared with the recommended exercises presented in [15]. In addition, it was taken into account that exercises are supposed to be performed in a home environment. Due to that, the 'run' class is excluded. Moreover, there are two classes with waving action, therefore the 'wave with one hand' class is excluded as well. The remaining action classes may be associated with the following exercises (based on [15]):

- Aerobic/endurance exercise, in which the body's large muscles move in a rhythmic manner (e.g., walking, skipping, jumping jack);
- Balance training, which includes various activities that increase lower body strength (e.g., galloping sideways, jumping in place, jumping forward on two legs);
- Flexibility exercise, which preserves or extends motion range around joints (e.g., waving, bending).

Several experiments were carried out to investigate the best combination of methods and parameters for the approach. Twenty simple shape descriptors were tested in combination with three matching measures and the use of up to 256 Fast Fourier Transform coefficients. In order to focus only on the highest results and be able to appropriately present them, for each matching measure an experiment is performed with several tests, in which a selected simple shape descriptor and different number of coefficients are used. The results of exercise recognition are provided in Table 1. The best result is considered as the highest accuracy and the smallest action representation. The highest accuracy for actions performed in place is 100% if MBR width is used and the action representation contains 54 elements. It means that each action sequence, regardless of the number of frames, is represented using 54 Fast Fourier Transform coefficients and a representation has a form of a vector with 54 real values. The matching process can be then performed using Euclidean distance or C1 correlation. In total, 100% accuracy is also obtained for X/Y Feret, but more DFT coefficients are required. Actions with changing location of a silhouette are most successfully classified if MBR area is used and action representation contains 32 values—an accuracy of 94.44% is yielded. Again, either Euclidean distance or C1 correlation can be employed. Ultimately, the averaged correct classification rate for exercise types recognition is 97.2% (8 action classes).

Table 1. Experimental results for the recognition of exercise types using 8 classes of the Weizmann dataset. The results are presented separately for actions performed in place and actions with changing location of a silhouette. The highest accuracy values are listed with the indication of the applied simple shape descriptor and the size of an action representation (given in brackets).

8 Classes	Actions Performed in Place	Actions with Changing Location of a Silhouette
Euclidean distance	100.00% MBR width (54)	94.44% MBR area (32)
Correlation Coefficient	97.22% circle variance (48)	83.33% MBR perimeter (35)
C1 Correlation	100.00% MBR width (54)	94.44% MBR area (32)

A second set of experiments concerned the use of the Weizmann database as a benchmark and the comparison of the results for 10 classes with the previous version of our approach, presented in [29]. The results are presented in Table 2. For actions performed in place the highest accuracy is 86.67% (MBR perimeter, 52 DFT coefficients, Euclidean distance) and for actions with changing location of a silhouette the accuracy equals 95.56% (MBR area, 33 DFT coefficients, C1 correlation). If the use of different methods and parameters for each subgroup is assumed, the averaged accuracy for the entire database is 91.12%. This outperforms our previous approach based on simple shape descriptors, that resulted in 83.3% accuracy for actions performed in place (MBR width) and 85.4% (PAM eccentricity) for the other subgroup. The averaged accuracy equalled then 84.35%, which means that the current approach improves the accuracy by nearly 7%.

Table 2. Experimental results for the Weizmann dataset used as a benchmark, presented separately for actions performed in place and actions with changing location of a silhouette. The highest accuracy values are listed with the indication of the applied simple shape descriptor and the size of an action representation (given in brackets).

10 Classes	Actions Performed in Place	Actions with Changing Location of a Silhouette
Euclidean distance	86.67% MBR perimeter (52)	91.11% MBR area (32)
Correlation Coefficient	86.67% width/length ratio (53)	84.44% perimeter and ellipse variance (66)
C1 Correlation	82.22% MBR perimeter (56)	95.56% MBR area (33)

5. Discussion and Conclusions

In the second section of the paper, a description of related works is given. The methods described there, that is [8–10,23–28], were chosen for two main reasons—they concern action recognition and use the Weizmann database. However, the approaches used to represent a frame or a silhouette are diverse. In [23] a set of rectangular patches is used to represent a shape and in [26] only contour points are applied. Some researchers use various transforms, e.g., the authors of [24] apply Trace transform to binarized silhouettes, while in [28] the two-dimensional Fourier transform is applied to the original frames. The opposite approach is the use of cumulative silhouettes [8,9] or cumulative skeletons [10]. Some other techniques are dense trajectories based on salient points [27] and time series [25].

The approach proposed in this paper combines simple shape descriptors with the one-dimensional Fourier transform and standard leave-one-out classification procedure. Each action sequence is firstly described by a set of simple features and represented using a predefined number of Fourier coefficients. Classification is two-stage: firstly, actions are divided into two subgroups based on trajectory length, and secondly, leave-one-sequence-out cross-validation is performed. The proposed approach yields 97.2% accuracy in the assumed application scenario and 91.12% accuracy on the entire Weizmann database. The best results were obtained with the use of features based on a minimum bounding rectangle—its area, width and perimeter. These features are simple; however, if observed over time, they carry much more information about an action. Therefore, the input data can be limited to rectangular objects of interest, representing regions where silhouettes are located. These areas can be tracked over time to extract centroid locations. With these assumptions, the calculation of convex hulls may be omitted.

A comparison of some recognition rates of the proposed approach to other methods tested on the Weizmann dataset is presented in Table 3. Although our approach does not provide a perfect accuracy, it can be compared with some other solutions. It should be mentioned that the presented methods may assume other application scenarios and experimental conditions. Moreover, if we limit the number of classes, it does not always improve the results, which is proven in our experiments. When the classification of actions with changing location of a silhouette is performed for 10 classes, the highest accuracy is 95.56%, while for the limited number of classes it decreases to 94.44%. According to that, we especially refer to the results presented in [8,23,28] that outperformed our results obtained in the experiment concerning the assumed application scenario. The authors of [28] also employ spectral domain features, however these features are extracted from video frames using the two-dimensional Fourier Transform. In our approach the frames are represented using simple shape descriptors, which for each sequence are concatenated into a vector, and the one-dimensional Fourier Transform is applied. Therefore, the initial data dimensionality is lower. The descriptor proposed in [23] requires the extraction of rectangular regions from a human silhouette which may be problematic in case of imperfect silhouettes. The approach proposed in [8] uses accumulated silhouette representation, which requires all foreground masks from an action sequence. In our approach each foreground mask is represented separately, therefore in the case of the real-time scenario the proposed approach can be adjusted to utilise fewer frames.

The proposed approach has some advantages—it can be adapted to different action types by selecting other shape features and matching measures. Action representations are small and easy to calculate because simple algorithms are applied. Moreover, if another distinctive feature is found, instead of centroid or in addition to it, the recognition space could be narrowed in a more efficient manner and eliminate misclassifications. The use of different methods in each subgroup improves overall results. The presented version of the approach is promising; however, an improvement is needed. Our future works include experiments using other databases with larger number of classes corresponding to different exercises. Moreover, recently popular solutions based on deep learning will be tested as well.

Table 3. Comparison of recognition rates obtained on the Weizmann database (cited methods are explained in Section 2. Related works).

Reference	Number of Actions	Accuracy
[28]	10	100%
[23]	9 (without skip)	100%
[8]	10 (93 videos)	98.24%
Proposed	8	97.20%
[9]	10	96.64%
[24]	10	95.42%
[26]	10	93.50%
Proposed	10	91.12%
[25]	10	89.00%
[10]	10	87.52%
[27]	10	78.88%

Author Contributions: Conceptualization, K.G. and D.F.; methodology, K.G.; software, K.G., validation, K.G. and D.F.; investigation, K.G.; writing—original draft preparation, K.G.; writing—review and editing, K.G. and D.F.; visualization, K.G.; supervision, D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vishwakarma, S.; Agrawal, A. A survey on activity recognition and behavior understanding in video surveillance. *Vis. Comput.* **2013**, *29*, 983–1009. [CrossRef]
2. Herath, S.; Harandi, M.; Porikli, F. Going deeper into action recognition: A survey. *Image Vis. Comput.* **2017**, *60*, 4–21. [CrossRef]
3. Bobick, A.; Davis, J. The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 257–267. [CrossRef]
4. Yilmaz, A.; Shah, M. Actions sketch: A novel action representation. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 984–989.
5. Laptev, I. On Space-Time Interest Points. *Int. J. Comput. Vis.* **2005**, *64*, 107–123. [CrossRef]
6. Baysal, S.; Kurt, M.C.; Duygulu, P. Recognizing Human Actions Using Key Poses. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 1727–1730.
7. Chaaraoui, A.A.; Climent-Pérez, P.; Flórez-Revuelta, F. Silhouette-based human action recognition using sequences of key poses. *Pattern Recognit. Lett.* **2013**, *34*, 1799–1807. [CrossRef]
8. Al-Ali, S.; Milanova, M.; Al-Rizzo, H.; Fox, V.L., Human Action Recognition: Contour-Based and Silhouette-Based Approaches. In *Computer Vision in Control Systems-2: Innovations in Practice*; Favorskaya, M.N., Jain, L.C., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 11–47. [CrossRef]
9. Vishwakarma, D.; Dhiman, A.; Maheshwari, R.; Kapoor, R. Human Motion Analysis by Fusion of Silhouette Orientation and Shape Features. *Procedia Comput. Sci.* **2015**, *57*, 438–447. [CrossRef]
10. Mliki, H.; Rabàa, Z.; Mohamed, H. Human action recognition based on discriminant body regions selection. *Signal Image Video Process.* **2018**, *12*, 845–852. [CrossRef]
11. Yao, G.; Lei, T.; Zhong, J. A review of Convolutional-Neural-Network-based action recognition. *Pattern Recognit. Lett.* **2019**, *118*, 14–22. [CrossRef]
12. World Health Organization. Active Ageing: A Policy Frame-Work. 2002. Available online: http://www.who.int/ageing/publications/active_ageing/en/ (accessed on 15 July 2020).
13. Ross, R.; Blair, S.; Arena, R.; Church, T.; Després, J.P.; Franklin, B.; Kaminsky, L.; Levine, B.; Lavie, C.; Myers, J.; et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation* **2016**, *134*, e653–e699. [CrossRef]
14. Lachman, M.; Lipsitz, L.; Lubben, J.E.; Castaneda-Sceppa, C.; Jette, A.M. When Adults Don't Exercise: Behavioral Strategies to Increase Physical Activity in Sedentary Middle-Aged and Older Adults. *Innov. Aging* **2018**, *2*, igy007. [CrossRef]

15. Thaxter-Nesbeth, K.; Facey, A. Exercise for Healthy, Active Ageing: A Physiological Perspective and Review of International Recommendations. *West Indian Med. J.* **2018**, *67*, 351–356. [CrossRef]
16. Chaaraoui, A.A.; Climent-Pérez, P.; Flórez-Revuelta, F. A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living. *Expert Syst. Appl.* **2012**, *39*, 10873–10888. [CrossRef]
17. Poppe, R. A survey on vision-based human action recognition. *Image Vis. Comput.* **2010**, *28*, 976–990. [CrossRef]
18. Aggarwal, J.; Ryoo, M. Human Activity Analysis: A Review. *ACM Comput. Surv.* **2011**, *43*, 16. [CrossRef]
19. Borges, P.V.K.; Conci, N.; Cavallaro, A. Video-Based Human Behavior Understanding: A Survey. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1993–2008. [CrossRef]
20. Cheng, G.; Wan, Y.; Saudagar, A.N.; Namuduri, K.; Buckles, B.P. Advances in Human Action Recognition: A Survey. *arXiv* **2015**, arXiv:1501.05964.
21. Zhang, H.B.; Zhang, Y.X.; Zhong, B.; Lei, Q.; Yang, L.; Du, J.X.; Chen, D.S. A Comprehensive Survey of Vision-Based Human Action Recognition Methods. *Sensors* **2019**, *19*, 1005. [CrossRef]
22. Rodríguez-Moreno, I.; Martinez-Otzeta, J.M.; Sierra, B.; Rodriguez Rodriguez, I.; Jauregi Iztueta, E. Video Activity Recognition: State-of-the-Art. *Sensors* **2019**, *19*, 3160. [CrossRef]
23. Ikizler, N.; Duygulu, P. Histogram of oriented rectangles: A new pose descriptor for human action recognition. *Image Vis. Comput.* **2009**, *27*, 1515–1526. [CrossRef]
24. Goudelis, G.; Karpouzis, K.; Kollias, S. Exploring trace transform for robust human action recognition. *Pattern Recognit.* **2013**, *46*, 3238–3248. [CrossRef]
25. Junejo, I.N.; Junejo, K.N.; Aghbari, Z.A. Silhouette-based human action recognition using SAX-Shapes. *Vis. Comput.* **2014**, *30*, 259–269. [CrossRef]
26. Chaaraoui, A.; Flórez-Revuelta, F. A Low-Dimensional Radial Silhouette-Based Feature for Fast Human Action Recognition Fusing Multiple Views. *Int. Sch. Res. Not.* **2014**, *2014*, 1–11. [CrossRef] [PubMed]
27. Garzon Villamizar, G.; Martinez, F. A Fast Action Recognition Strategy Based on Motion Trajectory Occurrences. *Pattern Recognit. Image Anal.* **2019**, *3*, 447–456. [CrossRef]
28. Imtiaz, H.; Mahbub, U.; Schaefer, G.; Zhu, S.Y.; Ahad, M.A.R. Human Action Recognition based on Spectral Domain Features. *Procedia Comput. Sci.* **2015**, *60*, 430–437. [CrossRef]
29. Gościewska, K.; Frejlichowski, D. Silhouette-Based Action Recognition Using Simple Shape Descriptors. In Proceedings of the International Conference, ICCVG 2018, Warsaw, Poland, 17–19 September 2018; pp. 413–424. [CrossRef]
30. Blank, M.; Gorelick, L.; Shechtman, E.; Irani, M.; Basri, R. Actions As Space-Time Shapes. In Proceedings of the Tenth IEEE International Conference on Computer Vision—Volume 2, ICCV '05, Beijing, China, 17–21 October 2005; IEEE Computer Society: Washington, DC, USA, 2005; pp. 1395–1402. [CrossRef]
31. Yang, L.; Albregtsen, F.; Lønnestad, T.; Grøttum, P. Methods to estimate areas and perimeters of blob-like objects: A comparison. In Proceedings of the IAPR Workshop on Machine Vision Applications, Kawasaki, Japan, 13–15 December 1994; pp. 272–276.
32. Rosin, P. Computing global shape measures. In *Handbook of Pattern Recognition and Computer Vision*; World Scientific Publishing Co. Pte. Ltd.: Singapore, 2005; pp. 177–196. [CrossRef]
33. Zhang, D.; Lu, G. Review of shape representation and description techniques. *Pattern Recognit.* **2004**, *37*, 1–15. [CrossRef]
34. Yang, M.; Kpalma, K.; Ronsin, J. A Survey of Shape Feature Extraction Techniques. *Pattern Recognit.* **2008**, *15*, 43–90.
35. Frigo, M.; Johnson, S.G. The Design and Implementation of FFTW3. *Proc. IEEE* **2005**, *93*, 216–231. [CrossRef]
36. Kpalma, K.; Ronsin, J. An Overview of Advances of Pattern Recognition Systems in Computer Vision. In *Vision Systems*; Obinata, G., Dutta, A., Eds.; IntechOpen: Rijeka, Croatia, 2007; Chapter 10. [CrossRef]
37. Chwastek, T.; Mikrut, S. The problem of automatic measurement of fiducial mark on air images (in polish). *Arch. Photogramm. Cartogr. Remote Sens.* **2006**, *16*, 125–133.
38. Brunelli, R.; Messelodi, S. Robust estimation of correlation with applications to computer vision. *Pattern Recognit.* **1995**, *28*, 833–841. [CrossRef]