Advances in Intelligent Systems and Computing 314

Aleksander Zgrzywa
Kazimierz Choroś
Andrzej Siemiński *Editors*

New Research in Multimedia and Internet Systems

Springer

Chapter 6

Human Behavior Recognition Using Negative Curvature Minima and Positive Curvature Maxima Points

Krzysztof Kowalak, Łukasz Kamiński, Paweł Gardziński,
Sławomir Maćkowiak, and Radosław Hofman

Abstract. Recently, automated human behavior recognition are studied in the context of many new applications such as content-based video annotation and retrieval, highlight extraction, video summarization and video surveillance. In this chapter a novel description of human pose – a combination of negative curvature minima (NCM) and positive curvature maxima (PCM) points are proposed. Experimental results are provided in the chapter in order to demonstrate precision of the human activity recognition versus size of the descriptor (a temporal interval durations between the nodes of the model). The experimental results are focused on recognition of call for help behavior. The results prove high score of recognition of the proposed method.

6.1 Introduction

The past decade has witnessed a rapid proliferation of video cameras in all walks of life and has resulted in a tremendous explosion of video content. Several applications such as content-based video annotation and retrieval, highlight extraction and video summarization require recognition of the activities occurring in the video. The analysis of human activities in videos is an area with increasingly important consequences from security and surveillance to entertainment and personal archiving. In the area of surveillance, automated systems to observe pedestrian traffic areas and detect dangerous action are becoming important. This type of observation task is not well suited to humans, as it requires careful concentration over long periods of

Krzysztof Kowalak · Łukasz Kamiński · Paweł Gardziński · Sławomir Maćkowiak
Poznań University of Technology, Poland
e-mail: {kkowalak,lkaminski,pgardzinski,smack}@multimedia.edu.pl

Radosław Hofman
Smart Monitor sp. z o.o., Szczecin, Poland
e-mail: radekh@smartmonitor.pl

A. Zgrzywa et al. (eds.), *New Research in Multimedia and Internet Systems*,
Advances in Intelligent Systems and Computing 314, DOI: 10.1007/978-3-319-10383-9_6

time. Therefore, there is clear motivation to develop automated, intelligent, vision-based monitoring systems that can aid a human user in the process of risk detection and analysis. The system developed at Poznań University of Technology is able to perform fully automatic analysis of human behavior and recognition such behaviors as fainting, a fight or a call for help.

The scope of review is limited to some well-known graphical models that have been successfully applied on complex human activity modeling and behavior description in crowded public scenes. Activity modeling using single camera can be realized by many different methods, including: probabilistic graphical models (e.g. Bayesian networks [1,2], dynamic Bayesian networks [3-5], propagation nets [6]), probabilistic topic models (e.g. probabilistic latent semantic analysis models [7], latent Dirichlet allocation model [8, 9]), Petri nets [10], syntactic approaches [11] or rule-based approaches [12]. Bayesian networks are discussed in more detail because the approach proposed in this chapter belongs to them. Detailed reviews on other approaches are such as Petri nets, neural networks, synthetic approaches can be found in survey by Turaga et al. [13].

A Bayesian network or belief network is a probabilistic directed acyclic graphical model. Model consist of nodes which represent set of random variables (e.g. consecutive states of an event) and links which represent conditional dependencies between nodes. The strength of a dependency is parameterized by conditional probabilities that are attached to each cluster of parent-child nodes in the network. Such network has powerful capabilities in representing and reasoning state-oriented visual observations, so Bayesian network has been a popular tool for activity modeling. Bayesian network can be extended by introducing temporal dependencies between random variables. This extended network is called a dynamic Bayesian network.

Many different graphical models have been proposed for activity modeling. For instance, propagation nets [14], a subset of dynamic Bayesian networks with ability to explicitly model temporal interval durations, have been employed to capture the duration of temporal subintervals of multiple parallel streams of events.

This chapter is organized into 4 main sections. Section 2 presents the whole human activity recognition system and explained required blocks of video processing. Proposed approach uses a directed graphical model based on propagation nets, a subset of dynamic Bayesian networks approaches, to model the behaviors. Section 3 provides detailed explanations on the novel description of human pose – a combination of NCM and PCM points which are the main topics of this chapter. Section 4 presents the assumptions of the experiments and achieved results for exemplary behavior that is a $Callforhelp$. Section 5 provides conclusions and suggests a number of areas to be pursued as further work.

6.2 Human Activity Recognition System

The behavior of a person can be understood as a set of poses described by characteristic points of the person. A set of characteristic points at a given time defines a pose. A set of poses defined for consecutive time points or a set of time vectors for individual points forms a descriptor.

The set of points to define a pose may have a different configuration for different types of behavior to be detected. In other words, for at least one type of behavior, a set of points is generated having a configuration different than a set of points for another type of behavior. For consecutive frames, the positions of points belonging to the set are traced and form trajectories of points.

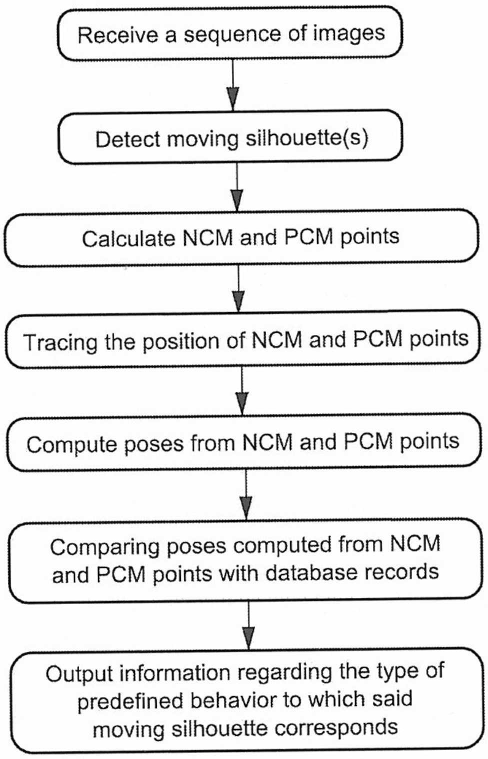

Fig. 6.1 General diagram of human activity recognition system

Figure 6.1 presents in general the human activity recognition system according to the proposed solution. Presented system used single stationary camera.

The main idea of the solution is to track the position of the NCM and PCM points on detected moving silhouette of the human. Position of points in the sets of points on consecutive images are traced in order to generate poses and next, the poses are compared with the predefined descriptors corresponding to the behavior. The comparison is performed by calculating the Euclidean distance for pairs of corresponding points. Each pose shall fit within a predetermined range.

6.3 Negative Curvature Minima and Positive Curvature Maxima Points

There are different ways of describing shapes in images. The invention presented in this chapter is based on the contours of objects in the scene, which are well characterized using the so-called concavity minima or negative curvature minima (NCM) points and convexity maxima or positive curvature maxima (PCM). The NCM points alone may be used, inter alia, to recognize persons in video sequences, as described in [15]. The definition of minimum concavity is as follows: an NCM point is a point of the contour between the points ($P1$, $P2$ in Fig. 6.2) of a convex hull, for which the distance from the segment $|P1P2|$ is largest. The points $P1$ and $P2$ are suitably distant from each other which will be described in details in the subsequent sections of the present description. Fig. 6.1 shows an example of an NCM point.

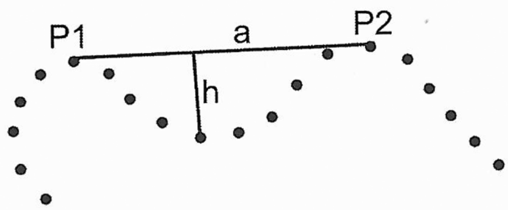

Fig. 6.2 Negative Curvature Minimum found based on hull points $P1$, $P2$

Like the concavity minima, contour convexity may be used to describe the curvature. The method utilizing contour convexity is called positive curvature maxima (PCM) points. This time, the extreme points $P1$ and $P2$ in Fig. 6.2 are selected from the convex hull so that $P1$ is the point closing the $P1$ is the point closing the i-th concavity and $P2$ is a point opening the $i+1$-th concavity. Fig. 6.3 shows an example of a PCM point. Among so selected pair of points, from the contour a PCM point is selected so that the distance from the segment $|P1P2|$ is as high as possible.

6.3.1 Points Selection

This method is based on NCM points and PCM points and has been depicted in Fig. 6.4. Because of such combination it is possible to describe silhouettes with data defining curvature of the hull NCM for negative curvature minima and PCM for positive curvature maxima. This combination contains more information than a standalone NCM point description.

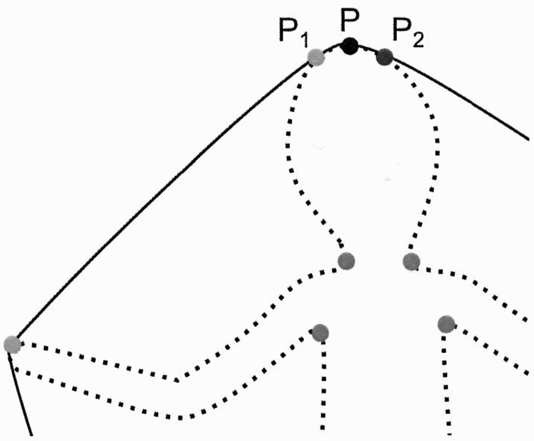

Fig. 6.3 Positive Curvature Maximum P (yellow)

A procedure (Fig. 6.4) determining NCM points starts from step S1 with selecting a pair of consecutive points $\{h_i, h_{(i+1)}\}$ in a vector of a convex hull. If a complete vector of the convex hull has been analyzed, the procedure proceeds from step S2 to step S9. If not, then in step S3 there is determined a length of a segment a between the points $\{h_i, h_{(i+1)}\}$. Next, at step S4, it is verified whether the length a is greater than a threshold γ. In case the length a is greater than the threshold, the procedure proceeds to step S5. Otherwise the procedure returns to step S1 in order to select another pair of points. At step S5, there is selected a point C from the contour vector such that point C is between points h_i and $h_{(i+1)}$ such that its distance h from the segment a is the greatest.

At step S6 it is verified whether an update hull condition is fulfilled, so that point C may be added to the convex hull. The parameters of the condition are as follows: AH_{th} is a concavity depth threshold, $cncArea$ is an area of concavity defined by the section of the hull between h_i and $h_{(i+1)}$ points, $cntArea$ is an inner area of the currently analyzed hull and the $Area_{th}$ is a threshold defining minimum ratio of concavity area to the inner area of the currently analyzed hull.

If the condition is fulfilled the procedure moves to step S7 where the update hull algorithm is executed and then moves to step S8 where point C is added to the NCM output vector.

Lastly, at step S9, the number of iterations is being checked as an end condition. In case it has not reached a required count, the process returns to step S1 and selects another pair of points from the vector of convex hull. The process is repeated from the beginning. The aforementioned update hull method utilizes an implementation of a known algorithm called "Gift wrapping" or "Jarvis march". Its task is to include in the convex hull a previously selected NCM point so that its definition is maintained. During execution of this method there is added a minimum number of contour points to the hull such that the hull vector maintains its continuity.

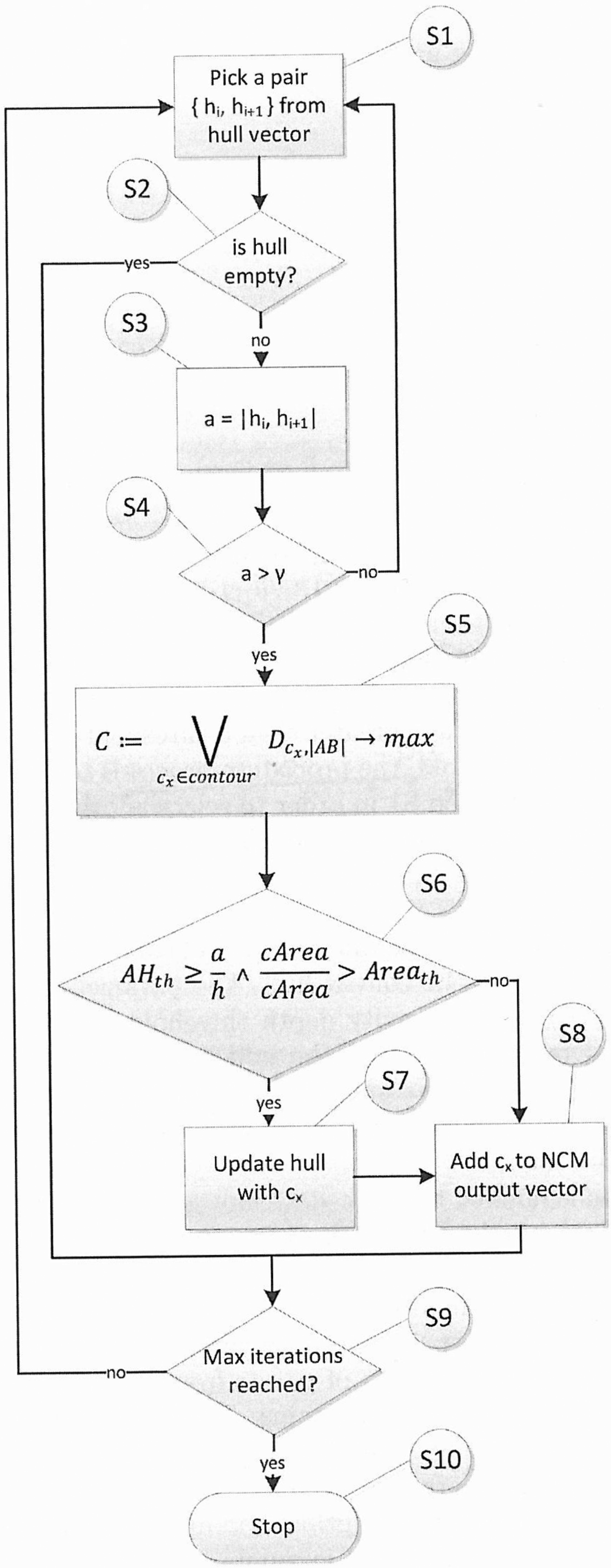

Fig. 6.4 NCM point selection algorithm

The other part of the fourth method relates to the PCM points that are determined similarly to the NCM points. The method may also be applied to pairs of points of a convex hull and is executed as follows:

1. First there are selected pairs of points $\{left_i, right_i\}$ until a pair fulfilling the condition for the length of a segment $D_{(|left_i right_i|)} > \gamma$ is obtained, thereby obtaining $\{left_0, right_0\}$ pair shown in Fig. 6.5
2. The second step is to move the $\{left_i, right_i\}$ until a next pair is found that fulfills the NCM condition thereby arriving at $\{left_1, right_1\}$ shown in Fig. 6.4 variant (b). The index of $left_1$ in the hull vector is stored as $idxRight$.
3. The third step of the procedure is to select a point K_0 from the convex hull between $idxLeft$ and $idxRight$, for which the distance h_0 from the segment $|right_0 left_1|$ is the greatest.
4. Lastly, as the fourth step set the $idxLeft = idx(right_1)$ and continue from the second step.

The subsequent K_i points are computed in an analogous way by maximizing their corresponding distances h_i from the segments $|right_i left_{(i+1)}|$. The process executes its last iteration when $left_n = left_0$. The vector of calculated points is added to previously determined NCM points thereby creating a silhouette descriptor.

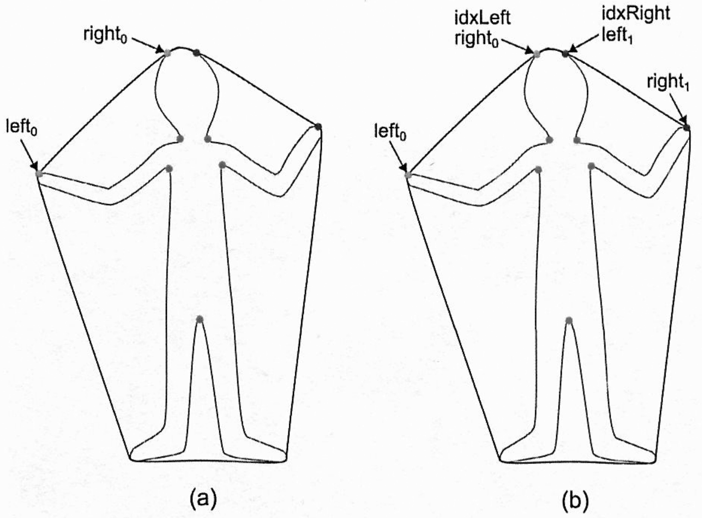

Fig. 6.5 PCM point selection example

6.4 Experimental Results

In our experiments all sequences were divided with respect to the subjects into a training set (9 persons), a validation set (8 persons) and a test set (9 persons)[1].

[1] The test sequence used in the experiments is available at
`http://www.multimedia.edu.pl/missi2014-human-behavior`

The classifiers were trained on the training set while the validation set was used to optimize the parameters of each method (γ parameter mentioned in Section 3). The recognition results were obtained on the test set. This chapter contains results concerning $Callforhelp$ behavior.

Efficiency of recognition was analyzed and for evaluating classification algorithms the *precision* and *recall* performance metrics were used. The metrics are defined as follows:

$$precision = \frac{TP}{TP+FP} \quad (1)$$
$$recall = \frac{TP}{TP+FN} \quad (2)$$

where TP is the set of true positives, FP is the set of false positives and FN is the set of false negatives.

The effectiveness of the proposed approach was validated using four descriptors that consisted of varied number of states (10, 18, 24 and 36 states). Each of them was checked for effectiveness depending on the varied threshold.

The best results were obtained for 36 states descriptor that reached over 90% recall ratio with nearly 80% precision while second promising descriptor achieved a result of roughly 90% recall ratio at the cost of 60% precision. Those experiments show that more complicated descriptors require higher threshold of acceptance to achieve good results. Additionally, it has been observed that the points distribution over human silhouette is important. Some configurations prove better results than others. The recognition result is presented on Fig. 6.6 and Fig. 6.7.

Fig. 6.6 Visual evaluation results

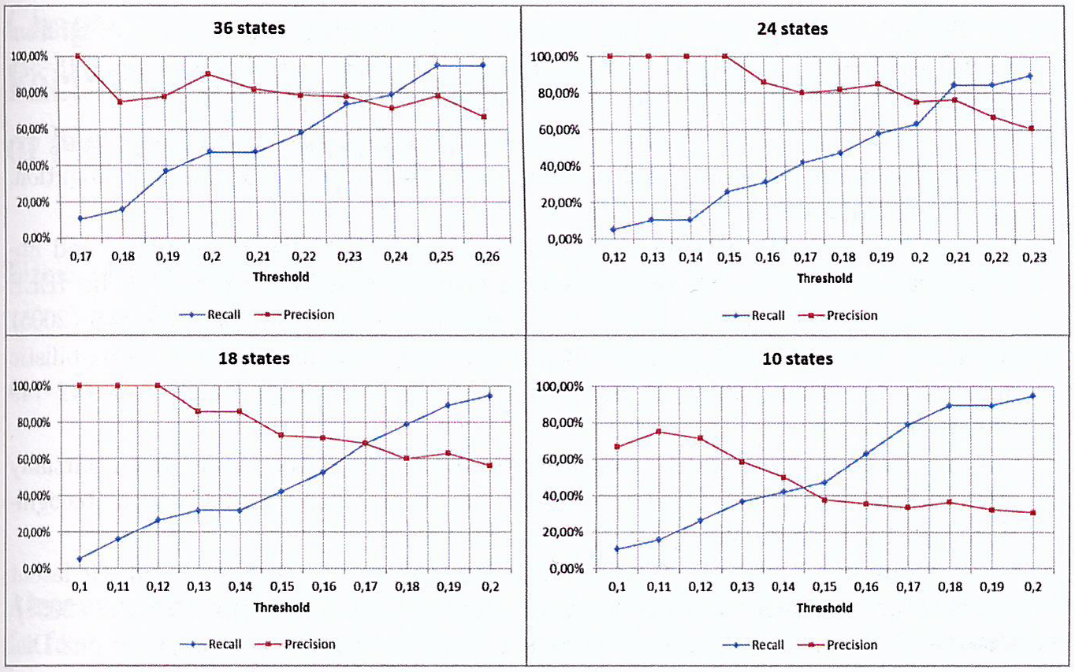

Fig. 6.7 Human activity recognition system evaluation results

6.5 Conclusions

In this chapter, a system for human activity recognition from monocular video is proposed. This system uses a novel description of human pose – a combination of NCM and PCM points on human contour. The points are used in classification process of the behaviour. Results prove that proposed solution seems to achieve a high detection efficiency. Moreover, the experiment results show that some work regarding characteristic points distribution and their relation to the specific behaviors is worth further research in the proposed approach.

Acknowledgements. The project *Innovative security system based on image analysis – SmartMonitor prototype construction (original title: Budowa prototypu innowacyjnego systemu bezpieczeństwa opartego o analize obrazu – SmartMonitor)* is the project co-founded by European Union (project number: UDA-POIG.01.04.00-32-008/10-02, value: 9.996.604 PLN, EU contribution: 5.848.560 PLN, realization period: 07.2011-04.2013). European funds for the development of innovative economy (Fundusze Europejskie – dla rozwoju innowacyjnej gospodarki).

References

1. Buxton, H., Gong, S.: Visual surveillance in a dynamic and uncertain world. Artificial Intelligence 78(1-2), 431–459 (1995)

2. Intille, S.S., Bobick, A.F.: A framework for recognizing multi-agent action from visual evidence. In: AAAI Conference on Artificial Intelligence, pp. 518–525 (1999)
3. Du, Y., Chen, F., Xu, W., Li, Y.: Recognizing interaction activities using dynamic Bayesian network. In: International Conference on Pattern Recogntion, pp. 618–621 (2006)
4. Duong, T., Bui, H., Phung, D., Venkatesh, S.: Activity recognition and abnormality detection with the switching hidden semi-Markov model. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 838–845 (2005)
5. Gong, S., Xiang, T.: Recognition of group activities using dynamic probabilistic networks. In: IEEE International Conference on Computer Vision, pp. 742–749 (2003)
6. Shi, Y., Bobick, A., Essa, I.: Learning temporal sequence model from partially labeled data. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1631–1638 (2006)
7. Li, J., Gong, S., Xiang, T.: Global behaviour inference using probabilistic latent semantic analysis. In: British Machine Vision Conference, pp. 193–202 (2008)
8. Kuettel, D., Breitenstein, M.D., Gool, L.V., Ferrari, V.: Whats going on? Discovering spatio-temporal dependencies in dynamic scenes. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1951–1958 (2010)
9. Wang, X., Ma, X., Grimson, W.E.L.: Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. IEEE Transactions on Pattern Analysis and Machine Intelligence 48(2), 31(3), 539–555 (2009)
10. Albanese, M., Chellappa, R., Moscato, V., Picariello, A., Subrahmanian, V.S., Turaga, P., Udrea, O.: A constrained probabilistic Petri net framework for human activity detection in video. IEEE Transactions on Multimedia 10(6), 982–996 (2008)
11. Ivanov, Y.A., Bobick, A.F.: Recognition of visual activities and interactions by stochastic parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 852–872 (2000)
12. Dee, H., Hogg, D.: Detecting inexplicable behaviour. In: British Machine Vision Conference, pp. 477–486 (2004)
13. Turaga, P., Chellappa, R., Subrahmanian, V.S., Udrea, O.: Machine recognition of human activities – a survey. IEEE Transactions on Circuits and Systems for Video Technology 18(11), 1473–1488 (2008)
14. Shi, Y., Bobick, A., Essa, I.: Learning temporal sequence model from partially labeled data. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1631–1638 (2006)
15. Zhao, L.: Dressed Human Modeling, Detection, and Parts Localization. The Robotics Institute, Carnegie Mellon University, Pittsburgh (2001)