Lying Human Activity Recognition Based On Shape Characteristics

Amin Hakimi Rad
Department of Computer Sciences
Tarbiat Modares University
Tehran, Iran
a.hakimi@modares.ac.ir

Nasrollah Moghadam Charkari
Department of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
moghadam@modares.ac.ir

Jalal A. Nasiri
Department of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
j.nasiri@modares.ac.ir

Hamid Reza Shayegh Broojeni
Department of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
shayegh@modares.ac.ir

***Abstract*—This paper proposes a markerless video analytic system for quantifying body parts movement while lying. These movements include: hand, leg, both hand & leg and turning to left or right movements. Combination of pixel intensity and area difference of both segmented and the whole parts of each silhouette compared with the following silhouettes would provide a useful cue for detection of different body parts movement while lying. Extracted feature vectors after applying PCA algorithm for dimension reduction are finally fed to a multiclass support vector machine for precise classification of proposed movements. Unlike most of the existent human action detection systems that only deal with human movements while standing, we have considered movements that a person does while lying, which has a wide range of application in sport and medical science. Reliable recognition rate of experimental results underlies satisfactory performance of our system.**

Keywords-video surveillance; human shape; multiclass support vector machine; human action recognition; feature reduction;

I. INTRODUCTION

Recognizing human action is a key component in many computer vision applications, such as video surveillance, human computer interface, video retrieval and virtual reality. Surveys are, for example, [1], [2] and [3].

Developing a system that can identify any type of human action is a challenging problem. In this paper a reliable method based on multiclass support vector machine has been developed for the recognition of human actions while lying. There are lots of approaches which advocate the use of single feature for human action classification. Some works are, for example, [4] and [5].

As it is obvious that single feature is insufficient for real action classification, The need for more features has been observed in [6] and [7]. We have used intensity and area difference shape features which are extracted from each silhouette and it's segmentation in different ways.

Human motion recognition works can be classified into model-based methods, and appearance-based methods [8].

Since a model-based method need a lot amount of computation cost for pose estimation, we have used an appearance-based method which attempts to recognize human motions with no prior model action [9] and [10]. The appearance-based features have been extracted from the segmented foreground part of each video frame.

Since it is more difficult to monitor a person's different body parts movement while lying we need to have a good camera place management while installing. Our expriments show that a camera which is installed on the wall near the ceiling is a good choice to capture the human's body parts movement while lying.

II. RELATED WORKS

Considerable works have been done in the field of human action detection. In [11], [12], [13] and [14] an action detection system has been proposed to detect human basic actions such as: walking, bending, hand waving and boxing. In [15], [16], [17] and [18] human action detection systems have been proposed to detect falling down events. As it is obvious most human action detection systems have been proposed to detect human actions while standing. In this paper we propose a human action detection system with a reliable method to detect human different actions while lying.

III. PROPOSED METHOD

This paper proposes a novel method for distinguishing human different actions while lying using a single camera. Our approach exploits computer vision techniques to detect people inside a single room. After applying a static key-frame selection method, we do a background subtraction process and extract a binary silhouette on each selected video frame. After the silhouettes are acquired, the next step involves extracting features and tracking their pattern over time. Due to the high dependence of classification results on efficient image features, this procedure is a crucial factor in the system performance.

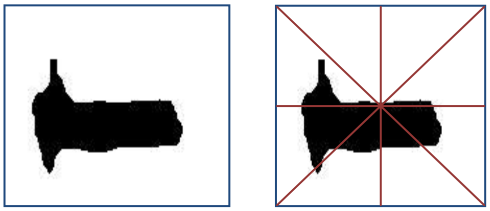

Figure 1. An example of a silhouette and it's segmentation by the proposed method: (a) sample silhouette, (b) segmented sample silhouette.

Since different body parts movement produce a different shape representation, we have analyzed the shape changes of extracted silhouettes in the video sequence. It appears that combination of pixel intensity and area difference shape features of each silhouette compared with the following silhouettes would provide a useful cue for detecting different body parts movement while lying. To increase the accuracy of the classification results using the proposed extracted features, we have also done a segmentation process on each selected silhouette in such a way that we have devided each silhouette into eight segments and then extracted the proposed shape features from each segment of every silhouette separately. We have then compared the extracted features of each segment with the similar segments of the consecutive frames. Fig. 1 shows a silhouette and it's segmentation with the proposed method. The extracted features will then be fed to a multiclass support vector machine for precise classification of movements.

A. *Proposed System Overview*

Overview of the system is shown in Fig. 2.

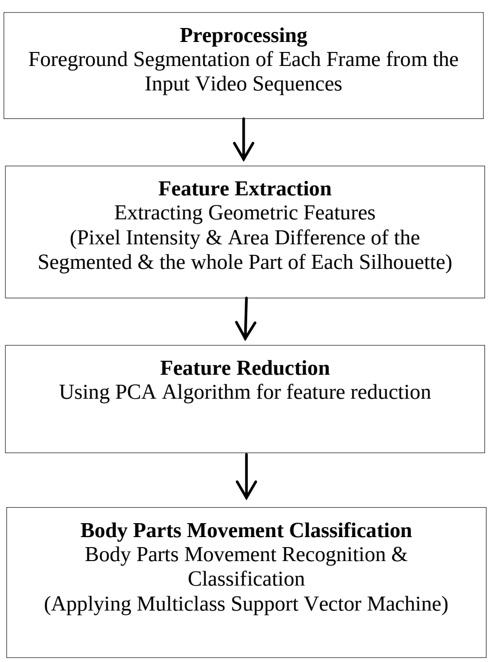

Figure 2. Proposed system flow.

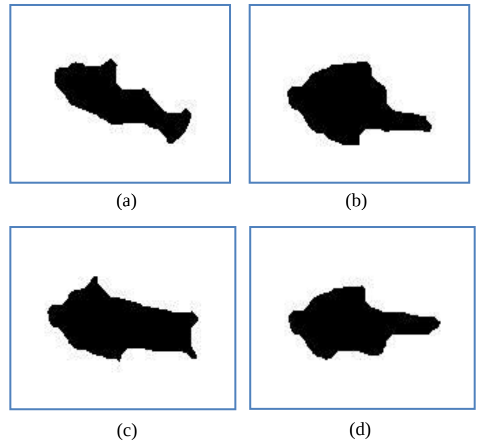

Figure 3. Sample extracted silhouettes of the proposed body parts movements: (a) turning left, (b) left leg movement, (c) right hand movement, (d) simultaneous movement of hand and leg: right leg and right hand.

B. *Foreground Segmentation and keyframe selection*

We have used the background subtraction method which is a popular way for motion segmentation and foreground subtraction by differencing between current image and a reference background image in a pixel by pixel fashion. The method which has been used for background subtraction is a fairly robust one which gives appropriate results on image sequences. Fig. 3 shows some extracted silhouettes of the four proposed body parts movements.

We have also applied a static key-frame selection method which selects one frame of every K arbitrary consecutive frame for feature extraction. An implementation of our work shows that setting $K = 5$ will provide a good implementation result.

C. *Feature Extraction*

Feature extraction is a main step in any recognition system. We have used geometrical features extracted from human shapes to form the final feature vector. So we have analyzed the shape changes of the selected silhouettes of the video sequence. Our expriments show that human shape is a good feature In order to capture different body parts movement while lying. In this research we have extracted two types of geometrical shape features to create the final feature vector.

1) Pixel intensity difference: We have extracted the number of pixel value differences of the same place from the consecutive frames as the pixel intensity difference. When a person moves, his/her different body parts pixel intensities of the silhouette in the consecutive frames will change a lot. To create the part of feature vector which is related to this feature, we have extracted it from the segmented parts of the consecutive frames and also from the whole part of two consecutive frames. For the proposed different body parts it appears that pixel intensity difference is a good feature to extract and evaluate.

2) Area Difference: To extract the area difference feature we calculate the number of pixels which are in the subtracted foreground part of each silhouette. We also extract the defined feature for segmented parts of the silhouettes. The calculated pixel number difference between consecutive frames could be a good feature to create the second part of our final feature vector.

D. Feature Reduction

Although, the extracted feature vector of the motion features can be used directly in body parts movement classification, many studies, in the field of data analysis and feature selection, suggest that not all the features are useful for classification accuracy. In this research we have used the Principal Components Analysis (PCA) technique for feature reduction.

PCA is a statistical method for reducing the dimensions of the data [19]. It selects a set of variables that are uncorrelated with each other and, at the same time, each one is linear combination of the original variables. Principal components are derived from the original data such that the first principal component accounts for the maximum proportion of the variance of the original data set, and subsequent orthogonal components account for the maximum proportion of the remaining variance. The process steps of PCA are as follows:

Step 1) Compute the mean vector of data
Step 2) Compute the covariance matrix of data
Step 3) Compute the eigenvalue and eigenvector matrix of covariance matrix.
Step 4) Form the components using the eigenvectors of the covariance matrix as weighting coefficients.

The exprimental results show that the PCA classifier performs well for our data set.

E. Body Parts Movement Classification

Support vector machines (SVMs) [20] are very popular and powerful in pattern learning because of supporting high dimensional data and at the same time, providing good generalization properties. Moreover, SVMs have many usages in pattern recognition and data mining applications such as text catagorization [21] and [22] phoneme recognition [23], 3D object detection [24], image classification [25], bioinformatics [26] etc. At the beginning, SVM was formulated for two-class (binary) classification problems. The extension of this method to multi-class problem is neither straightforward nor unique. DAG SVM [27] is one of the methods that have been proposed to extend SVM classifier to support multi-class classification.

1) Binary support vector machine formulation: Let $X = \{(x_i, y_i)\}_{i=1}^{n}$ be a set of *n* training samples, where $x_i \in \Re^m$ is an m-dimensional sample in the input space, and $y_i \in \{-1,1\}$ is the class label of sample x_i. SVM finds the optimal separating hyperplane (OSH) with the minimal classification errors. The linear separation hyperplane is in the form of:

$$f(x) = w^T x + b \qquad (1)$$

where w and b are the weight vector and bias, respectively. The optimal hyperplane can be obtained by solving the optimization problem (4), where ξ_i is slack variable for obtaining a soft margin while variable C controls the effect of the slack variables. Separation margin increases by decreasing the value of C .

In a support vector machine, the optimal hyperplane is obtained by maximizing the generalization ability of the SVM. However if the training data are not linearly separable, the obtained classifier may not have high generalization ability, eventhough the hyperplanes are determined optimally. To enhance linear seperability, the original input space is mapped into a high-dimensional dot-product space called the feature space. Now using the nonlinear vector function $\varphi(x) = (\varphi_1(x), \ldots, \varphi_l(x))^T$ that maps the *m*-dimensional input vector x into the *l*-dimensional feature space, the OSH in the feature space is given by:

$$f(x) = w^T \varphi(x) + b, \qquad (2)$$

The decision function for a test data is:

$$D(x) = sign(w^T \varphi(x) + b), \qquad (3)$$

The optimal hyperplane can be found by solving the following quadratic optimization problem:

$$\begin{aligned} &Minimize \ \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i \\ &subject\ to \ \ y_i (w^T \varphi(x_i) + b) \geq 1 - \xi_i \qquad (4) \\ &\xi_i \geq 0, \ \ i = 1, \ldots, n \end{aligned}$$

2) Multiclass support vector machine: As described before, SVMs are intrinsically binary classifiers, but, the classification of body parts movment involve more than two classes. In order to face this issue, a number of multiclass classification strategies can be adopted [28] and [29]. The most popular ones are the one-against-all (OAA) and the one-against-one (OAO) strategies.

The one-aginst-one constructs $n(n-1)/2$ decision functions for all the combinations of class pairs. Exprimental results indicate that the one-agaist-one is more suitable for practical use. We use OAO for body parts movement classification.

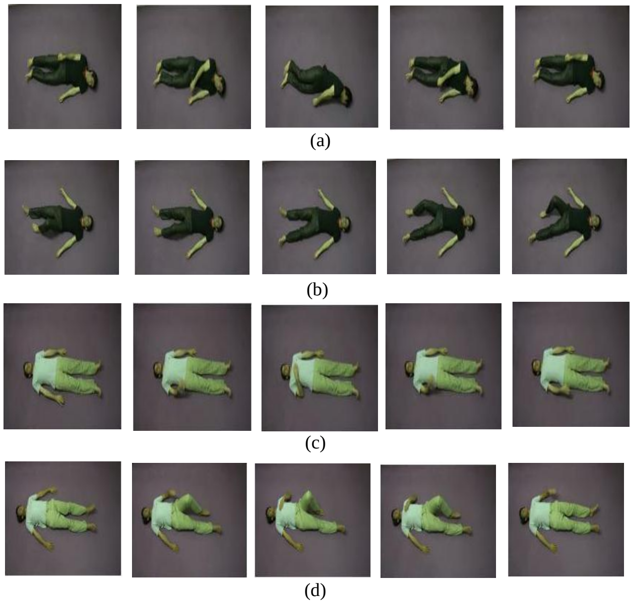

Figure 4. Examples of each body parts movement: (a) turning left, (b) right leg movement, (c) right hand movement, (d) simultaneous movement of hand and leg: left hand and left leg.

IV. EXPRIMENTAL RESULTS AND DISSCUSION

A. *Data Acquisition*

The dataset which we have prepared and used for human action recognition while lying includes four types of actions obtained from three persons with different clothings. The video sequence has been prepared using a fixed place *16 Mega Pixel SONY Digital Camera* while the person is lying toward it. Fig. 4 shows examples of each body parts movement.

B. *Performance Evaluation*

The experimental results show that the system has a robust recognition rate in detecting proposed body parts movements. TABLE I represents the experimental results. N_a, STD *and* R respectively refer to number of actions, standard deviation and the recognition rate.

Also, for better understanding of the wrong classification results we have illustrated the confusion matrix of the classifier output in TABLE II. Notice that M1-M4 are respectively representative of these body parts movements: Turning Left or Right, Leg, Hand, both Hand & Leg movements.

TABLE I. RECOGNITION RATE FOR VARIOUS MOVEMENTS

Movements	N_a	*STD*	*R*
Turning Left or Right	21	0	100.00
Leg Movement	23	0.1800	91.66
Hand Movement	28	0.1610	90.00
Hand & Leg Movement	20	0.2582	80.00

TABLE II. CONFUSION MATRIX

	Classified as			
	M1	M2	M3	M4
M1	21	0	0	0
M2	0	21	1	1
M3	0	1	25	2
M4	0	1	3	16

V. CONCLUSIONS

In this paper a novel efficient approach based on support vector machines for different body parts movement detection is proposed. One of the main advantages of the proposed system in comparison with other human action detection system is to detect human different body parts movement which includes: hand, leg, both hand & leg and turning to left or right whlie lying by extracting efficient shape features. Exprimental results show that feature selection greatly improves the quality of classification. The combination of pixel intensity and area difference shape features, give useful information on the proposed movements. Our expriments indicate that multiclass SVM methods are more suitable for human body parts movement recognition than the other methods because of their capacity to solve an optimization problem in one step.

For future works extracting and combining more amounts of efficient shape features will provide better classification results.

REFERENCES

[1] J. Steffens, E. Elagin, H. Neven, "Person Spotter-fast and robust system for human detection, tracking and recognition", Proc. of IEEE Intl. Conf. on Automatic Face and Gesture Recognition. 1998, pp. 516-521.

[2] M-H. Yang, N. Ahuja, "Recognizing hand gesture using motion trajectories", Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition. 1999, pp. 468-472.

[3] D-S. Jang, H-I. Choi, "Active models for tracking moving objects", Pattern Recognition, 33, (7), 2000, pp. 1135-1146.

[4] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri,Actions as space-time shapes, PAMI,29(12):2247- 2253,2007.

[5] M. D. Rodriguez, J. Ahmed, M. Shah,Action MACH: a spatiotemporal maximum average correlation height filter for action recognition, CVPR, pages 1-8, 2008.

[6] K. Schindler, L. v. Gool, Action snippets: how many frames does human action recognition require?, CVPR, pages 1-8, 2008.

[7] Schuldt, C. Laptev, B. I. Caputo, Recognizing human actions: a local SVM approach. ICPR, pages 32-36, 2004.

[8] Takehito OGATA, JooKooi TAN Seiji ISHIKAWA, "High-Speed Human Motion Recognition Based on a Motion History Image and an Eigenspace", IEICE - Transactions on Information and Systems,Volume E89-D , Issue 1 (January 2006), Pages 281-289.

[9] H. Weiming, T. Tieniu, W. Liang, S Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 34, no. 33, pp. 334-352, Aug. 2004.

[10] E. Ardizzone, A. Chella, and R. Pirrone, "Pose classification using support vector machines," Int. Joint Conf. Neural Networks, Italy, July, 2000, pp. 317-322.

[11] V. Syrris and V. Petridis, "A lattice-based neuro-computing methodology for real-time human action recognition," Information Sciences, vol. 181, pp. 1874-1887, 2011.

[12] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 23, pp. 257-267, 2001.

[13] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," 2005, pp. 65-72.

[14] M. Rodriguez, J. Ahmed, and M. Shah. Action MACH: a spatio-temporal maximum average correlation height filter for action recognition. In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2008.

[15] P. Rajendran, A. Corcoran, B. Kinosian, and M. Alwan, "Falls, fall prevention, and fall detection technologies," Eldercare Technology for Clinical Practitioners, pp. 187-202, 2008.

[16] B. Jansen and R. Deklerck, "Context aware inactivity recognition for visual fall detection," 2006, pp. 1-4.

[17] S. G. Miaou, P. H. Sung, and C. Y. Huang, "A customized human fall detection system using omni-camera images and personal information," 2006, pp. 39-42.

[18] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," 2007, pp. 875-880.

[19] R. O. Duda, P. E. Hart and D. G. Stork, Pattern classification (2nd Edition), Wiley-Interscience, 2000

[20] Vapnik, V., “The Nature of Statistical Learning Theory”, New York: Springer-Verlag, 1995.

[21] Joachims, T., “Text categorization with support vector machines: Learning with many relevant features”, Technical report, University of Dortmund, 1997.

[22] Wang, T.-Y., Chiang, H.-M., “Fuzzy support vector machine for multi-class text categorization”, Information Process and Management, 43, 914–929, 2007.

[23] Salomon, J., “Support vector machines for phoneme classification”, M.Sc Thesis, University of Edinburgh, 2001.

[24] Pontil, M., Verri, A., “Support Vector Machines for 3D Object Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 6, 1998.

[25] Takeuchi, K., Collier, N., “Bio-Medical Entity Extraction using Support Vector Machines”, Proceedings of the ACL 2003 Workshop on Natural Language Processinge in Biomedicine, 57-64, 2003.

[26] Foody, M.G., Mathur, A., “A Relative Evaluation of Multiclass Image Classification by Support Vector Machines”, IEEE Transactions on Geoscience and Remote Sensing, 42, 1335– 1343, 2004.

[27] Platt, J., Cristianini, N., Shawe-Taylor, J., “Large margin DAGs for multiclass classification”, Advances in Neural Information Processing Systems 12. MIT Press, 543–557, 2000.

[28] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machine,” IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[29] C.-W.Hsu and C.-J. Lin, “A comparison of methods formulticlass support vector machines,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.