J Intell Robot Syst
DOI 10.1007/s10846-011-9540-1

Convex Deficiencies for Human Action Recognition

Mabel Iglesias-Ham · Edel Bartolo García-Reyes · Walter George Kropatsch · Nicole Maria Artner

Received: 10 August 2010 / Accepted: 9 January 2011

Abstract A human action can be identified by visualizing the sequence of 2D binary projections over time. Here, one of the most representative features is shape and a wide range of algorithms have been proposed using its descriptors. This paper proposes convex deficiencies, the difference between an object and its convex hull, to be considered as a representation for the human action classification problem. A simple description using the centroids of the convex deficiencies over time is presented. Recognition of human actions is done with a fast matching algorithm that considers the spatial distribution of the centroid trajectories and the shape of the clusters in its 2D projection. The proposed representation is robust to deformations, scale, speed of the performed action and to the starting point of the movement sequence. Experiments using the videos of the Weizmann database show promising results demonstrating the effectiveness of the proposed methodology in classifying simple human actions, e.g. walking and running. The new proposed methodology should be extendable to a broader set of actions.

Keywords Convex deficiencies · Human motion classification · Pattern recognition

This work was supported by the Austrian Science Fund under grant P18716-N13.

M. Iglesias-Ham (✉) · E. B. García-Reyes
Pattern Recognition Department, CENATAV, Havana, Cuba
e-mail: miglesias@cenatav.co.cu, mabel@prip.tuwien.ac.at

E. B. García-Reyes
e-mail: egarcia@cenatav.co.cu

M. Iglesias-Ham · W. G. Kropatsch · N. M. Artner
PRIP, Vienna University of Technology, Vienna, Austria

W. G. Kropatsch
e-mail: krw@prip.tuwien.ac.at

N. M. Artner
e-mail: artner@prip.tuwien.ac.at

Published online: 19 January 2011

1 Introduction

Human action recognition is a highly active area of research in computer vision due to its application in video surveillance, human-computer interaction, video indexing and browsing, automatic video-annotation and so forth. The first studies in action recognition were based on the use of markers attached to the human body to accurately capture the motion patterns in a binary camera [22].

The present work focuses on the analysis and representation of the global pattern of human motion (not the motion of particular parts such as the face or hands to recognize expressions or gestures usually used in human-computer interaction applications) and the usage of this representation to do action recognition (e.g. running and walking). We propose the difference between an object and its convex hull, the *convex deficiencies*, to be considered as a representation for the human action classification problem. There is not, to the authors knowledge, any other published approach that uses the convex deficiencies to recognize human actions. A simple action description using the centroids of the convex deficiencies over time is presented. Recognition of human actions is done with a fast matching algorithm that considers the spatial distribution of the centroid trajectories and the shape of the resulting centroid clusters in its 2D projection. This is a novel representation, which aims to capture the dynamics of the human shape in time, robust to boundary noise and some deformations e.g. included by changes in appearance. The representation is simple and efficient and could be employed for practical applications like real time action recognition. The proposed feature is robust to deformations, scale, speed of the performed action and to the starting point of the movement sequence. Our contribution is to present a new methodology which is demonstrated by a few specific action recognition tasks (full body). The proposed methodology should be applicable to other actions than running and walking, and could be also feasible to apply in a more general task as shape recognition.

1.1 Related Work

Previous works use the *convex hull* to obtain a hierarchical representation of concavities [7, 10] and match 2D shapes [3].Those approaches aim at describing single 2D shapes, with no motion representation. Convex deficiencies have been applied to hand printed character recognition tasks [28, 30]. The human silhouette in time, can be seen as a deformable shape that changes from person to person performing the same action, or even from different occurrences of the same action performed by the same person. Nevertheless, the general structure of the shape is suitable to represent and identify actions.

Another possibility to represent actions is to use *optical flow* estimations [9, 19]. The optical flow information enables to know the probable displacement of the objects present in a scene. Optical flow computation can give undesired results in the presence of soft or homogeneous surfaces and discontinuities, particularly for tracking purposes. Nevertheless, the optical flow feature is independent from changes in appearance, compared to other measures like spatial or temporal gradients.

Spatio Temporal Templates, first introduced by [4], are static vector-images where the vector value at each point is a function of the motion properties at the corresponding spatial location in an image sequence. However, such templates are too sensitive to the different movement durations. In [6], a star figure enclosed by a convex polygon is built to represent the extremities of the silhouette of the human body. A video is then represented as a sequence of those star figure's parameters, which is regarded as a spatio-temporal template. The proposed representation could be enclosed into this approach, as we represent the whole sequence by a single *template/model* containing structural features of the foreground centroid and its spatial relation to the trajectories of the convex deficiencies centroids. However, using the shape of the resulting centroid clusters in its 2D projection, allows the feature to be more stable to different movement durations.

Key frames have also been used to identify an action [5, 12], reducing the whole sequence to a minimum set of non frequent frames (usually 1) and avoiding redundancies. The drawback of this representation is that it lacks information about motion, and there is a limited set of actions that can be identified by a single frame. Even a human observer could have problems in identifying the correct action with only a single frame.

In the *Bag of Words* approach a pose is understood as a word in a vocabulary (set of predefined poses), and a set or sequence of these words is used as a representation for actions. Likewise, any other local feature could be used as a word, and the set of local features in the whole sequence as a representation [11, 16, 20].

Finally, there is a group of works approaching the human action recognition task using *volumetric representations*. These volumes are obtained by concatenating 2D slices over time. Numerous publications show the usefulness of analyzing actions as spatio-temporal volumes of some local feature like gradient, optical flow, intensity, etc [14, 15, 32]. Unfortunately, these approaches do not consider the sequential ordering of the spatial shape in time in the representation. For that reason, symmetrical actions performed backwards can not be differentiated from the same action performed forwards. Space-time interest points computed on the volumes, have been proposed to compensate variations in relative motion between an object and the camera when computing space-time descriptors [25]. This representation uses interest points to adapt features to the local velocity of the image pattern to make it stable to different amounts of camera motion. Interest points are affected by the camera view point product of the variations in shape of the projected 2D human figure. However, space-time interest points have been used as a stable feature for velocity and scale variations in the video capture process.

This paper is organized as follows: Section 2 shows the proposed representation based on convex deficiencies. In Section 3 the matching algorithm is described, followed by experimental results using the benchmark database Weizmann in Section 4. Section 5 concludes the paper and gives an outlook on future work.

2 Convex Deficiencies: A Representation for Human Actions

We assume that we can obtain a binary sequence of images from every video representing the human silhouette (2D connected region) as foreground and the rest of the scene as background as provided in [14]. We used the provided silhouettes as we do not intend to address the foreground detection issue in this work. The advantages of using silhouettes are that they contain rich shape information, a series of body silhouette images is independent of the speed of movement, and silhouettes are easier to extract compared with tracking body parts.

The **convex hull** H of an object (e.g. human) O is the minimal convex set containing the object [2], i.e. $\nexists H'$ convex, $O \subseteq H'$ s.t. $H' \subset H$. It is also common to call the boundary of the minimal convex set containing the given object a convex hull [2], i.e. ∂H. We consider the first definition, in which the convex hull of an object has the same dimension as the object, and a convex object coincides with its convex hull. The **convex deficiencies** are the connected components of the difference between the object and its convex hull (has the same dimension as the object). Then, the convex deficiencies of a foreground region are 2D connected components representing concavities and holes of the shape.

We compute the convex hull to differentiate the convex deficiencies from the rest of the background. In order to obtain the connected components of the convex deficiencies and its boundary relations with the foreground we can build a combinatorial pyramid [21, 23]. A combinatorial pyramid can be constructed with $log(m)$ height in the number of vertices in the base level [17, 24]. The advantage of using this hierarchical, structural representation is that topology is preserved in every level, and it is a feature we can use to support recognition. Additionally, we can use the top level as a reduced representation of the original image to perform further computations [13, 29]. The relative position of the convex deficiencies with respect to the foreground is an important feature for the representation and is invariant to scale changes. In Fig. 1, a selection of frames of a video sequence are used to show the top level boundary graphs of the pyramids, representing the variations of every convex deficiency and its global structure around the shape. The gray regions around the body represent the biggest and most representative convex deficiencies as a demonstration.

The sequence of convex deficiencies in time has properties that allow to identify the performed action. Considering general features of its dynamics makes the description robust to scale (e.g. relative position with respect to the center of gravity of the foreground region), to deformations (e.g. centroid of the convex deficiency region), and small variations in view point (e.g. the number of convex deficiencies). Using this information, it is not necessary to know accurate trajectories of joint points of the human body beforehand to obtain a meaningful structure in recognition tasks.

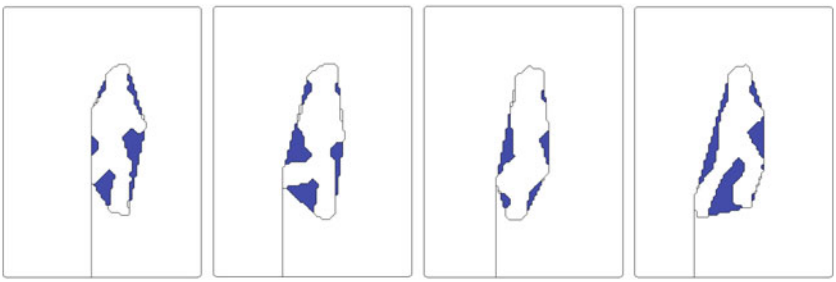

Fig. 1 Sequence of human silhouettes represented by the *top level* boundary graphs in a pyramid with regions including: convex deficiencies arranged around the shape of the object (*gray*), background and human shape (*foreground*)

It is not necessary either to perform morphological operations to get a skeleton of the human shape [1], which saves time.

The description uses the centroids of the convex deficiencies over time. Figure 2 shows the usefulness of the centroids of the convex deficiencies to distinguish actions. In the figure, one can see how the bending action could be easily differentiated from walking and running. However, the centroid trajectories (positions over time) of running and walking are similar, which is consistent with the consideration that they use very similar movements. By projecting the whole set of points in the 2D plane it is possible to achieve a more differentiating pattern for actions, which allows to even distinguish between walking and running (see Fig. 4).

Figures 3 and 4 visualize how the 2D projections of the centroid trajectories show a similar pattern for videos of different persons performing the same action. The projection removes the temporal order of the centroids and therefore it is robust to the starting point of the motion cycle in the sequence and to the speed of the performed action. Hence, there is no need to normalize the cycle length or starting point to compare two videos (two actions).

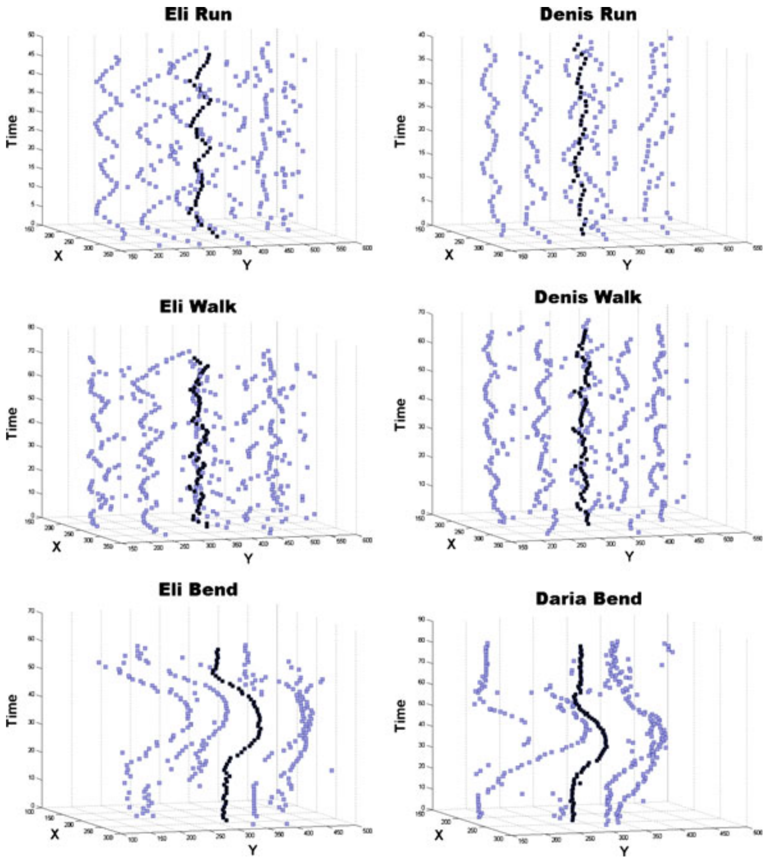

Fig. 2 Centroids of the convex deficiencies of 6 video sequences (*light colored squares*) and the *center of gravity* of the foreground region in the *middle* (*dark colored squares*)

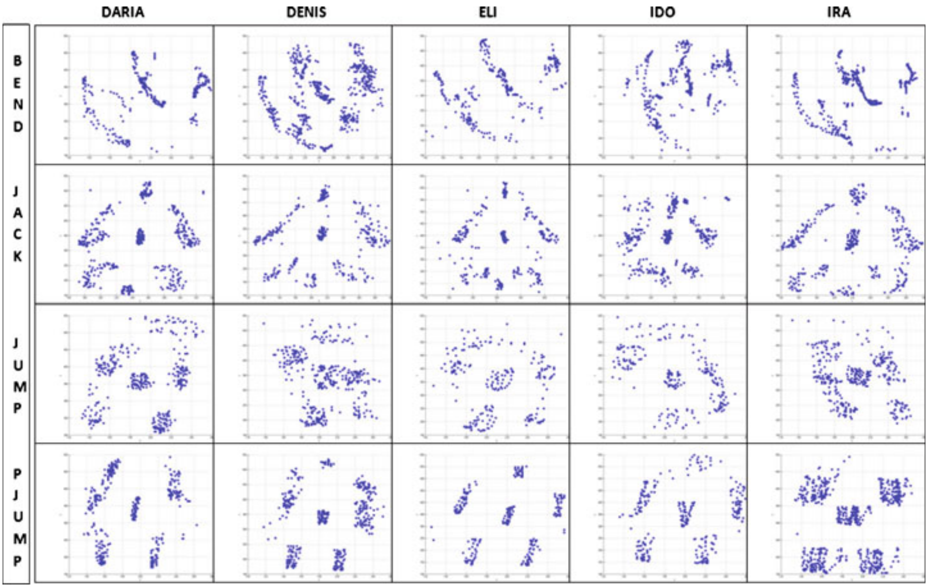

Fig. 3 2D projections of the centroid trajectories of the convex deficiencies and the foreground, for 5 persons and 4 actions in the Weizmann database

The spatial arrangement of the convex deficiencies around the silhouette is an important property as it represents its position around the body. In Figs. 3 and 4, the group in the center of the points represents the center of gravity of the foreground

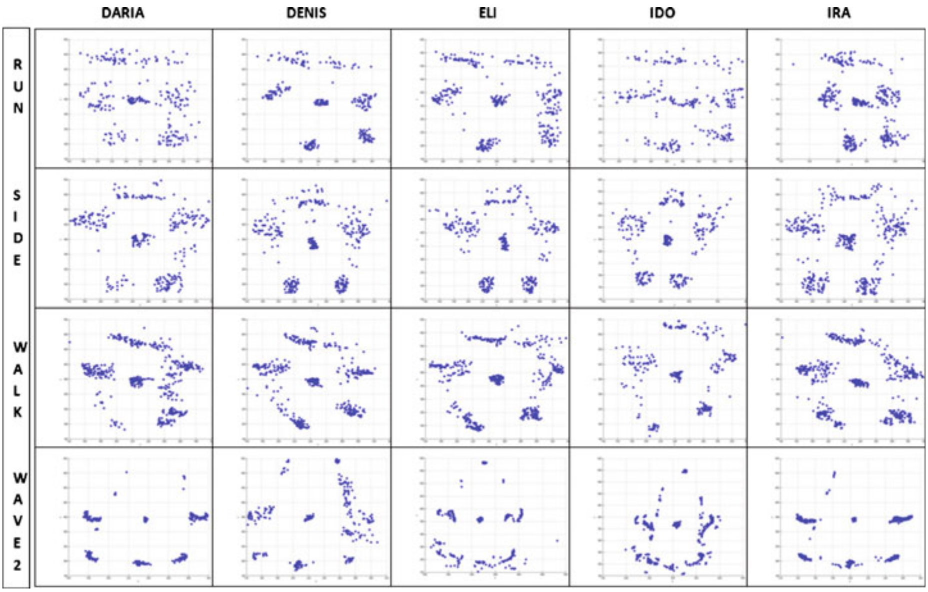

Fig. 4 2D projections of the centroid trajectories of the convex deficiencies and the foreground, for 5 persons and 4 actions in the Weizmann database. See walking and running actions

regions. For the description we group the centroids (points in figures) associated to the same convex deficiency in time (trajectory) and describe its shape in 2D. The grouping/clustering is performed by associating each centroid point in frame t, to the closest points in frame $t+1$ (distance less than 40 pixels, threshold was determined empirically). The result of the clustering is a robust separation of the different groups of centroids as the convex deficiencies are clearly separated in space (in the experiments with the Weizmann database).

Then, the variance/covariance matrix for each group of points and its spectral decomposition is computed to estimate the magnitude of its main directions r_1 and r_2, its main orientation O_{r_1}, and the correlation value *corr*. Finally, for every convex deficiency group the orientation between its centroid and the centroid of the foreground region group O_c is determined, which results in a 5-dimensional feature vector $(r_1, r_2, O_{r_1}, O_c, corr)$. To allow the description to be invariant to scale, r_1 and r_2 are normalized with respect to the average area of the foreground region in the video.

The proposed representation has as an advantage that uses motion information, when approaches like key-frames lack of this important feature. Unlike spatio temporal templates, it is robust to variations in length of the performed action, since it is based on the estimation of the probability function of the 2D-projections of the convex deficiencies centroid.

3 Human Action Matching Algorithm

In order to show preliminary results of our novel description for action recognition, we focus on the challenging problem of distinguishing walking from running and employ a simple matching algorithm. The algorithm finds the similarity between the two sets of 5-dimensional feature vectors associated to the groups of convex deficiency centroids. The groups of convex deficiency centroids appears as a star like shape in the 2D projections, with the center being the group of centroids from the foreground region. Every group of centroids of a convex deficiency, can be associated with another that has a similar relative position with respect to the foreground centroids (respective O_c with difference less than 9°), similar orientations and magnitude in its main directions (respective O_{r_1} with difference less than 15°, and r_1, r_2 with difference less than 0.03), and the difference in the correlation values *corr* must be less than 0.2. One convex deficiency group can not be associated with more that one group in the second video, neither two convex deficiencies groups can be associated to the same group in the second video.

The final similarity takes into account the number of matching convex deficiencies groups (*match*) regarding the video with the lowest number of convex deficiencies groups (*lowest_CD*), and the number that did not match (*non_match*) regarding the video with the biggest number of convex deficiencies groups (*biggest_CD*):

$$lowest\_CD = min(|video1.Groups|, |video2.Groups|);$$

$$biggest\_CD = max(|video1.Groups|, |video2.Groups|);$$

$$non\_match = biggest\_CD - match;$$

$$Sim = match/lowest\_CD - non\_match/biggest\_CD;$$

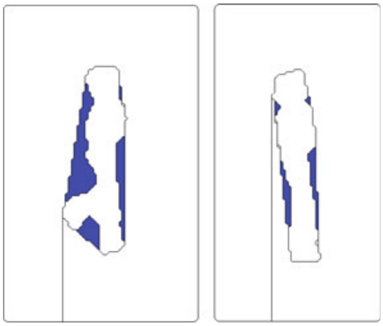

Fig. 5 Difference of convex deficiencies between skip (*left*) and jump (*right*) actions. Most of the silhouettes during the rest of the action show a similar pattern

It is important to consider the number of matching convex deficiencies groups, but also the number of convex deficiencies groups that did not match. Consider for example, the case of jumping and skipping actions, where most of the convex deficiencies appear in both actions. Jump and skip are the more conflicting actions in the Weizmann database (usually used as a benchmark), because of their similarity. However, there is a split in the convex deficiencies of the skip action which is not going to find a match in the jump shape and this is a key difference (see Fig. 5).

The similarity measure is going to have values between −1 and 1. −1 when non of the convex deficiencies groups found a match, and 1 when all the convex deficiencies found a match.

4 Experiments

The Weizmann database [14] contains 10 actions including bend (bend), jumping-jack (jack), jump-in-place (pjump), jump-forward (jump), run (run), gallop-sideways (side), jump-forward-one-leg (skip), walk (walk), wave one hand (wave1) and wave two hands (wave2) performed by 9 actors. There are many previous works using the selected database including [8, 14, 18, 26, 27, 31, 33]. We used the described database as its videos are accesible without restrictions and provides binary silhouettes (which is convenient for us as we do not focus in the segmentation step), to demonstrate the feasibility of the new methodology to describe actions. However, we are not aiming to compare our results with the state of the art as a final action recognition algorithm yet.

In the first experiment the task is to distinguish walking and running actions, which are common movements in video surveillance applications. This experiment uses the silhouettes subtracted from the background provided with the Weizmann dataset. There are 20 walking and running videos in total instead of 18 (9×2), as one of the actors (Lena) has two binary run videos and two walk videos. We performed a leave-one-out test, comparing one video sequence with the remaining 19. The most similar was always a video with another occurrence of the same action (100% recognition rate). We also performed a sligtly different test including the rest of the videos in the database, which is a total of 93 videos that includes an extra sequence for the lena_skip action. Table 1 shows that the walking action was classified correctly 8 out of 10 videos, and the other 2 were erroneosly classified as skip actions. However, the run action videos were all correctly clasified (100%).

Table 1 Recognition results for walking and running actions in the Weizmann database

Action	1st position	Misclassified
Walk	8	2 (skip)
Run	10	0

Table 2 First and second best results for each test sequence regarding the set of walking and running actions

Test sequence	1st best	2nd best
Normal walk	Walk	Walk
Walking in a skirt	Walk	Walk
Carrying briefcase	Walk	Run
Limping man	Walk	Run
Occluded Legs	Walk	Walk
Knees Up	Run	Walk
Walking with a dog	Run	Run
Sleepwalking	Walk	Run
Swinging a bag	Walk	Walk
Occluded by a pole	Walk	Walk

Table 3 First and second best results for each video regarding the whole set of 93 videos

Test sequence	1st best	2nd best
Normal walk	Walk	Walk
Walking in a skirt	Walk	Walk
Carrying briefcase	Side	Walk
Limping man	Walk	Side
Occluded Legs	Walk	Side
Knees Up	Skip	Run
Walking with a dog	Run	Run
Sleepwalking	Walk	Run
Swinging a bag	Walk	Walk
Occluded by a pole	Walk	Side

Table 4 Table showing the behavior of the representation in different changes of viewpoint

Test sequence	1st best	2nd best
Walking in 0°	Walk	Walk
Walking in 9°	Walk	Walk
Walking in 18°	Walk	Run
Walking in 27°	Walk	Walk
Walking in 36°	Walk	Run
Walking in 45°	Run	Run
Walking in 54°	Run	Walk
Walking in 63°	Walk	Run
Walking in 72°	Run	Run
Walking in 81°	Walk	Walk

In the field of video surveillance there is the problem of detecting abnormal behavior. The presented approach provides a simple method to distinguish walking people (normal behavior) from running people (abnormal behavior). It would allow to efficiently identify suspicious people and emit an alarm, which is very useful in practical video surveillance applications. The aim of this set of experiments based mainly on walk and run actions, is to show the potencial of the new representation as a very simple description which can be computed very fast, and its able to differentiate actions usually confusing as the selected ones.

The second experiment uses videos provided in [14] to show results in a preliminary robustness test. Here we have 10 videos in different scenarios including noise that affects the motion capture ('Normal walk', 'Walking in a skirt', 'Carrying briefcase', 'Limping man', 'Occluded legs', 'Knees up', 'Walking with a dog', 'Sleepwalking', 'Swinging a bag' and 'Occluded by a pole'). We compared each of the sequences with the 20 videos of persons walking and running. Results show that the presented approach is able to correctly recognize most of the videos, except for the sequences 'walking with a dog' and 'walking with the knees up', which considerably affects the convex deficiencies of the extracted shape (see Table 2). After the test with the 20 videos of running and walking, we extended the database to the whole set of actions and repeated the queries now with all 93 videos. The results in Table 3 shows that the result is comparable to the ones with only running and walking actions except for a misclassification in the case of the 'carrying briefcase' video. The results of this experiment testify that our solution is robust to some partial occlusions, changes in appearance, and non-rigid deformations in the extracted space-time shape. We plan to extend these experiments of robustness to a larger database.

Finally, we performed a recognition test changing the viewing angle (Table 4). We obtained correct recognition until the variations were less than 36° when comparing with running and walking actions. Starting from 45° the results were not stable. When adding the whole set of actions we get again a correct identification until 36° of variation. The results show that our representation has limitations in the angle of the performed action toward the camera. Yet, for small variations we can still obtain correct recognition rates.

5 Conclusions and Future Work

This paper presented a representation for human actions based on the convex deficiencies of the silhouette of humans in time. Convex deficiencies are the difference between an object and its convex hull. The convex deficiencies are clustered over time and the trajectories of their centroids are used to build a simple and efficient description, which can be used in practical video surveillance applications. We showed preliminary results of its effectiveness using the extracted silhouettes from the Weizmann database for classifying simple human actions (e.g. walking and running). Experiments demonstrate the potential of the new methodology, with a description which can be computed very fast, and its able to differentiate actions usually confusing as the selected ones.

Future work will aim to extend the proposed methodology using additional features extracted from its convex deficiencies (besides the centroids) and studying its behavior in time for the different human actions, extending the experiments to a larger database to evaluate its robustness in different scenarios (scale, appearance, etc.), and also to recognize other actions of interest. We plan to study the differences between convex deficiencies resulting from cavities and holes (e.g. influence on spatial arrangement, birth and death of cavities and holes). Furthermore, as combinatorial maps have proven useful for the efficient computation of topological invariants, we plan to explore the usefulness of topological features extracted in top of a combinatorial pyramid in the human action recognition problem, combining the topology of the human shape with the topology of its convex deficiencies.

Using only the centroids of the convex deficiencies we could achieve good recognition rates for walking and running actions. We believe, the proposed representation could be satisfactorily combined with previous solutions to improve the general problem of human action recognition. The human shape could be decomposed in meaningful parts (limbs and center), and associate each convex deficiency with its position around the body, to be considered in addition. However, as in this paper we used properties from the 2D projection of the centroid trajectories, we miss possibly relevant information from the sequential order of the convex deficiencies. Therefore, we plan to work on obtaining a more complete representation of the centroid trajectories to improve the recognition results.

References

1. Ali, S., Basharat, A., Shah, M.: Chaotic invariants for human action recognition. In: IEEE International Conference on Computer Vision, pp. 1–8 (2007). doi:http://doi.ieeecomputersociety.org/10.1109/ICCV.2007.4409046
2. Arcelli, C., Sanniti, G., Svensson, S.: Computing and analysing convex deficiencies to characterise 3d complex objects. Image Vis. Comput. **23**(2), 203–211 (2005)
3. Badawy, O.E., Kamel, M.: Matching concavity trees. In: Fred, A.L.N., Caelli, T., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.) SSPR/SPR. LNCS, vol. 3138, pp. 556–564. Springer (2004)
4. Bobick, A., Davis, J.: Real-time recognition of activity using temporal templates. In: WACV '96: Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV '96), p. 39. IEEE Computer Society, Washington, DC, USA (1996)
5. Carlsson, S., Sullivan, J.: Action recognition by shape matching to key frames. In: IEEE Computer Society Workshop on Models Versus Exemplars in Computer Vision (2001)
6. Chen, D.Y., Shih, S.W., Liao, H.Y.M.: Human action recognition using 2-d spatio-temporal templates. In: ICME, pp. 667–670 (2007)
7. Cohn, A.: A hierarchical representation of qualitative shape based on connection and convexity. In: Proc COSIT95. LNCS, pp. 311–326. Springer (1995)
8. Dollár, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal features. In: In VS-PETS, pp. 65–72 (2005)
9. Efros, A.A., Berg, A.C., Mori, G., Malik, J.: Recognizing action at a distance. In: Proc. Int. Conf. Computer Vision ICCV03 (2003)
10. El Badawy, O., Kamel, M.S.: Hierarchical representation of 2-d shapes using convex polygons: a contour-based approach. Pattern Recogn. Lett. **26**(7), 865–877 (2005). doi:10.1016/j.patrec.2004.09.031
11. Farhadi, A., Tabrizi, M.K.: Learning to recognize activities from the wrong view point. In: ECCV '08: Proceedings of the 10th European Conference on Computer Vision, pp. 154–166. Springer, Berlin, Heidelberg (2008)
12. González, J., Varona, J., Roca, F., Villanueva, J.J.: A human action comparison framework for motion understanding. In: In 6th Catalan Conference for Artificial Intelligence (CCIA'2003), vol. 100, pp. 168–177. Palma de Mallorca, Spain (2003)

13. González, R., Ion, A., Iglesias, M., Kropatsch, W.G.: Irregular graph pyramids and representative cocycles of cohomology generators. In: Torsello, A., Escolano, F., Brun, L. (eds.) Proc. 7th GbR. LNCS, vol. 5534, pp. 263–272. Springer, Venice, Italy (2009)
14. Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. Trans. Pattern Anal. Mach. Intell. **29**(12), 2247–2253 (2007). www.wisdom.weizmann.ac.il/∼vision/SpaceTimeActions.html
15. Grundmann, M., Meier, F., Essa, I.: 3d shape context and distance transform for action recognition. In: 19th International Conference on Pattern Recognition, ICPR2008, pp. 1–4. IEEE Computer Society Press (2008)
16. Hatun, K., Duygulu, P.: Pose sentences: a new representation for action recognition using sequence of pose words. In: 19th International Conference on Pattern Recognition, ICPR2008, pp. 1–4. IEEE Computer Society Press (2008)
17. Haxhimusa, Y., Saib, M., Langs, G., Kropatsch, W.G.: Logarithmic tapering graph pyramid. In: In Proc. of 24th DAGM Symposium LNCS (2002)
18. Huang, W., Wu, Q.M.J.: Incremental discriminative-analysis of canonical correlations for action recognition. In: In IEEE 12th International Conference on Computer Vision (ICCV) (2010)
19. Ikizler, N., Cinbis, R.G., Duygulu, P.: Human action recognition with line and flow histograms. In: ICPR, pp. 1–4. IEEE (2008)
20. Ikizler, N., Duygulu, P.: Human action recognition using distribution of oriented rectangular patches. In: Workshop on Human Motion—Understanding, Modeling, Capture and Animation. Lecture Notes in Computer Science, vol. 4814, pp. 271–284. Springer (2007)
21. Illetschko, T.: Minimal combinatorial maps for analyzing 3d data. Tech. Rep. PRIP-TR-110, Vienna University of Technology, Inst. of Computer Aided Automation, Pattern Recognition and Image Processing Group (2006)
22. Johansson, G.: Visual perception of biological motion and a model for its analysis. Percept. Psychophys. **14**, 201–211 (1973)
23. Kropatsch, W.G.: Building irregular pyramids by dual graph contraction. IEE Proc. Vis. Image Signal Process. **142**(6), 366–374 (1995)
24. Kropatsch, W.G., Haxhimusa, Y., Pizlo, Z., Langs, G.: Vision pyramids that do not grow too high. Pattern Recogn. Lett. **26**(3), 319–337 (2005). doi:10.1016/j.patrec.2004.10.026
25. Laptev, I., Lindeberg, T.: Velocity adaptation of space-time interest points. In: ICPR ’04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR’04), vol. 1, pp. 52–56. IEEE Computer Society, Washington, DC, USA (2004). doi:10.1109/ICPR.2004.971
26. Niebles, J.C., Fei-Fei, L.: A hierarchical model of shape and appearance for human action classification. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
27. Niebles, J.C., Wang, H., Fei-fei, L.: Unsupervised learning of human action categories using spatial-temporal words. In: In Proc. BMVC (2006)
28. Parker, J.R.: The use of convex deficiencies for the recognition of hand printed digits. In: Melter, R.A., Wu, A.Y. (eds.) Vision Geometry III, vol. 2356, pp. 169–175 (1994)
29. Peltier, S., Ion, A., Kropatsch, W.G., Damiand, G., Haxhimusa, Y.: Directly computing the generators of image homology using graph pyramids. Image Vis. Comput. **27**, 846–853 (2009)
30. Stavrianopoulou, A., Anastassopoulos, V.: The euler feature vector. Int. Conf. Pattern Recognit. **3**, 7034 (2000)
31. Wu, X., Liang, W., Jia, Y.: Human action recognition based on self organizing map. In: In ICASSP (2009)
32. Yilmaz, A., Shah, M.: A differential geometric approach to representing the human actions. Comput. Vis. Image Underst. **109**(3), 335–351 (2008). http://dblp.uni-trier.de/db/journals/cviu/cviu109.html#YilmazS08
33. Zelnik-manor, L., Irani, M.: Event-based analysis of video. In: In Proc. CVPR, pp. 123–130 (2001)