applied sciences

MDPI

Article

Action Classification for Partially Occluded Silhouettes by Means of Shape and Action Descriptors

Katarzyna Gościewska and Dariusz Frejlichowski *

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

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

Abstract: This paper presents an action recognition approach based on shape and action descriptors that is aimed at the classification of physical exercises under partial occlusion. Regular physical activity in adults can be seen as a form of non-communicable diseases prevention, and may be aided by digital solutions that encourages individuals to increase their activity level. The application scenario includes workouts in front of the camera, where either the lower or upper part of the camera's field of view is occluded. The proposed approach uses various features extracted from sequences of binary silhouettes, namely centroid trajectory, shape descriptors based on the Minimum Bounding Rectangle, action representation based on the Fourier transform and leave-one-out cross-validation for classification. Several experiments combining various parameters and shape features are performed. Despite the presence of occlusion, it was possible to obtain about 90% accuracy for several action classes, with the use of elongation values observed over time and centroid trajectory.

Keywords: shape analysis; foreground occlusion; action recognition; exercise classification; silhouette sequences

check for updates

Citation: Gościewska, K.; Frejlichowski, D. Action Classification for Partially Occluded Silhouettes by Means of Shape and Action Descriptors. *Appl. Sci.* **2021**, *11*, 8633. https://doi.org/10.3390/app11188633

Academic Editor: Francesco Bianconi

Received: 30 July 2021
Accepted: 14 September 2021
Published: 16 September 2021

1. Introduction

Human action recognition is a very popular topic in a field of computer vision. There is a large variety of applications associated with action recognition and classification based on Visual Content Analysis, some examples include surveillance systems [1,2], video retrieval and annotation [3], human–computer interaction [4,5], and quality-of-live improvement systems [6,7]. The variety of applications implies different approaches for the selection and description of action features due to varied action types. In [8], actions are defined as ''simple motion patterns usually executed by a single person and typically lasting for a very short duration (order of tens of seconds)." In turn, the author of [9] indicates diverse characteristics of an action, ranging from very primitive to cyclic body movements. As it is then stated in [10], actions are primitive movements lasting up to several minutes—an action is less complicated than an activity and more complex than a single motion (a gesture). Taking the above into account, an action is a collection of simple movements that are organized in a short period of time and can have periodic or non-periodic characteristics. Examples of actions include walking, hand waving and bending, and it is easy to notice that their basic movements are common to some physical exercises. In this paper, a focus is put on the recognition of actions that are classified based on exercise types. However, a problematic aspect is introduced—occlusion—which is a very challenging problem in computer vision [11,12].

According to the recommendations of the World Health Organization (WHO) [13], a healthy adult should do at least 150 min of moderate-intensity (or 75 min of vigorous-intensity) physical activity every week to keep general health and lower the risk of non-communicable diseases (e.g., cardiovascular disease, hypertension, diabetes, mental health conditions). A physical activity refers to a variety of exercises, such as jogging, weight lifting, bike riding, swimming and many more. However, the latest situation in the world related to the COVID-19 pandemic has been forcing people to change their daily routines and lockdowns, in particular, have been limiting physical activity [14,15]. There are some studies from all over the world researching the adverse consequences of the pandemic on different aspects of life and well-being that indicate the importance of physical activities for health improvement [16–19]. The WHO proposed some updated general recommendations on physical activity during pandemic, emphasizing several activity types that can be done at home, such as online exercise classes, dancing, playing active video games, jumping rope as well as muscle strength and balance training [20]. A narrative review in [15] summarises physical activity guidelines from 29 papers and gives some final joint recommendations. Generally, exercise types should include aerobic activity (e.g., running, walking), muscle strength activity (e.g., squats, various jumps), flexibility-stretching and relaxation (e.g., stretching arms, bending, yoga), and balance training. A brief report presented in [14] provides the results of a survey on the preferred home exercises for digital programs. Nearly 70% out of 15,261 participants from different countries were willing to work out at home, and were mostly interested in flexibility, resistance (strength training) and endurance (e.g., aerobic training) exercises.

In our works, a specific application scenario is assumed in which a person performs exercises in front of a computer camera, based on the displayed examples. Camera captures exercises and the system analyses them during recognition process. This can be related to active video games and exercise classes online. In the examined scenario, a focus is put on incomplete action data, where a part of a silhouette is occluded as a result of the improper camera positioning or the person being at the wrong distance from the camera. Then, usually the upper or lower part of the video frame is occluded and a part of foreground silhouette is missing (above the shoulders or below the knees). Due to the fact that each pose can be affected in different manner, an action descriptor should relay on the pose variability (changes between frames) rather than an exact shape contour itself. Effective recognition needs robust shape features which can be calculated despite the lack of a part of a silhouette. This paper addresses the problem by applying shape generalization and a combination of simple shape features that are analysed over time. Exercise types are recognised by means of binary silhouette sequences extracted from the Weizmann database [21] with added occlusions and two-stage classification. Firstly, coarse classification based on centroid trajectory divides actions into two subgroups—actions performed in place and actions with changing location of a silhouette. Since all exercises can be performed with repetitions, periodicity is not taken into account. Then, each silhouette is represented using selected shape descriptor and all silhouette descriptors of a sequence compose an action representation. Action representations are transformed using Discrete Fourier Transform (DFT) and classified using leave-one-sequence-out procedure. The proposed action recognition algorithm can be applied in a physical activity training system offering digital programs for home exercises. Types of exercises can be selected according to personal preferences and age group. For instance, in older adults, recommended exercises include mobility, balance and flexibility activities that lead to fall prevention [15]. Some of these exercises can be performed while seating or holding a chair, which can lead to occlusions. Given that the proposed approach can employ various shape features, its parameters can be adapted to specific applications.

Using a taxonomy presented in [12], the proposed action representation belongs to the category of holistic representations that are based on global features of a human body shape and movements. Body parts are not localised and discriminative information is extracted from regions of interest. Section 2 covers some examples of holistic approaches and other related works, as well as describes several action recognition challenges related to object occlusion in video sequences. 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 on Action Recognition and the Problem of Object Occlusion

Methods for recognizing human activities are widely applied in various systems and approaches in order to recognize and classify gestures, actions and behaviours. A multitude of applications and activity types results in a variety of features that can be used in the recognition process. However, an exemplary human activity recognition system consists of several common modules that perform motion segmentation, object classification, human tracking, action recognition and semantic description [2]. If a holistic representation is considered, action recognition is then based on the detected objects of interest (foreground binary masks, usually the shapes of human silhouettes) with known locations (trajectory or other motion information). Popular solutions include space-time techniques, where all silhouettes from a single sequence are accumulated and then features are extracted, or shape-based methods where features are extracted from each shape and combined into action representation afterwards.

2.1. Selected Related Works

One of the first works on spatio-temporal action recognition is reported in [22]. The authors propose Motion History Image (MHI) and Motion Energy Image (MEI) templates that represent how an object is moving and where the motion is present, respectively. In [23], the variants of MEI and MHI are combined to propose a new representation called temporal key poses. Instead of using one representation for a whole video sequence, a temporal template is assigned to each video frame. Then, k-nearest neighbour (KNN) and majority voting are applied for action classification. MHI is also used in [24] to create binary patterns. A texture operator, Local Binary Pattern (LBP), extracts the direction of motion from MHI and then the Histogram of Oriented Gradients (HOG) is used to represent features from LBP. Classification is performed with a Support Vector Machine (SVM).

Another solution that uses aggregated silhouettes from the entire sequence is described in [25]. Each sequence is represented using Average Silhouette Energy Images, where higher intensity pixels refer to static motion and low intensity values represent changing movements. Based on these, Edge Distribution of Gradients, Directional Pixels and R-transform produce feature vectors that are combined in a final action representation. The authors of [26] investigate several contour and shape based descriptors indicating superiority of silhouette-based approaches. Again, action features are extracted from accumulated silhouette images, called ASI. It is indicated that almost perfect accuracy can be obtained using HOG and KNN with Euclidean distance.

In [27], binary silhouettes are represented separately. Shape contours are converted into time series and then encoded into short symbolic representation called SAX-Shape. A set of SAX vectors represent an action, and classification is performed using a random forest algorithm. A popular concept for data reduction is the selection of key poses [28,29]. The authors of [29] use silhouette contours and divide them into equal radial bins using centroid location. Then, a summary value of each bin is calculated based on the Euclidean distances from centroid to every contour point. The proposed feature is used in a learning algorithm to build a bag of key poses, and sequences of key poses are classified using Dynamic Time Warping and leave-one-out procedure. The idea of radial scheme is applied in [30] as well; however, the entire shape area is used (contour with its interior). Firstly, human silhouette is divided into smaller regions in a radial fashion. Then, each radial bin is represented using region-based shape descriptors—geometrical and Hu moments. Obtained feature vectors are fed into multi-class SVM that indicates action classes. Radial coordinates and centroid location are also related to polar transform, which is proposed for shape description in [31]. Three polar coordinate systems are employed to represent the whole human body, and upper and lower parts of the body. For each of these systems, a histogram is generated based on radial bins. The normalized histograms are concatenated and represent a human posture. Classification is performed on a predefined number of frames.

2.2. Action Recognition under Occlusion

The approaches for action recognition mentioned in Section 2.1 yield high classification accuracy on popular benchmark datasets, however, have not been demonstrated on partially occluded action sequences so far. Occlusion is an important and challenging problem for vision-based systems, and foreground object occlusion can be perceived as a loss of data. In a single-camera setup, self-occlusion or inter-object occlusion may be present [11]. Moreover, the object of interest can remain partially occluded (temporarily or not) by the edge of the video frame—in other words, a part of the silhouette is outside the camera's field of view. Occlusion can coexist with other problems, such as changes in scene illumination, cluttered background, moving shadows or similar foreground to background, all of which can contribute to artefacts in images resulting from the background subtraction process. Artefacts in binary images have a form of extra pixels or a lack of some pixels in a silhouette area, thus it can be referred to as false data or data loss, respectively.

Occlusions are very problematic in real environments and busy scenes, making it difficult to detect and track objects. The selection of experimental database depends mainly on the recognition problem, application and required action types [11,32]. Literature sources indicate various ways to evaluate the action recognition method in the presence of occlusion, some of which add the occlusion to existing videos or foreground masks (e.g., [33–35]). In [33], a method for human action recognition based on sequences of depth maps is proposed. Each depth map consists of a human silhouette on a black background, where the area of a silhouette is coloured according to depth information. In order to avoid joint tracking, temporal dynamics are modelled using expandable graphical model (action graph) and the postures are represented using a bag of 3D points. A test dataset was collected by the authors, and include actions such as high arm wave, hand clap, bend, side kick, jogging or tennis serve, grouped into three subsets. The proposed solution is tested in each subset under simulated occlusion, where a depth map is divided into quadrants and one or two various quadrants of a silhouette are occluded. Compared to the experiments on unoccluded data, the recognition accuracy did not decrease significantly, except the case where the upper body was under heavy occlusion.

Another action recognition technique for corrupted silhouettes is proposed in [34]. It extracts normalized regions of interest containing as little background as possible (called a silhouette block) and represents each region as a binary vector—an image sequence of frames is transformed into vector sequence. Each silhouette block is partitioned into 12 subparts and partial template matching is applied. The results are integrated using voting or the sum of distances. During the experiments, two types of regions are superimposed on image sequences to simulate occlusion. These include horizontal and vertical stripes of different width and height, which increased the test dataset. The proposed approach obtained almost perfect accuracy for the original Weizmann database and is quite effective for horizontal occlusions.

In [35], several local descriptors are tested in the presence of static occlusion, namely Trajectory, Histogram of Oriented Gradients, Histogram of Orientation Flow and Motion Boundary Histogram. These methods are tested in combination with a standard bag-of-features representation and classified using SVM. The experiments were performed on the KTH benchmark dataset. Silhouette regions of interest were extracted and 10%, 25%, 50% or 75% of an area of each region was occluded. A combination of Trajectory and Motion Boundary Histogram (MBH) yielded average accuracy of 90% for partially occluded silhouettes.

3. Proposed Approach for Extraction and Classification of Action Descriptors

In this paper, an approach for action recognition is evaluated in the presence of static occlusion. A procedure consisting of several processing steps is proposed. It combines centroid locations that capture motion information, simple shape descriptors that represent characteristics of silhouettes and the Fourier transform to extract features from action descriptors. A previous version of the approach was presented in [36], where it was tested using more than twenty shape features and three matching measures. The convex hulls of silhouettes were used as input data in a scenario involving the recognition of eight types of actions corresponding to home exercises. The aim of the experiments was to indicate the most accurate shape features for action classification without any occlusions. In turn, here the evaluation of the approach is carried out in the presence of occlusion for a specific scenario concerning the recognition of various exercises performed by a single person in front of a static camera. It is assumed that either the upper or lower part of the silhouette is occluded because of the camera being positioned incorrectly or a person being not far enough from the camera. Only foreground masks extracted from the video sequences are used for recognition, due to the fact that they carry information about an object's shape, pose and position in the video frame. A static occlusion is added to every image, which results in removing a part of a silhouette from a foreground binary mask. It is assumed that all images are occluded; however, in several cases, a silhouette could temporarily be out of occlusion. For example, if there is an upper occlusion and a person bends down, a silhouette is fully visible for several frames. The similar situation may occur for jumping in place when a person bends the legs. This can be seen as a dynamic occlusion. However, since static occlusion is added to every frame and whole sequences are analysed, the small amount of dynamic occlusion is of minor importance and simply results from the intrinsic characteristics of some action classes. The following subsections explain in more detail how silhouette sequences are processed and classified. Figure 1 gives an overview of the proposed approach.

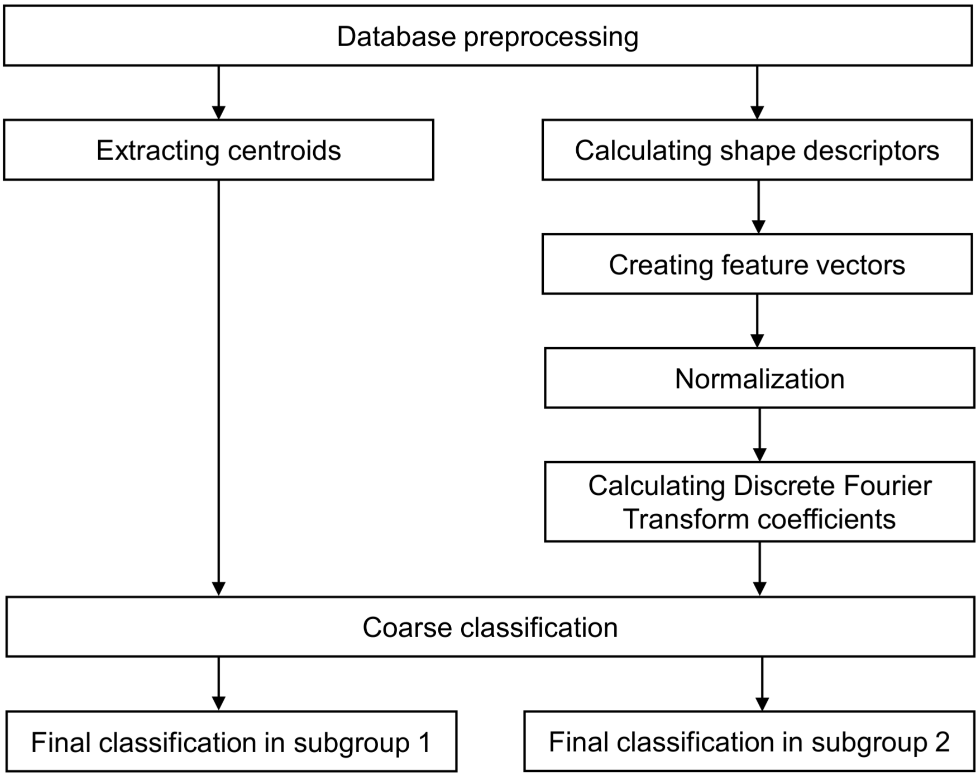

Database preprocessing
Extracting centroids
Calculating shape descriptors
Creating feature vectors
Normalization
Calculating Discrete Fourier Transform coefficients
Coarse classification
Final classification in subgroup 1
Final classification in subgroup 2

Figure 1. A block diagram representing subsequent steps of the proposed approach.

3.1. Database and Preprocessing

The experiments on the proposed approach are performed using the Weizmann database [21] (the results are presented in Section 3). Here, the reasons for data selection are explained and the preprocessing steps are described. The Weizmann dataset has several advantages. Actions are captured on a static background and human silhouettes are easily distinguishable, fully visible and of equal scale. There are ten action types in the database which correspond to recommended exercises [14,15,37] as follows:

- Aerobic exercise, e.g., running, walking, skipping, jumping jack (activities in which the body's large muscles move in a rhythmic manner);
- Balance training, e.g., galloping sideways, jumping in place, jumping forward on two legs (activities increasing lower body strength);
- Flexibility exercise, e.g., one or two-hand waving, bending (activities preserving or extending motion range around joints).

The database consists of 90 video sequences (144 × 180 px, 50 fps, 1–3 s) and corresponding foreground masks extracted for each video frame, obtained using background subtraction. Each mask is a binary image containing single human silhouette (white pixels) on a black background (see Figure 2 for examples). Some silhouettes are incomplete or have additional pixels (Figure 3)—artefacts resulting from the foreground extraction process. Depending on the shape descriptor used, artefacts may have insignificant influence on the classification results. Binary images with background only or in which a silhouette is too close to the left or right edge of the video frame are removed. The direction of movement is standardized, so all actors move from left to right.

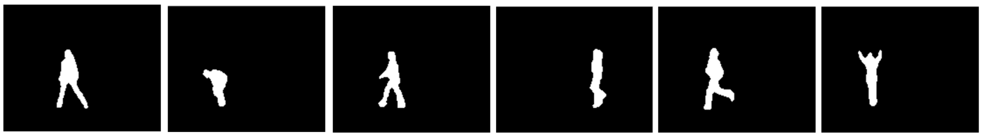


Figure 2. Exemplary foreground masks from the Weizmann database [21].

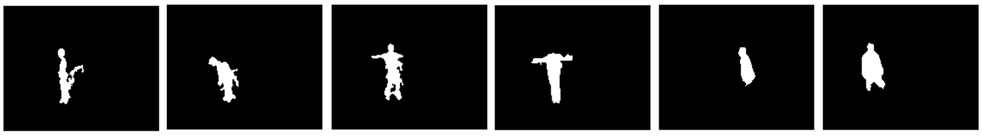


Figure 3. Examples depicting various artefacts resulting from the background subtraction.

3.2. Extracting Motion Information and Adding Occlusion

Preprocessed binary images are used to obtain the centroid trajectory. The centroid coordinates are calculated as the average coordinates of all pixels belonging to the silhouette's area. The trajectory length (relative to the bottom edge of the video frame) is used as a condition for dividing the database into two subgroups. This step is referred to as coarse classification. Actions are often distinguished into periodic and non-periodic ones, but in case of physical activity all exercises can be repeated periodically. Therefore, here the database is divided automatically into actions performed in place (a trajectory is short, e.g., bending) and actions during which a person changes location across the video frame (a trajectory is long, e.g., walking). For a single video sequence, centroid positions are calculated for each silhouette and accumulated in a separate image whose size is equal to the size of a video frame (144 × 180 px). The trajectory length is measured relative to the bottom edge of the video frame and if it is shorter than 20 pixels an action is classified as performed in place. Longer trajectories refer to actions with changing location of a silhouette. Coarse classification allows for the use of different features and parameters for each subgroup.

The assumed application scenario requires the input data to be occluded in the specific manner; therefore, the foreground masks from the Weizmann database were modified in two different ways to occlude the upper and lower parts of the camera's field of view. This type of occlusion simulates a situation in which a person is too close to the camera, and either the person's legs or head with neck and shoulders do not appear in the video frame. Foreground silhouettes on the original masks were located in slightly different places vertically. Therefore, in order to determine the cut-off point, the centroid coordinates of all silhouettes in the given sequence were averaged. The average value was then decreased or increased to experimentally indicate the final cut-off point, which translates to the size of the occlusion. For example, if the averaged centroid row coordinate equals 50 (the image matrix origin is in the upper left corner), then it was decreased by 20 for upper occlusion and increased by 20 for the lower occlusion. The cut-off point differs only between sequences due to changes in centroid location, but it is constant for all frames within a given sequence (a simulation of a static camera). Figure 4 depicts several examples of occluded silhouettes.

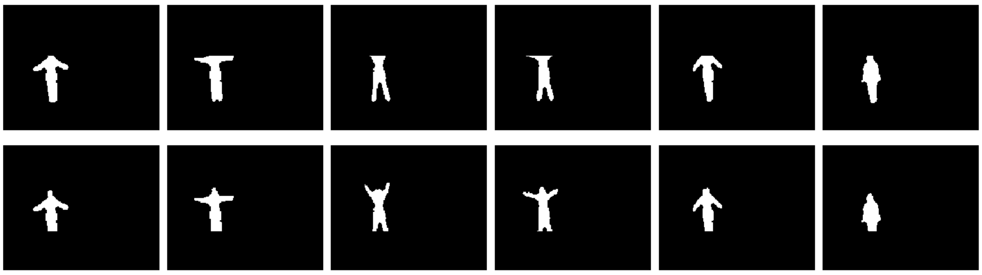


Figure 4. Foreground masks for a jumping-jack action with added occlusions.

3.3. Shape Description

The shape description algorithm enables obtaining a numerical representation of a shape. It extracts the most distinctive features and reduces two-dimensional image matrix into a vector with numbers, or a single value in case of simple shape descriptors. Such simple descriptors are basic shape measurements or shape factors that capture general characteristics and usually are not used alone [38]. However, when calculated for more shapes and observed over time, they are more valuable. An appropriate selection of the representation helps to limit the influence of outliers, and shape generalization can be used additionally. Here, the Minimum Bounding Rectangle (MBR) is employed.

An MBR defines a smallest rectangular region that contains all points of a shape [38,39]. The algorithm creates a blob-like object that covers a foreground silhouette's area and is described by the coordinates of four corner points. The idea of using rectangular blobs is also known from the object detection and tracking approaches. Based on the corner coordinates, several characteristics can be calculated and combined into shape ratios. Basic MBR measurements include its length and width. The length of the rectangle can be calculated in two different ways. Firstly, as a distance between two fixed corners where width is a horizontal measurement and length is vertical (in relation to the coordinate axes). This approach is used to calculate other MBR-based descriptors. However, in the second method, the length is always the longer side of the rectangle, and width is the shorter one, no matter which corner points it concerns. These are referred to as 'shorterMBR' and 'longerMBR'. Width and height are further used to estimate perimeter, area, eccentricity, elongation and rectangularity. Eccentricity is a ratio of width to length of the MBR, while elongation is a value of eccentricity subtracted from 1. Rectangularity shows the similarity of an original silhouette to its MBR and is obtained as the ratio of the area of a shape to the area of its MBR. Ultimately, nine different MBR measurements and ratios are considered as shape descriptors for the evaluated approach.

In the shape description step, each foreground mask is represented using selected shape descriptor. As a result, an image is reduced to a single value. The descriptors of all frames from a given sequence are put into a vector and subjected to the normalization to a range of 0 to 1. Normalized vector can be plotted as a line graph to observe how shape descriptors change over time. The number of shape descriptors of a single sequence equals the number of its video frames; therefore, these vectors are still of different length.

3.4. Action Representation

This step aims to prepare action representations of equal size, based on the normalized vectors with shape descriptors. To achieve this, the one-dimensional Discrete Fourier Transform is applied. The number of the resultant Fourier coefficients is predefined and various options are tested to indicate the smallest and most accurate one. If the number of coefficients is smaller than the size of a description vector, the latter one is truncated. Otherwise, zero-padding is applied; therefore, the descriptor vectors are appended with zeros. Adding zeros in the time domain is equal to the interpolation in the frequency domain. Final action representations contain absolute values of the Fourier coefficients.

3.5. Action Classification

The proposed approach employs two-stage classification. A coarse classification that creates two subgroups—actions performed in place and actions with changing location of a silhouette—can be performed at the earlier step. This is due to that different shape descriptors and action representations can be applied. The actual classification is performed in each subgroup separately based on a standard leave-one-sequence-out cross-validation. It means that each action representation is matched with the rest of the representations from the database and a matching measure is calculated. This procedure is repeated for all instances and the results are accumulated in terms of the percentage of correct classifications. The closest match indicates probable class of an action under investigation. If a correlation is used for matching (C1 correlation based on L1-norm [40]), then the most similar indication is taken. In turn, if a distance is calculated (Euclidean distance [41]), a nearest neighbour (less dissimilar one) indicates the recognized class.

3.6. List of Processing Steps

Sections 3.1–3.5 provided a detailed description of the proposed approach, including how to prepare the input data. To summarize the most important elements, here a list of consecutive processing steps is given, as follows:

1. The database is preprocessed as explained in Section 3.1, and occlusion is added based on Section 3.2.
2. Motion and shape features are extracted from all sequences. A sequence is composed of binary foreground masks $BM_i = \{bm_1, bm_2, \ldots, bm_n\}$ and represented by a vector with normalized shape descriptors, $SD_i = \{sd_1, sd_2, \ldots, sd_n\}$, where n is the number of frames. Shape descriptors are based on the Minimum Bounding Rectangle measurements which are explained in Section 3.3. To collect motion information, centroid locations are stored as trajectories (Section 3.2). Centroid coordinates are calculated as an average of coordinates of all points included in the shape area.
3. Each SD vector is transformed into action representation AR using the Discrete Fourier Transform (Section 3.4). The one-dimensional DFT of an exemplary vector $u(t)$ of length t ($0 < t < T$; T is a period of t) is as follows [42]:

$$a_k = \frac{1}{N} \sum_{t=0}^{N-1} u(t) \exp^{\left(\frac{-j2\pi kt}{N}\right)}, k = 0, 1, \ldots, N-1 \quad (1)$$

   Then, a selected number of absolute coefficients is used for classification.
4. Coarse classification is performed based on trajectory length. The database is divided into two subgroups.
5. Final classification (Section 3.5) is performed in each subgroup separately using the leave-one-sequence-out procedure and one of the following matching measures—Euclidean distance d_E [41] or C1 correlation c_1 [40]. The respective formulas for two compared vectors $x = x_1, x_2, \ldots, x_n$ and $y = y_1, y_2, \ldots, y_n$ are:

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \quad (2)$$

$$c_1(x, y) = 1 - \frac{\sum\limits_i |x_i - y_i|}{\sum\limits_i (|x_i| - |y_i|)}. \quad (3)$$

4. Experimental Conditions and Results

The experiments were carried out with the use of the Weizmann database [21], which contains ten action classes with two action types: the first type includes actions performed in place, such as bending (referred to as 'bend'), jumping jack ('jack'), jumping in place ('pjump'), waving with one hand ('wave1') and waving with two hands ('wave2'), and the second type refers to actions with changing location of a silhouette, namely walking ('walk'), running ('run'), galloping sideways ('side'), jumping forward on two legs ('jump') and skipping ('skip'). Action sequences are divided into subgroups during coarse classification and the division is made based on the trajectory length; therefore, the subgroups are referred to as 'short trajectory' and 'long trajectory', respectively. All foreground images from 90 sequences were preprocessed in accordance with Section 3.1. Based on Section 3.2, two versions of the database were prepared by occluding the upper or lower part of the silhouette, and are referred to as the 'upper-occlusion' and 'lower-occlusion'. Each database is tested separately.

In a single experiment, there are several tests investigating classification accuracy of the combinations of the shape descriptor, matching measure and the size of action representation. A sequence of binary silhouettes (less than 150 images per action) is taken as an input data for a single action. Firstly, all silhouettes are represented using selected shape descriptor and the results are combined into a vector which is then normalized. Vectors from all silhouette sequences are transformed into action representations that are subjected to the classification process. This includes coarse classification into two subgroups and final classification into action classes using leave-one-sequence-out procedure. A test representation is matched with the rest of the representations and the most similar one indicates the probable action class. Two matching measures are compared: Euclidean distance and C1 correlation. The experiments employed nine shape descriptors and various action representations based on from 2 to 256 DFT coefficients. The best result relates to the experiment with the highest accuracy (correct classification rate) and the smallest representation size. Then, a combination of a shape descriptor and matching measure that is used during this experiment is considered as the most effective version of the approach. The top results for each shape descriptor, obtained for occluded data, are presented in Tables 1 and 2. In turn, Table 3 contains a comparison of the selected results achieved using both occluded databases and the original Weizmann database.

Table 1 contains the results of the experiments performed on the 'lower-occlusion' database. The highest classification accuracy for actions performed in place is 91.11%. The second best is 88.89%, and the third one is 84.44%. The most accurate experiment employed elongation for shape description. The accuracy of 91.11% is then achieved either for Euclidean distance or C1 correlation, and action representations are composed of 57 and 55 coefficients, respectively. For the second group of actions, the maximum recognition rate is 73.33% and was obtained for two shape descriptors—rectangularity (84 coefficients) and shorterMBR (71 coefficients). In both experiments, the C1 correlation is applied. The difference between the highest results in both subgroups is large and equals nearly 20%. In case of actions with changing location of a silhouette, this should not be surprising since the lower part of the silhouette is occluded and feet cannot be localised continuously.

The results for the experiments performed with the use of the 'upper-occlusion' database are not so strongly diversified between the subgroups (Table 2). The highest classification rate for actions performed in place equals 88.89% and is attributed to the experiment using elongation, 51 Fourier coefficients and C1 correlation. For the other subgroup, an accuracy of 84.44% is achieved in three experiments using area, shorterMBR and rectangularity. However, if rectangularity is employed, the representation size is smaller (46 coefficients using C1 correlation).

Table 1. Results for the experiments using the 'lower-occlusion' database. Correct classification rates are given for coarsely classified actions, nine shape descriptors based on the Minimum Bounding Rectangle and two matching measures (EU—distance, C1—correlation). The number of the DFT coefficients is given in brackets.

Shape Descriptor	**Actions Performed in Place**		**Actions with Changing Location of a Silhouette**	
	EU	**C1**	**EU**	**C1**
Rectangularity	68.89% (68)	71.11% (110)	68.89% (79)	73.33% (84)
Elongation	91.11% (57)	91.11% (55)	68.89% (79)	64.44% (39)
Eccentricity	71.11% (88)	75.56% (108)	64.44% (37)	60.00% (38)
Width	82.22% (54)	84.44% (54)	55.56% (17)	60.00% (57)
Length	62.22% (41)	62.22% (18)	64.44% (36)	60.00% (65)
Area	68.89% (39)	71.11% (106)	62.22% (66)	64.44% (70)
LongerMBR	68.89% (41)	71.11% (41)	66.67% (68)	66.67% (89)
ShorterMBR	88.89% (55)	84.44% (56)	71.11% (30)	73.33% (71)
Perimeter	77.78% (52)	80.00% (52)	66.67% (40)	62.22% (39)

Table 2. Results for the experiments using the 'upper-occlusion' database. Correct classification rates are given for coarsely classified actions, nine shape descriptors based on the Minimum Bounding Rectangle and two matching measures (EU—distance, C1—correlation). The number of the DFT coefficients is given in brackets.

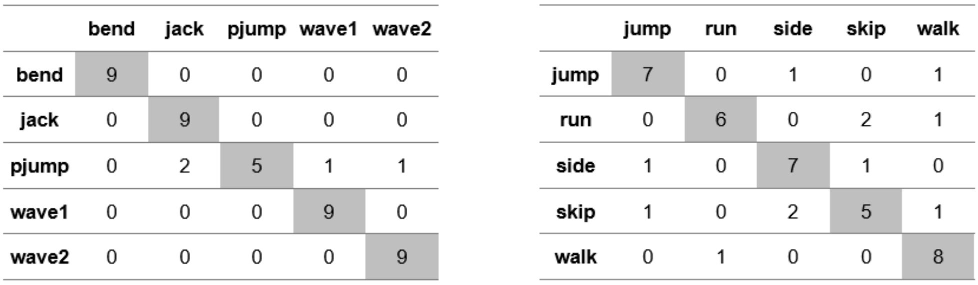
Shape Descriptor	**Actions Performed in Place**		**Actions with Changing Location of a Silhouette**	
	EU	**C1**	**EU**	**C1**
Rectangularity	77.78% (55)	75.56% (50)	82.22% (28)	84.44% (46)
Elongation	82.22% (47)	88.89% (51)	82.22% (27)	80.00% (28)
Eccentricity	68.89% (32)	71.11% (34)	73.33% (41)	77.78% (168)
Width	73.33% (53)	71.11% (81)	60.00% (39)	60.00% (18)
Length	55.56% (38)	64.44% (42)	75.56% (40)	75.56% (68)
Area	68.89% (38)	75.56% (83)	84.44% (70)	77.78% (38)
LongerMBR	75.56% (53)	80.00% (58)	60.00% (27)	60.00% (40)
ShorterMBR	80.00% (124)	80.00% (65)	82.22% (18)	84.44% (69)
Perimeter	68.89% (99)	68.89% (56)	75.56% (93)	75.56% (75)

Table 3. Summary of the results obtained using databases with or without occlusions.

Shape Descriptor	Actions Performed in Place		Actions with Changing Location of a Silhouette	
	EU	C1	EU	C1
'Lower-occlusion' database	Elongation 91.11% (57)	Elongation 91.11% (55)	ShorterMBR 71.11% (30)	ShorterMBR 73.33% (71)
'Upper-occlusion' database	Elongation 82.22% (47)	Elongation 88.89% (51)	Area 84.44% (70)	Rectangularity 84.44% (46)
Database without occlusion	ShorterMBR 86.67% (56)	Perimeter 91.11% (51)	Area 86.67% (31)	Area 84.44% (33)

The proposed approach is also tested using the Weizmann database without any occlusions. The comparison of the experimental results yielding the highest classification accuracy is given in Table 3. Some conclusions can be drawn. Firstly, in several cases, the use of C1 correlation gives better results or the results are equal to the experiments with Euclidean distance. Secondly, in the subgroup of actions with changing location of a silhouette, the lower occlusion has stronger influence on the results. It can be concluded that, for action classification purposes, the information about feet positions is more important than the presence of head and shoulders. It is also confirmed by the results obtained for the 'upper-occlusion' database, which were very similar to those achieved on the unoccluded database. Different conclusions can be drawn from the results obtained for actions performed in place. Since actors do not move across the camera's field of view, feet positions do not influence the results. Furthermore, the same accuracy is achieved for the 'lower-occlusion' database and the database without any occlusions. It is also important to notice that different shape descriptors are used; however, the size of action representation is similar. In some cases, the number of DFT coefficients is smaller than the number of images in an input sequence. Therefore, the proposed approach offers data reduction, both by representing each image by a single value and each action by a selected number of the absolute spectral domain coefficients.

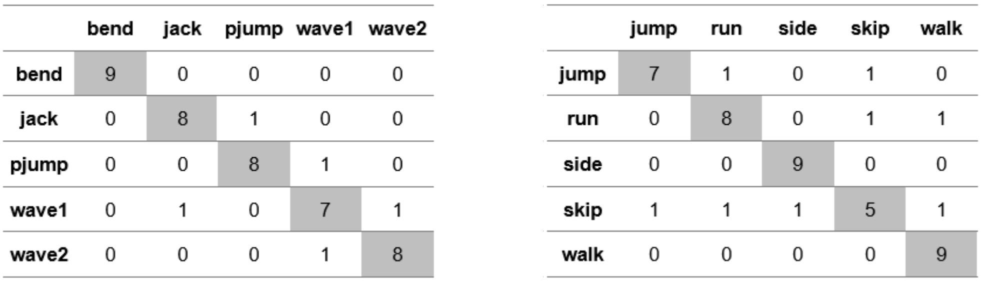
Figures 5–7 present confusion matrices for selected experiments which yielded the highest accuracy results. In each subgroup, there are 45 action sequences—five action classes, nine sequences in each class. Each figure contains two confusion matrices: the left one refers to 'actions performed in place' and the other one to 'actions with changing location of a silhouette'. Values in matrices correspond to the number of classifications, where correct classifications are on a grey background, and misclassifications are on a white background. Among the occluded actions, 'skip' is most often confused with other classes. However, when there is no occlusion, this class is recognized correctly, but other action types are confused with 'skip'. In the experiment with the 'lower-occlusion' database, besides 'skip', jump in place ('pjump') was less accurate classified, with 4 false classifications out of 9 in total. Among actions without occlusion, jumping forward ('jump') is the least distinguishable class. Appendix A contains additional confusion matrices for the experimental results which are given in Tables 1 and 2.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	9	0	0	0
pjump	0	2	5	1	1
wave1	0	0	0	9	0
wave2	0	0	0	0	9

	jump	run	side	skip	walk
jump	7	0	1	0	1
run	0	6	0	2	1
side	1	0	7	1	0
skip	1	0	2	5	1
walk	0	1	0	0	8

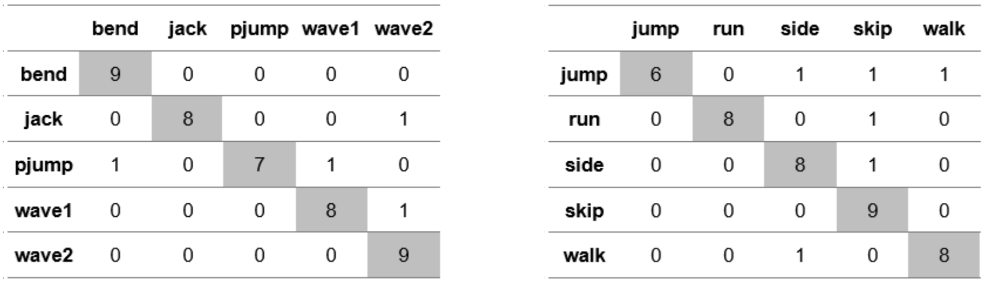
Figure 5. Confusion matrices for the best experiment using the 'lower-occlusion' database. (**Left table**): 'actions performed in place' subgroup, Elongation, C1 correlation, 55 DFT coefficients, accuracy 91.11%. (**Right table**): 'actions with changing location of a silhouette' subgroup, ShorterMBR, C1 correlation, 71 DFT coefficients, accuracy 73.33%.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	8	1	0	0
pjump	0	0	8	1	0
wave1	0	1	0	7	1
wave2	0	0	0	1	8

	jump	run	side	skip	walk
jump	7	1	0	1	0
run	0	8	0	1	1
side	0	0	9	0	0
skip	1	1	1	5	1
walk	0	0	0	0	9

Figure 6. Confusion matrices for the best experiment using the 'upper-occlusion' database. (**Left table**): 'actions performed in place' subgroup, Elongation, C1 correlation, 51 DFT coefficients, accuracy 88.89%. (**Right table**): 'actions with changing location of a silhouette' subgroup, Rectangularity, C1 correlation, 46 DFT coefficients, accuracy 84.44%.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	8	0	0	1
pjump	1	0	7	1	0
wave1	0	0	0	8	1
wave2	0	0	0	0	9

	jump	run	side	skip	walk
jump	6	0	1	1	1
run	0	8	0	1	0
side	0	0	8	1	0
skip	0	0	0	9	0
walk	0	0	1	0	8

Figure 7. Confusion matrices for the best experiment using the original database without occlusions. (**Left table**): 'actions performed in place' subgroup, Perimeter, C1 correlation, 51 DFT coefficients, accuracy 91.11%. (**Right table**): 'actions with changing location of a silhouette' subgroup, Area, Euclidean distance, 31 DFT coefficients, accuracy 86.67%.

5. Discussion and Conclusions

This paper presents an approach for action recognition in the scenario of exercise classification under partial occlusion. It uses a combination of shape descriptors, trajectory and spectral domain features to assign silhouette sequences to action classes. General shape features are tracked; therefore, small disturbances or artefacts in the foreground masks are of minor importance. The proposed approach has a form of a general procedure consisting of several processing steps, but some of them are using different methods. It enables this approach to be adapted to other applications or data. It also made it possible to test different combinations of shape descriptors, action descriptors and matching measures in the assumed scenario. In order to improve recognition accuracy, applying other shape features or classification procedures is considered.

The goal of the experiments was to indicate such combination that would give the highest accuracy for the recognition of occluded silhouettes. Results differ between subgroups of actions performed in place and actions with changing location of a silhouette. For the 'lower-occlusion' database, the highest accuracy is 91.11% (elongation, 55 coefficients, C1 correlation) and 73.33% (shorterMBR, 71 coefficients, C1 correlation), for the subgroups, respectively. In turn, in the 'upper-occlusion' database, the highest accuracy in the first subgroup is 88.89% (elongation, 51 coefficients, C1 correlation) and 84.44% in the second one (rectangularity, 46 coefficients, C1 correlation). This gives an average accuracy for the entire 'upper-occlusion' database equal to 86.65%, while, for the 'lower-occlusion' database, it is 82.22%. Compared to the averaged accuracy in data without occlusions, which is 88.89%, the decrease in classification rate is small. It can also be concluded that classification accuracy for actions performed in place is close to 90%, and higher than in the other subgroup, which makes the results promising, especially under the assumed scenario.

The proposed approach was tested using the datasets modified by the authors; therefore, there are no other results in the literature, presented on the same data. However, some general characteristics of the results can be compared to the conclusions on some other solutions handling occlusion. The authors of [35] investigated combinations of different descriptors as well. They experimented with partial occlusion (10% or 25% of the region of interest is occluded) and heavy occlusion (50% and 75% occluded), and compared the results with unoccluded data. The percentage results are similar to the results presented in this paper. For the experiments combining trajectory with MBH or HOG, the difference in accuracy between partially occluded and unoccluded data amounted to several percent. The authors of [33] proposed an action recognition approach that uses silhouettes with depth data, and achieves nearly perfect accuracy for three subgroups of activity classes. Occlusion is added by removing a quarter or a half of a silhouette. The highest accuracy decrease is observed when the top part of a silhouette is fully or partially occluded, especially in the subgroup where the following action classes are included: forward punch, high throw, hand clap, bend and more. In other cases, the differences were relatively small. In [34], several horizontal and vertical occlusions were added to the Weizmann database and tested with the use of the approach based on spatio-temporal blocks and partial template matching. The comparison of the results shows small differences in accuracy between the experiments performed on databases with horizontal occlusions and without any occlusion, while vertical occlusions cause larger accuracy decrease.

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. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

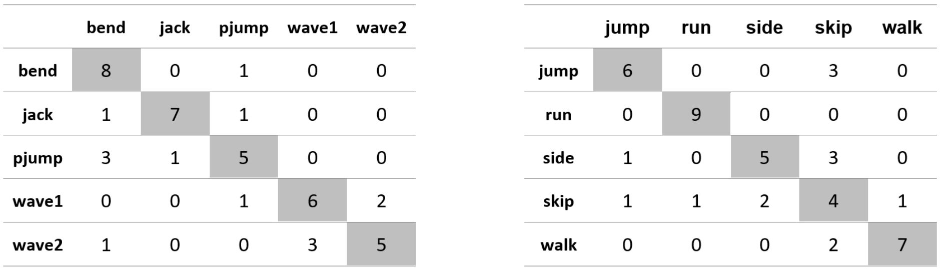
Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository (https://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html).

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

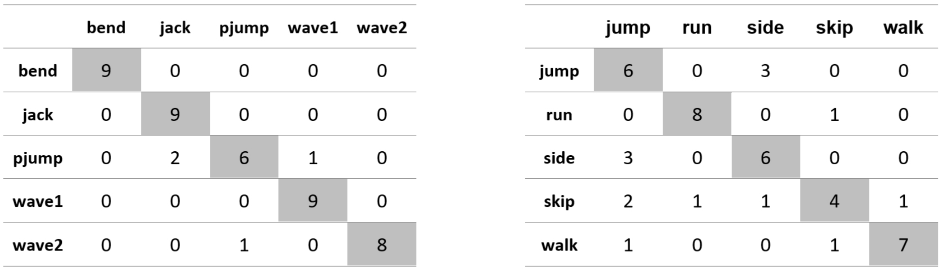
Appendix A. Figures Presenting Confusion Matrices for the Experiments the Results of Which Are Given in Tables 1 and 2



	bend	jack	pjump	wave1	wave2
bend	8	0	1	0	0
jack	1	7	1	0	0
pjump	3	1	5	0	0
wave1	0	0	1	6	2
wave2	1	0	0	3	5

	jump	run	side	skip	walk
jump	6	0	0	3	0
run	0	9	0	0	0
side	1	0	5	3	0
skip	1	1	2	4	1
walk	0	0	0	2	7

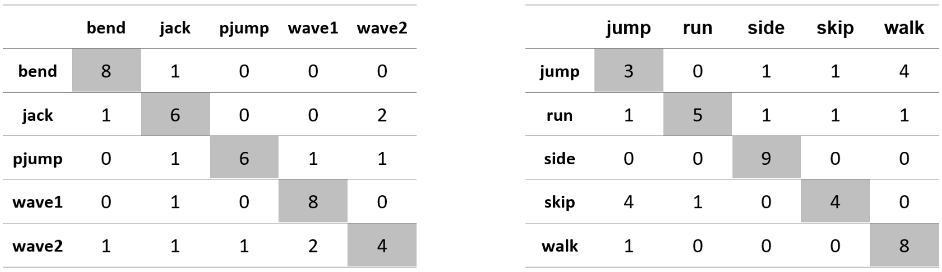
Figure A1. Confusion matrices for the experiment using the 'lower-occlusion' database, rectangularity and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	9	0	0	0
pjump	0	2	6	1	0
wave1	0	0	0	9	0
wave2	0	0	1	0	8

	jump	run	side	skip	walk
jump	6	0	3	0	0
run	0	8	0	1	0
side	3	0	6	0	0
skip	2	1	1	4	1
walk	1	0	0	1	7

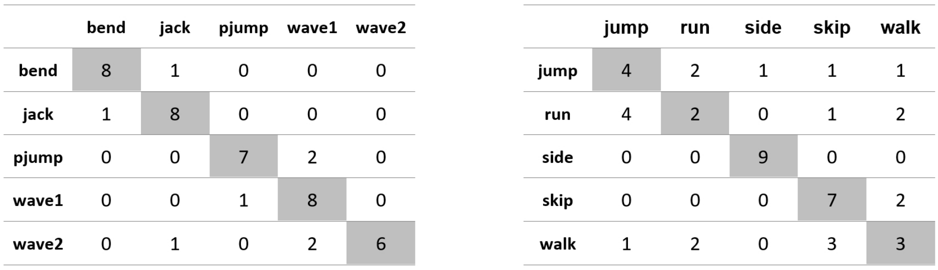
Figure A2. Confusion matrices for the experiment using the 'lower-occlusion' database, elongation and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	8	1	0	0	0
jack	1	6	0	0	2
pjump	0	1	6	1	1
wave1	0	1	0	8	0
wave2	1	1	1	2	4

	jump	run	side	skip	walk
jump	3	0	1	1	4
run	1	5	1	1	1
side	0	0	9	0	0
skip	4	1	0	4	0
walk	1	0	0	0	8

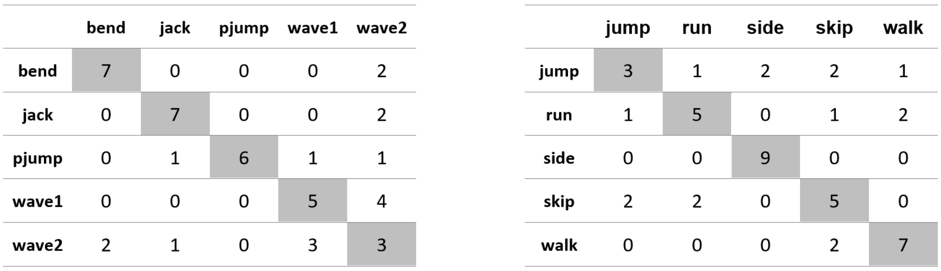
Figure A3. Confusion matrices for the experiment using the 'lower-occlusion' database, eccentricity and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	8	1	0	0	0
jack	1	8	0	0	0
pjump	0	0	7	2	0
wave1	0	0	1	8	0
wave2	0	1	0	2	6

	jump	run	side	skip	walk
jump	4	2	1	1	1
run	4	2	0	1	2
side	0	0	9	0	0
skip	0	0	0	7	2
walk	1	2	0	3	3

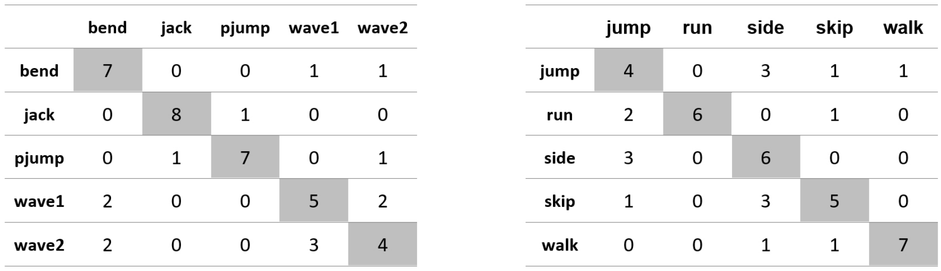
Figure A4. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR width and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	7	0	0	0	2
jack	0	7	0	0	2
pjump	0	1	6	1	1
wave1	0	0	0	5	4
wave2	2	1	0	3	3

	jump	run	side	skip	walk
jump	3	1	2	2	1
run	1	5	0	1	2
side	0	0	9	0	0
skip	2	2	0	5	0
walk	0	0	0	2	7

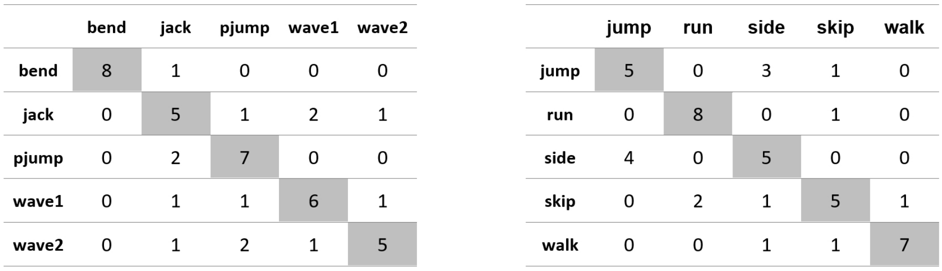
Figure A5. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR length and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	7	0	0	1	1
jack	0	8	1	0	0
pjump	0	1	7	0	1
wave1	2	0	0	5	2
wave2	2	0	0	3	4

	jump	run	side	skip	walk
jump	4	0	3	1	1
run	2	6	0	1	0
side	3	0	6	0	0
skip	1	0	3	5	0
walk	0	0	1	1	7

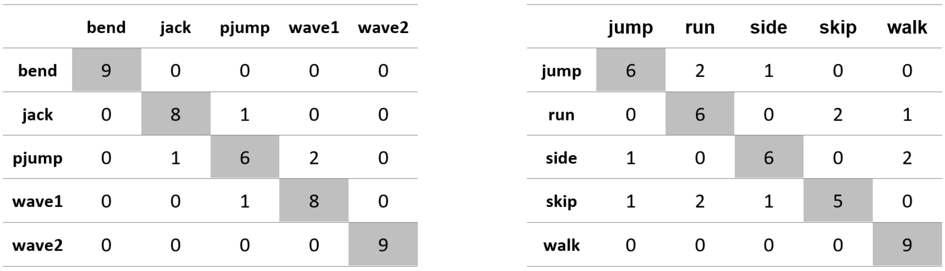
Figure A6. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR area and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	8	1	0	0	0
jack	0	5	1	2	1
pjump	0	2	7	0	0
wave1	0	1	1	6	1
wave2	0	1	2	1	5

	jump	run	side	skip	walk
jump	5	0	3	1	0
run	0	8	0	1	0
side	4	0	5	0	0
skip	0	2	1	5	1
walk	0	0	1	1	7

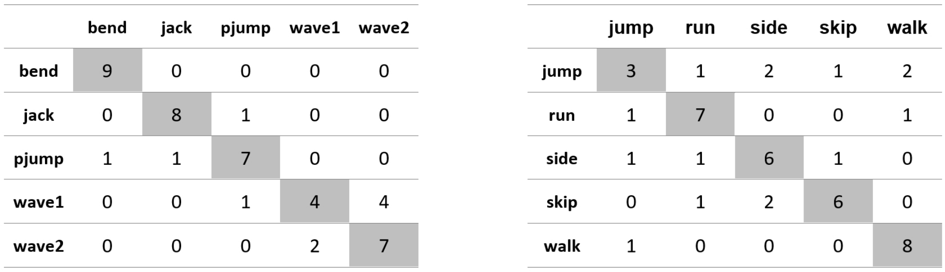
Figure A7. Confusion matrices for the experiment using the 'lower-occlusion' database, longerMBR and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	8	1	0	0
pjump	0	1	6	2	0
wave1	0	0	1	8	0
wave2	0	0	0	0	9

	jump	run	side	skip	walk
jump	6	2	1	0	0
run	0	6	0	2	1
side	1	0	6	0	2
skip	1	2	1	5	0
walk	0	0	0	0	9

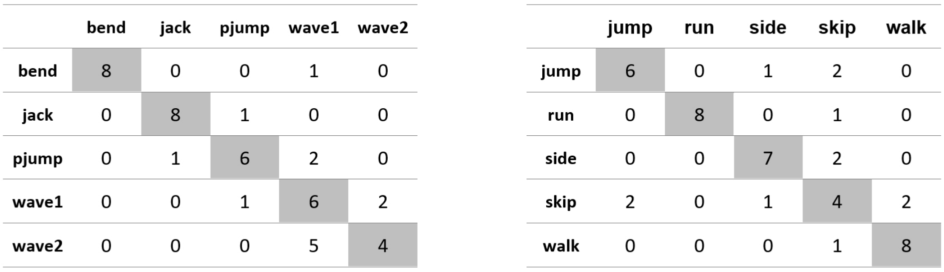
Figure A8. Confusion matrices for the experiment using the 'lower-occlusion' database, shorterMBR and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	8	1	0	0
pjump	1	1	7	0	0
wave1	0	0	1	4	4
wave2	0	0	0	2	7

	jump	run	side	skip	walk
jump	3	1	2	1	2
run	1	7	0	0	1
side	1	1	6	1	0
skip	0	1	2	6	0
walk	1	0	0	0	8

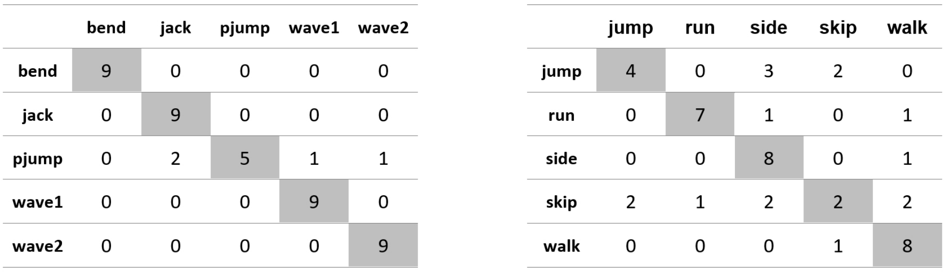
Figure A9. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR perimeter and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	8	0	0	1	0
jack	0	8	1	0	0
pjump	0	1	6	2	0
wave1	0	0	1	6	2
wave2	0	0	0	5	4

	jump	run	side	skip	walk
jump	6	0	1	2	0
run	0	8	0	1	0
side	0	0	7	2	0
skip	2	0	1	4	2
walk	0	0	0	1	8

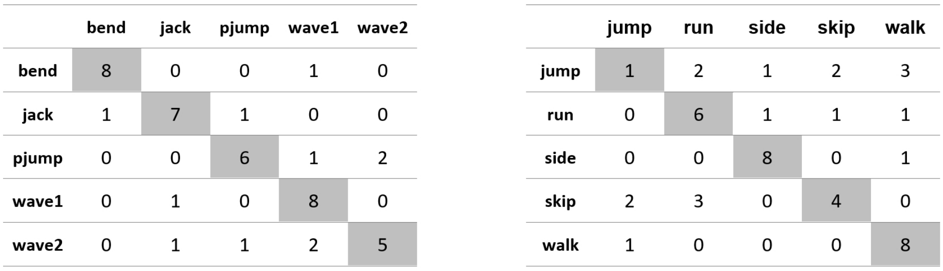
Figure A10. Confusion matrices for the experiment using the 'lower-occlusion' database, rectangularity and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	9	0	0	0
pjump	0	2	5	1	1
wave1	0	0	0	9	0
wave2	0	0	0	0	9

	jump	run	side	skip	walk
jump	4	0	3	2	0
run	0	7	1	0	1
side	0	0	8	0	1
skip	2	1	2	2	2
walk	0	0	0	1	8

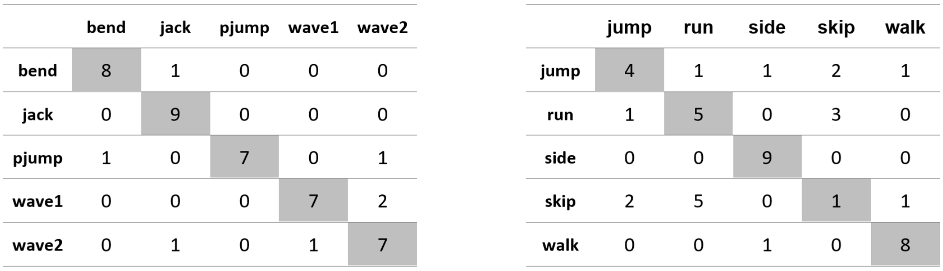
Figure A11. Confusion matrices for the experiment using the 'lower-occlusion' database, elongation and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	8	0	0	1	0
jack	1	7	1	0	0
pjump	0	0	6	1	2
wave1	0	1	0	8	0
wave2	0	1	1	2	5

	jump	run	side	skip	walk
jump	1	2	1	2	3
run	0	6	1	1	1
side	0	0	8	0	1
skip	2	3	0	4	0
walk	1	0	0	0	8

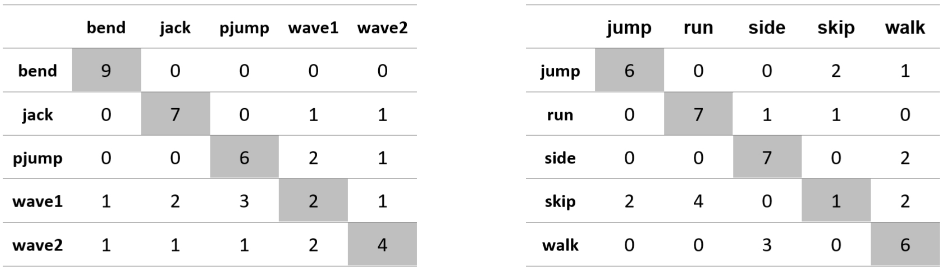
Figure A12. Confusion matrices for the experiment using the 'lower-occlusion' database, eccentricity and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	8	1	0	0	0
jack	0	9	0	0	0
pjump	1	0	7	0	1
wave1	0	0	0	7	2
wave2	0	1	0	1	7

	jump	run	side	skip	walk
jump	4	1	1	2	1
run	1	5	0	3	0
side	0	0	9	0	0
skip	2	5	0	1	1
walk	0	0	1	0	8

Figure A13. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR width and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	7	0	1	1
pjump	0	0	6	2	1
wave1	1	2	3	2	1
wave2	1	1	1	2	4

	jump	run	side	skip	walk
jump	6	0	0	2	1
run	0	7	1	1	0
side	0	0	7	0	2
skip	2	4	0	1	2
walk	0	0	3	0	6

Figure A14. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR length and C1 correlation.

	bend	jack	pjump	wave1	wave2
bend	8	0	0	0	1
jack	0	8	1	0	0
pjump	0	2	7	0	0
wave1	0	0	1	4	4
wave2	0	0	0	4	5

	jump	run	side	skip	walk
jump	5	0	1	3	0
run	0	8	0	1	0
side	2	0	5	2	0
skip	2	2	1	3	1
walk	0	0	0	1	8

Figure A15. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR area and C1 correlation.

	bend	jack	pjump	wave1	wave2
bend	7	0	0	1	1
jack	0	5	1	3	0
pjump	0	2	7	0	0
wave1	0	1	0	8	0
wave2	2	1	1	0	5

	jump	run	side	skip	walk
jump	5	0	2	1	1
run	0	8	0	1	0
side	3	0	6	0	0
skip	1	2	2	4	0
walk	0	0	2	0	7

Figure A16. Confusion matrices for the experiment using the 'lower-occlusion' database, longerMBR and C1 correlation.

	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	9	0	0	0
pjump	2	2	5	0	0
wave1	0	0	1	7	1
wave2	0	0	0	1	8

	jump	run	side	skip	walk
jump	7	0	1	0	1
run	0	6	0	2	1
side	1	0	7	1	0
skip	1	0	2	5	1
walk	0	1	0	0	8

Figure A17. Confusion matrices for the experiment using the 'lower-occlusion' database, shorterMBR and C1 correlation.

	bend	jack	pjump	wave1	wave2
bend	8	0	1	0	0
jack	0	8	1	0	0
pjump	0	1	7	0	1
wave1	0	0	1	5	3
wave2	0	0	0	1	8

	jump	run	side	skip	walk
jump	4	1	0	0	4
run	1	6	1	0	1
side	0	0	7	2	0
skip	1	1	2	4	1
walk	0	0	0	2	7

Figure A18. Confusion matrices for the experiment using the 'lower-occlusion' database, MBR perimeter and C1 correlation.

	bend	jack	pjump	wave1	wave2
bend	7	1	0	1	0
jack	0	9	0	0	0
pjump	0	1	7	0	1
wave1	0	0	0	5	4
wave2	0	0	0	2	7

	jump	run	side	skip	walk
jump	6	0	0	3	0
run	0	8	1	0	0
side	0	1	7	1	0
skip	1	0	1	7	0
walk	0	0	0	0	9

Figure A19. Confusion matrices for the experiment using the 'upper-occlusion' database, rectangularity and Euclidean distance.

	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	9	0	0	0
pjump	0	0	7	1	1
wave1	0	1	0	7	1
wave2	0	0	1	3	5

	jump	run	side	skip	walk
jump	7	0	1	1	0
run	0	8	0	1	0
side	1	1	6	0	1
skip	1	1	0	7	0
walk	0	0	0	0	9

Figure A20. Confusion matrices for the experiment using the 'upper-occlusion' database, elongation and Euclidean distance.

	bend	jack	pjump	wave1	wave2
bend	7	0	0	2	0
jack	2	7	0	0	0
pjump	2	0	5	1	1
wave1	0	1	0	5	3
wave2	0	0	1	1	7

	jump	run	side	skip	walk
jump	5	1	0	2	1
run	1	8	0	0	0
side	0	0	9	0	0
skip	5	0	1	3	0
walk	0	0	1	0	8

Figure A21. Confusion matrices for the experiment using the 'upper-occlusion' database, eccentricity and Euclidean distance.

	bend	jack	pjump	wave1	wave2
bend	8	1	0	0	0
jack	1	8	0	0	0
pjump	1	0	7	1	0
wave1	0	1	0	5	3
wave2	0	1	0	3	5

	jump	run	side	skip	walk
jump	4	1	1	3	0
run	0	5	1	3	0
side	0	0	8	0	1
skip	4	2	0	2	1
walk	0	0	1	0	8

Figure A22. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR width and Euclidean distance.

	bend	jack	pjump	wave1	wave2
bend	6	1	0	0	2
jack	2	5	1	0	1
pjump	1	2	3	1	2
wave1	1	0	2	5	1
wave2	2	0	0	1	6

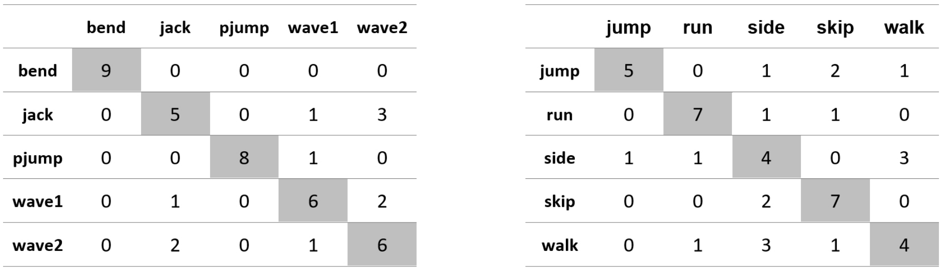
	jump	run	side	skip	walk
jump	5	0	0	2	2
run	0	9	0	0	0
side	0	0	8	0	1
skip	4	0	0	5	0
walk	1	0	1	0	7

Figure A23. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR length and Euclidean distance.

	bend	jack	pjump	wave1	wave2
bend	8	0	0	0	1
jack	0	9	0	0	0
pjump	0	2	5	2	0
wave1	1	0	1	4	3
wave2	3	0	0	1	5

	jump	run	side	skip	walk
jump	5	1	2	1	0
run	1	8	0	0	0
side	0	0	9	0	0
skip	0	0	1	8	0
walk	0	0	1	0	8

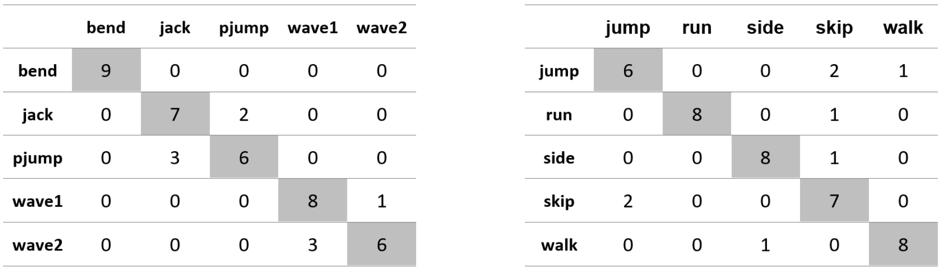
Figure A24. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR area and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	5	0	1	3
pjump	0	0	8	1	0
wave1	0	1	0	6	2
wave2	0	2	0	1	6

	jump	run	side	skip	walk
jump	5	0	1	2	1
run	0	7	1	1	0
side	1	1	4	0	3
skip	0	0	2	7	0
walk	0	1	3	1	4

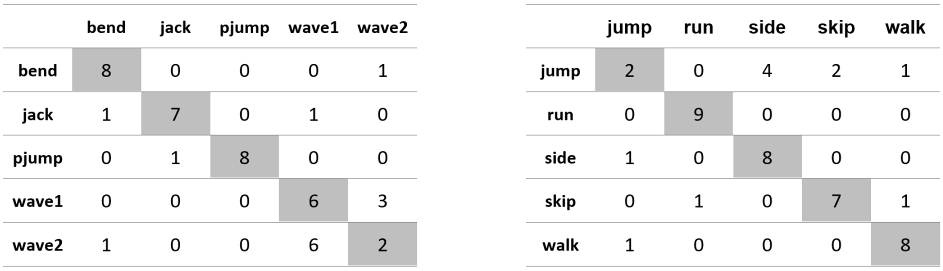
Figure A25. Confusion matrices for the experiment using the 'upper-occlusion' database, longerMBR and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	7	2	0	0
pjump	0	3	6	0	0
wave1	0	0	0	8	1
wave2	0	0	0	3	6

	jump	run	side	skip	walk
jump	6	0	0	2	1
run	0	8	0	1	0
side	0	0	8	1	0
skip	2	0	0	7	0
walk	0	0	1	0	8

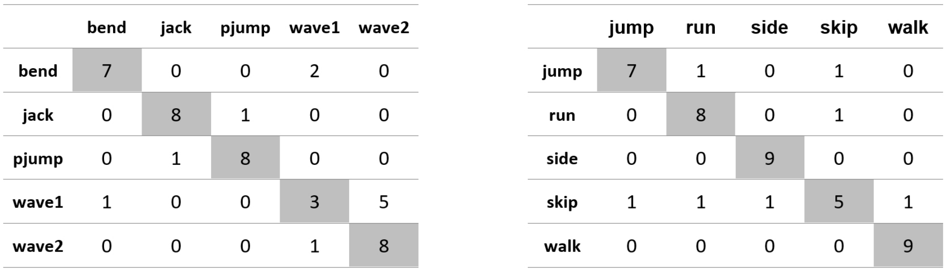
Figure A26. Confusion matrices for the experiment using the 'upper-occlusion' database, shorterMBR and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	8	0	0	0	1
jack	1	7	0	1	0
pjump	0	1	8	0	0
wave1	0	0	0	6	3
wave2	1	0	0	6	2

	jump	run	side	skip	walk
jump	2	0	4	2	1
run	0	9	0	0	0
side	1	0	8	0	0
skip	0	1	0	7	1
walk	1	0	0	0	8

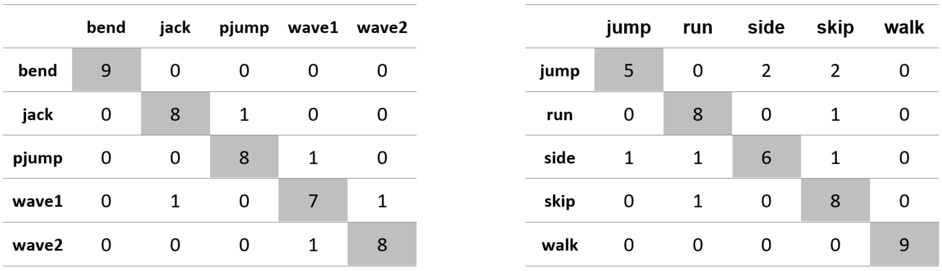
Figure A27. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR perimeter and Euclidean distance.



	bend	jack	pjump	wave1	wave2
bend	7	0	0	2	0
jack	0	8	1	0	0
pjump	0	1	8	0	0
wave1	1	0	0	3	5
wave2	0	0	0	1	8

	jump	run	side	skip	walk
jump	7	1	0	1	0
run	0	8	0	1	0
side	0	0	9	0	0
skip	1	1	1	5	1
walk	0	0	0	0	9

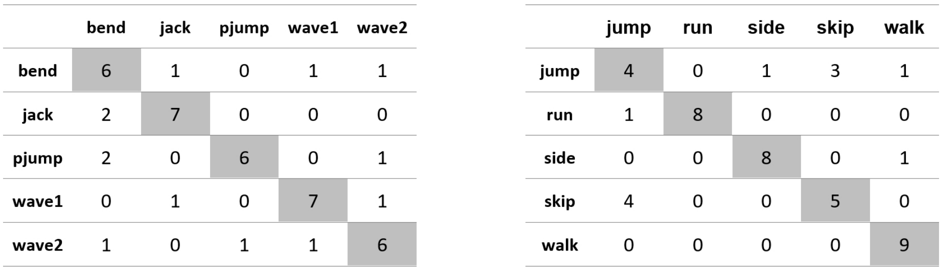
Figure A28. Confusion matrices for the experiment using the 'upper-occlusion' database, rectangularity and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	8	1	0	0
pjump	0	0	8	1	0
wave1	0	1	0	7	1
wave2	0	0	0	1	8

	jump	run	side	skip	walk
jump	5	0	2	2	0
run	0	8	0	1	0
side	1	1	6	1	0
skip	0	1	0	8	0
walk	0	0	0	0	9

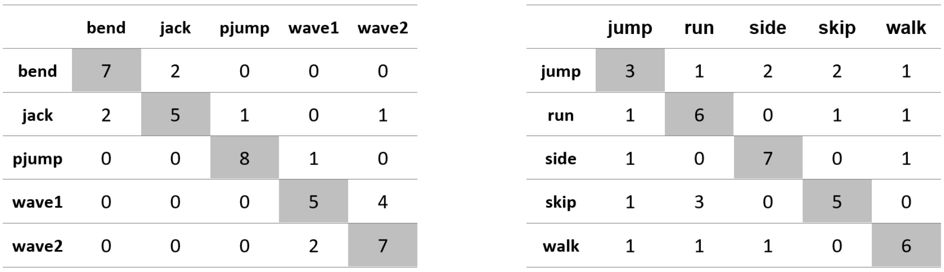
Figure A29. Confusion matrices for the experiment using the 'upper-occlusion' database, elongation and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	6	1	0	1	1
jack	2	7	0	0	0
pjump	2	0	6	0	1
wave1	0	1	0	7	1
wave2	1	0	1	1	6

	jump	run	side	skip	walk
jump	4	0	1	3	1
run	1	8	0	0	0
side	0	0	8	0	1
skip	4	0	0	5	0
walk	0	0	0	0	9

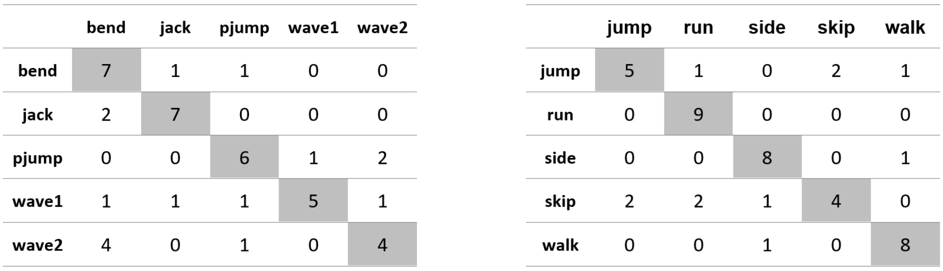
Figure A30. Confusion matrices for the experiment using the 'upper-occlusion' database, eccentricity and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	7	2	0	0	0
jack	2	5	1	0	1
pjump	0	0	8	1	0
wave1	0	0	0	5	4
wave2	0	0	0	2	7

	jump	run	side	skip	walk
jump	3	1	2	2	1
run	1	6	0	1	1
side	1	0	7	0	1
skip	1	3	0	5	0
walk	1	1	1	0	6

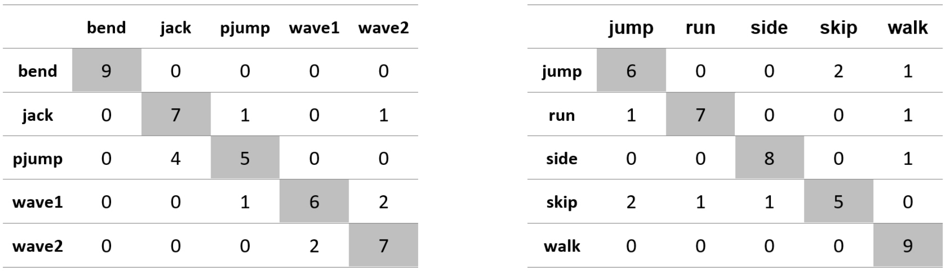
Figure A31. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR width and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	7	1	1	0	0
jack	2	7	0	0	0
pjump	0	0	6	1	2
wave1	1	1	1	5	1
wave2	4	0	1	0	4

	jump	run	side	skip	walk
jump	5	1	0	2	1
run	0	9	0	0	0
side	0	0	8	0	1
skip	2	2	1	4	0
walk	0	0	1	0	8

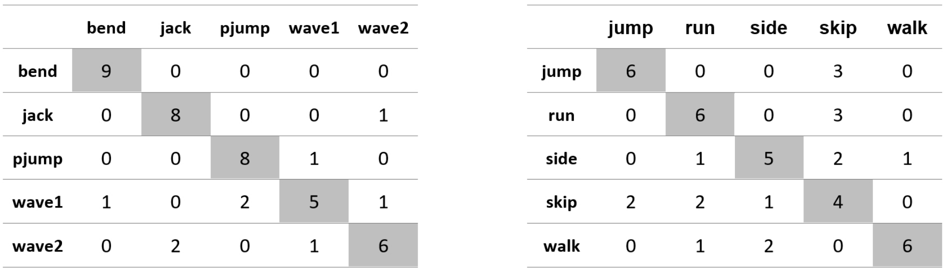
Figure A32. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR length and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	7	1	0	1
pjump	0	4	5	0	0
wave1	0	0	1	6	2
wave2	0	0	0	2	7

	jump	run	side	skip	walk
jump	6	0	0	2	1
run	1	7	0	0	1
side	0	0	8	0	1
skip	2	1	1	5	0
walk	0	0	0	0	9

Figure A33. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR area and C1 correlation.



	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	8	0	0	1
pjump	0	0	8	1	0
wave1	1	0	2	5	1
wave2	0	2	0	1	6

	jump	run	side	skip	walk
jump	6	0	0	3	0
run	0	6	0	3	0
side	0	1	5	2	1
skip	2	2	1	4	0
walk	0	1	2	0	6

Figure A34. Confusion matrices for the experiment using the 'upper-occlusion' database, longerMBR and C1 correlation.

	bend	jack	pjump	wave1	wave2
bend	9	0	0	0	0
jack	0	8	1	0	0
pjump	0	1	7	1	0
wave1	0	0	0	7	2
wave2	0	0	0	4	5

	jump	run	side	skip	walk
jump	6	0	2	1	0
run	0	9	0	0	0
side	1	0	8	0	0
skip	1	0	1	7	0
walk	0	0	1	0	8

Figure A35. Confusion matrices for the experiment using the 'upper-occlusion' database, shorterMBR and C1 correlation.

	bend	jack	pjump	wave1	wave2
bend	5	0	0	3	1
jack	0	9	0	0	0
pjump	0	0	9	0	0
wave1	1	1	0	3	4
wave2	1	0	0	3	5

	jump	run	side	skip	walk
jump	3	0	2	2	2
run	1	7	0	1	0
side	0	0	8	0	1
skip	0	1	0	7	1
walk	0	0	0	0	9

Figure A36. Confusion matrices for the experiment using the 'upper-occlusion' database, MBR perimeter and C1 correlation.

References

1. Lin, W.; Sun, M.T.; Poovandran, R.; Zhang, Z. Human activity recognition for video surveillance. In Proceedings of the 2008 IEEE International Symposium on Circuits and Systems, Seattle, WA, USA, 18–21 May 2008; pp. 2737–2740, doi:10.1109/ISCAS.2008.4542023.
2. Vishwakarma, S.; Agrawal, A. A survey on activity recognition and behavior understanding in video surveillance. *Vis. Comput.* **2013**, *29*, 983–1009, doi:10.1007/s00371-012-0752-6.
3. Duchenne, O.; Laptev, I.; Sivic, J.; Bach, F.; Ponce, J. Automatic annotation of human actions in video. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 1491–1498, doi:10.1109/ICCV.2009.5459279.
4. Papadopoulos, G.; Axenopoulos, A.; Daras, P. Real-Time Skeleton-Tracking-Based Human Action Recognition Using Kinect Data. In Proceedings of the International Conference on Multimedia Modeling, Dublin, Ireland, 6–10 January 2014; Volume 8325, pp. 473–483, doi:10.1007/978-3-319-04114-8_40.
5. Rautaray, S.; Agrawal, A. Vision based Hand Gesture Recognition for Human Computer Interaction: A Survey. *Artif. Intell. Rev.* **2015**, *43*, 1–54, doi:10.1007/s10462-012-9356-9.
6. El murabet, A.; Abtoy, A.; Touhafi, A.; Tahiri, A. Ambient Assisted living system's models and architectures: A survey of the state of the art. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *32*, 1–10, doi:10.1016/j.jksuci.2018.04.009.
7. Schrader, L.; Toro, A.; Konietzny, S.; Rüping, S.; Schäpers, B.; Steinböck, M.; Krewer, C.; Mueller, F.; Guettler, J.; Bock, T. Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People. *J. Popul. Ageing* **2020**, *13*, 139–165, doi:10.1007/s12062-020-09260-z.
8. Turaga, P.; Chellappa, R.; Subrahmanian, V.S.; Udrea, O. Machine Recognition of Human Activities: A Survey. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1473–1488, doi:10.1109/TCSVT.2008.2005594.
9. Poppe, R. A survey on vision-based human action recognition. *Image Vis. Comput.* **2010**, *28*, 976–990, doi:10.1016/j.imavis.2009.11.014.
10. 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, doi:10.1016/j.eswa.2012.03.005.
11. Chandel, H.; Vatta, S. Occlusion Detection and Handling: A Review. *Int. J. Comput. Appl.* **2015**, *120*, 33–38, doi:10.5120/21264-3857.
12. Al-Faris, M.; Chiverton, J.; Ndzi, D.; Ahmed, A.I. A Review on Computer Vision-Based Methods for Human Action Recognition. *J. Imaging* **2020**, *6*, 46, doi:10.3390/jimaging6060046.
13. World Health Organization. Global Recomendations on Physical Acitivity. 2011. Available online: https://www.who.int/dietphysicalactivity/physical-activity-recommendations-18-64years.pdf (accessed on 5 July 2021).
14. Wilke, J.; Mohr, L.; Tenforde, A.S.; Edouard, P.; Fossati, C.; González-Gross, M.; Ramirez, C.S.; Laiño, F.; Tan, B.; Pillay, J.D.; et al. Restrictercise! Preferences Regarding Digital Home Training Programs during Confinements Associated with the COVID-19 Pandemic. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6515, doi:10.3390/ijerph17186515.
15. Polero, P.; Rebollo-Seco, C.; Adsuar, J.; Perez-Gomez, J.; Rojo Ramos, J.; Manzano-Redondo, F.; Garcia-Gordillo, M.; Carlos-Vivas, J. Physical Activity Recommendations during COVID-19: Narrative Review. *Int. J. Environ. Res. Public Health* **2020**, *18*, 65, doi:10.3390/ijerph18010065.

16. Füzéki, E.; Schröder, J.; Carraro, N.; Merlo, L.; Reer, R.; Groneberg, D.A.; Banzer, W. Physical Activity during the First COVID-19-Related Lockdown in Italy. *Int. J. Environ. Res. Public Health* **2021**, *18*, 2511, doi:10.3390/ijerph18052511.
17. Robertson, M.; Duffy, F.; Newman, E.; Prieto Bravo, C.; Ates, H.H.; Sharpe, H. Exploring changes in body image, eating and exercise during the COVID-19 lockdown: A UK survey. *Appetite* **2021**, *159*, 105062, doi:10.1016/j.appet.2020.105062.
18. Stockwell, S.; Trott, M.; Tully, M.; Shin, J.; Barnett, Y.; Butler, L.; McDermott, D.; Schuch, F.; Smith, L. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: A systematic review. *BMJ Open Sport Exerc. Med.* **2021**, *7*, e000960, doi:10.1136/bmjsem-2020-000960.
19. Wolf, S.; Seiffer, B.; Zeibig, J.M.; Welkerling, J.; Brokmeier, L.; Atrott, B.; Ehring, T.; Schuch, F. Is Physical Activity Associated with Less Depression and Anxiety During the COVID-19 Pandemic? A Rapid Systematic Review. *Sport. Med.* **2021**, *51*, 1771–1783, doi:10.1007/s40279-021-01468-z.
20. World Health Organization. #HealthyAtHome—Physical Activity. 2021. Available online: https://www.who.int/news-room/campaigns/connecting-the-world-to-combat-coronavirus/healthyathome/healthyathome---physical-activity (accessed on 5 July 2021).
21. 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, Beijing, China, 17–21 October 2005; Volume 2, pp. 1395–1402, doi:10.1109/ICCV.2005.28.
22. Bobick, A.; Davis, J. The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 257–267, doi:10.1109/34.910878.
23. Eweiwi, A.; Cheema, M.S.; Thurau, C.; Bauckhage, C. Temporal key poses for human action recognition. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops, Barcelona, Spain, 6–13 November 2011; pp. 1310–1317, doi:10.1109/ICCVW.2011.6130403.
24. Ahad, M.A.R.; Islam, N.; Jahan, I. Action recognition based on binary patterns of action-history and histogram of oriented gradient. *J. Multimodal User Interfaces* **2016**, *10*, 335–344, doi:10.1007/s12193-016-0229-4.
25. 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, doi:10.1016/j.procs.2015.07.515.
26. 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, Switzerlands, 2015; pp. 11–47, doi:10.1007/978-3-319-11430-92.
27. Junejo, I.N.; Junejo, K.N.; Aghbari, Z.A. Silhouette-based human action recognition using SAX-Shapes. *Vis. Comput.* **2014**, *30*, 259–269, doi:10.1007/s00371-013-0842-0.
28. 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, doi:10.1109/ICPR.2010.427.
29. 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*, 547069, doi:10.1155/2014/547069.
30. Sargano, A.B.; Angelov, P.; Habib, Z. Human Action Recognition from Multiple Views Based on View-Invariant Feature Descriptor Using Support Vector Machines. *Appl. Sci.* **2016**, *6*, 309, doi:10.3390/app6100309.
31. Hsieh, C.H.; Huang, P.; Tang, M.D. Human Action Recognition Using Silhouette Histogram. In Proceedings of the 34th Australasian Computer Science Conference, Perth, Australia, 17–20 January 2011; Volume 113, pp. 11–16, doi:10.5555/2459296.2459298.
32. Beddiar, D.R.; Nini, B.; Sabokrou, M.; Hadid, A. Vision-based human activity recognition: A survey. *Multimed. Tools Appl.* **2020**, *79*, 30509–30555, doi:10.1007/s11042-020-09004-3.
33. Li, W.; Zhang, Z.; Liu, Z. Action recognition based on a bag of 3D points. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13-18 June 2010; pp. 9–14, doi:10.1109/CVPRW.2010.5543273.
34. Abe, T.; Fukushi, M.; Ueda, D. Primitive Human Action Recognition Based on Partitioned Silhouette Block Matching. In *Advances in Visual Computing*; Bebis, G., Boyle, R., Parvin, B., Koracin, D., Li, B., Porikli, F., Zordan, V., Klosowski, J., Coquillart, S., Luo, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 308–317, doi:10.1007/978-3-642-41939-3_30.
35. Jargalsaikhan, I.; Direkoglu, C.; Little, S.; O'Connor, N.E. An Evaluation of Local Action Descriptors for Human Action Classification in the Presence of Occlusion. In *MultiMedia Modeling*; Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N., Eds.; Springer International Publishing: Cham, Switzerlands, 2014; pp. 56–67, doi:10.1007/978-3-319-04117-9_6.
36. Gościewska, K.; Frejlichowski, D. The Analysis of Shape Features for the Purpose of Exercise Types Classification Using Silhouette Sequences. *Appl. Sci.* **2020**, *10*, 6728, doi:10.3390/app10196728.
37. 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, doi:10.7727/wimj.2018.177.
38. Yang, M.; Kpalma, K.; Ronsin, J. A Survey of Shape Feature Extraction Techniques. In *Pattern Recognition Techniques, Technology and Applications*; Peng-Yeng Yin, Ed.; I-Tech, Vienna, Austria, 2008; pp. 43–90.
39. Rosin, P. Computing global shape measures. In *Handbook of Pattern Recognition and Computer Vision*; World Scientific: Singapore, 2005; pp. 177–196, doi:10.1142/9789812775320_0010.
40. Brunelli, R.; Messelodi, S. Robust estimation of correlation with applications to computer vision. *Pattern Recognit.* **1995**, *28*, 833–841, doi:10.1016/0031-3203(94)00170-Q.

41. 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, doi:10.5772/4960.
42. Zhang, D.; Lu, G. A comparative Study of Fourier Descriptors for Shape Representation and Retrieval. In Proceedings of the 5th Asian Conference on Computer Vision, Melbourne, Australia, 22–25 January 2002; pp. 646–651.