

Automatic Classification of Human Body Postures Based on the Truncated SVD

N. Zerrouki and A. Houacine

Abstract—In this experimental study, we propose the use of Singular Value Decomposition (SVD) coefficients as features to automatically classify human body postures. The classification process uses images extracted from a fixed camera video. A background subtraction technique is applied for human body segmentation. A truncated SVD is performed by selecting significant magnitude coefficients. And the height-width ratio of the human body is also included in the set of features. The classification is then performed using an Artificial Neural Network (ANN). Four body postures are considered in our experiments, namely: standing, bending, sitting, and lying. Evaluation results show that the proposed method achieved 90.46% classification accuracy. Truncated SVD coefficients and height-width ratio as body posture features are thus appropriate descriptors to achieve high classification accuracy. Also, the proposed method yields the best classification accuracy compared to well-known classification methods.

Index Terms—Human body postures, classification, SVD coefficients, neural network.

I. INTRODUCTION

Human behavior understanding remains an interesting study area in intelligent video monitoring because it is closely related to computer vision applications and automatic surveillance. It is a very challenging and complex task due to some difficulties such as variations in people's appearance, skin color, illumination conditions, and the amount of data generated. To reduce the complexity, many previous works have focused on some specific topics such as face recognition [1], body tracking [2], and hand gesture recognition [3] from a single image or multiple images.

Behavior understanding includes mainly motion pattern identification or body posture analysis and classification. Many approaches have been proposed for human posture classification. Poppe *et al.* [4] use a degree of freedom to estimate the body posture for a presenter in meeting environments. Hooi *et al.* propose a computer vision-based approach to automatically detect human body parts and estimate the human body postures from a monocular video sequence [5].

However, most of the stated methods use skin color to locate hands and head. The person is also supposed to face the camera which can be considered as a limitation.

Capo *et al.* present a computer vision algorithm to automatically model the human body skeleton using Bayesian networks classifier [6]. In [7] the body is decomposed into its

main parts, it constructs a silhouette based body model to determine the location of the body parts [8]. But recognition of human parts is not an easy task as it requires processing based on only visible parts of the body. Lutz Goldmann *et al.* [9] use appearance-based features for posture description such as color, shape, texture, motion, and size. Fast Fourier Transform (FFT) is applied on projection histograms as human body features and artificial neural network (ANN) is used for posture classification by C. Juang and C. Chang [10].

Jun Yin *et al.* [11], present an automatic system for some human abnormal behavior such as shaking, falling down, bending, etc., based on Hidden Markov Model (HMM) for the purpose of behavior recognition. I. Cohen *et al.* [12], consider the use of the shape boundary, position, and mean color to characterize human posture.

Werghi *et al.* [13], use wavelet decomposition for feature extraction and a Bayesian approach as pattern classifier. Human gait recognition based on motion analysis and support vector machines (SVMs) is proposed in [14], [15].

In accordance with aforementioned studies, classification of four human postures including standing, bending, sitting, and lying, is the objective of this work. The classification of the four body postures is based on two-dimensional human body image from a single camera.

In the present paper, new body posture features are proposed. The classification is based on SVD coefficients and height-width body's ratio as features. The SVD coefficients are computed on segmented images. And an ANN is used for posture classification task due to its powerful learning ability.

This paper is structured as follows: Section II describes human body segmentation using background subtraction method. Section III introduces human body feature extraction. Section IV describes ANN classifier for posture classification. In Section V we discuss the experimental results obtained from the analysis of posture classification. Section VI draws the conclusion of this study.

II. HUMAN BODY SEGMENTATION USING BACKGROUND SUBTRACTION

The human body must first be segmented from an image scene for the posture classification problem. In our system we used a background subtraction technique for segmentation. This method uses a single Gaussian distribution for the conditional background probability distribution.

For every pixel in current image, we can determine whether it is a background pixel or not by comparing its probability $P(I|X)$ from the current value with a fixed threshold [16].

We need to save two parameters for our background distribution since single Gaussian distribution only relates to

Manuscript received August 25, 2013; revised October 29, 2013.

The authors are with LCPTS laboratory, University of Sciences and Technology Houari Boumédienne, Algeria (e-mail: nzerrouki@usthb.dz, ahouacine@usthb.dz).

its mean and variance. The Gaussian distribution is given by:

$$P(I[X]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(I[X] - \mu)^2}{2\sigma^2}\right], \quad (1)$$

where μ and σ are the mean and variance. In order to take into account changes in the background image, the background model is adapted progressively, as follows:

$$\mu_{k+1} = \alpha B_k + (1 - \alpha)\mu_k \quad (2)$$

$$\sigma_{k+1}^2 = \alpha(B_k - \mu_k)^2 + (1 - \alpha)\sigma_k^2 \quad (3)$$

where μ_k and σ_k are the mean and variance of the current pixel in the k^{th} image, B_k is the background image used for the k^{th} frame subtraction and α is the update parameter. The range of α is from 0 to 1. Updating the parameter α can correct some errors in the initialized background.

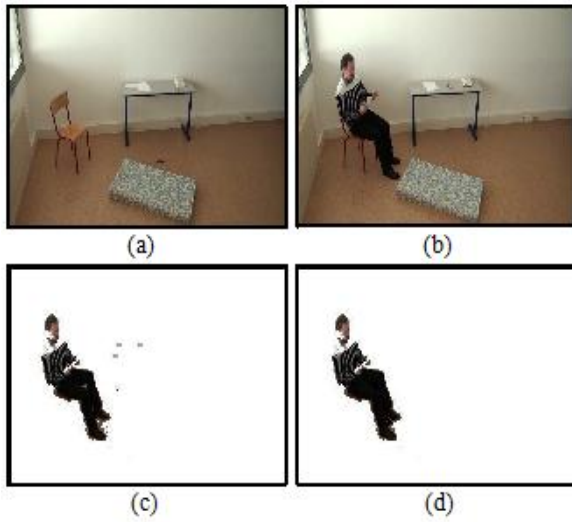


Fig. 1. Results of background subtraction algorithm.

Fig. 1 shows an example of background subtraction technique. In Fig. 1(a) and Fig. 1(b) we illustrate the background and the input image respectively. From the result in Fig. 1(c), some noise regions are observed. To eliminate noise, a morphological operator, which performs the erosion and dilation operator with 3×3 structuring elements, is applied. The segmented image after applying morphological operation is shown in Fig. 1(d).

III. HUMAN BODY FEATURE EXTRACTION

Feature extraction is a crucial task in the classification of human body posture. The extracted features have a direct impact on the recognition results as on the practical performances.

On one hand we would like features to be as descriptive of the posture as possible. On the other hand, we need to involve only limited size feature vectors to take into account the practical exigencies and assure a computational efficiency.

In our method, we propose the use of SVD coefficients as features. The essential concept of the truncated SVD is to use only a small number of singular values to approximate the original image [17]. Given m by n image I , SVD decomposition consists of the factorization of the image I into

three matrices as follows:

$$I = U * S * V^T \quad (4)$$

The matrix S is a diagonal matrix containing the non-negative singular values of the image in decreasing order. There are r singular values, where r is called the rank of image I ; U and V are orthogonal matrices.

We use only the singular values of S as the vector of feature; since the two other matrices (U and V) are rather used for the image reconstruction.

The critical issue consists in reducing the feature space dimension (r values) to allow an efficient training. To address this problem, we represent an image I by its approximation image I' , which corresponds to a truncated rank r' , as:

$$I' = U * S' * V^T, \quad \text{with } r' < r \quad (5)$$

S' is of the same dimensions as matrix S except that it contains only the r' first most significant magnitude singular values (the other singular values are replaced by zeros).

The choice of r' singular values is based on energy information of the image I' , which can be expressed as the sum of squared singular values as.

$$E = S_{11}^2 + S_{22}^2 + \dots + S_{r'r'}^2 \quad (6)$$

Moreover, we can define the energy ratio (ER), which represents the percentage of energy information represented by the truncated decomposition, by the following relation:

$$ER(\%) = \left(\frac{\sum_{i=1}^{r'} S_{ii}^2}{\sum_{i=1}^r S_{ii}^2} \right) \times 100 \quad (7)$$

The ER increases with the number of truncated singular values as represented in Fig. 2. However we can notice that most energy information is confined within the small first number of singular values [17].

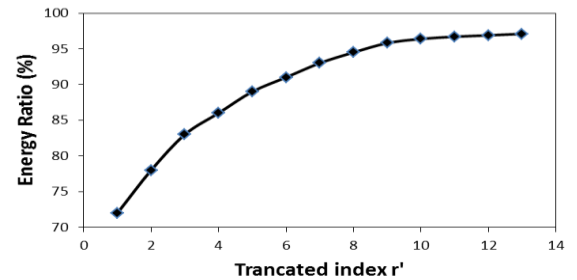


Fig. 2. Relationship of ER with truncated index.

Therefore, the amount of redundant data can be reduced using the truncation method to achieve a better computational efficiency. From our experiments, a value $r' = 10$ is selected.

Another two features are extracted from the segmented image. Which are height and width of the body [18]. These features also appear to reflect significant information for posture classification and their ratio is added to the truncated SVD features to constitute the whole feature vector to be used in the classification process.

IV. POSTURE CLASSIFICATION

After extracting features, an appropriate classifier is needed to classify patterns into one of four posture classes. In

our study we used an ANN classifier. Thus the complete proposed posture identification system consists of a feature extraction stage followed by the classifier, as shown in Fig. 3.

TABLE I: MLP TRAINING PARAMETERS

| | |
|---|--|
| Number of layers | 1 input layer, 1 hidden layer, 1 output layer |
| Number of neurons in different layers | Input : 11 Hidden: between 2 and 25 Output : 4 |
| Initial Weights | Random between -1 and +1 |
| Activation functions for hidden and output layers | Log sigmoid |
| Training parameters for learning | Back propagation |
| Number of iterations | 2560 |
| Acceptable mean square error | 0.001 |

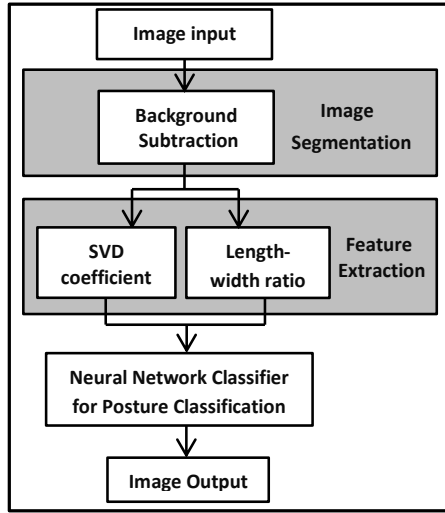


Fig. 3. The proposed system architecture.

The most widely used neural classifier today is the Multi-layer Perceptron (MLP) network with back propagation (BP) learning algorithm. BP is used for the optimization of MLP.

This classifier is a supervised neural network that can use multiple input, output and hidden layers with arbitrary number of neurons [19].

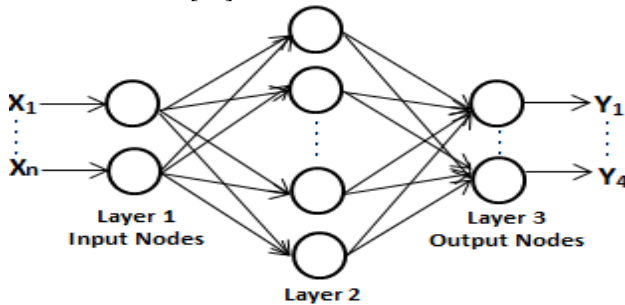


Fig. 4. Schematic diagram of the ANN model.

Here, we have chosen to use a single node for each feature in the input layer (input data). Output layer has four neurons, which gives the patterns of the four posture's classes. This output value should provide the corresponding class for the input data, i.e. we expect to obtain a high output value on the

correct class node and a low output value on all the rest. In the learning phase, a set of training input vectors is presented at the input layers by feature vectors and their corresponding desired output vectors.

Starting with small random weights for each node, the network adjusts the weights attached to the connections so that the difference between the network's output and the desired output for that input vector is decreased.

A single neuron in MLP can be represented as:

$$y_j = f_j \sum W_{ij} \times x_i \quad (8)$$

where, x_i is data input to NN, W_{ij} represents weights between i^{th} neuron of previous layer and j^{th} neuron of the current layer and f_j represents the activation function.

Fig. 4 shows a three-layered MLP network with back-propagation learning (input layer, hidden layer and output layer) used in this study.

To select the optimal MLP architecture, we varied the number of neurons in the hidden layer (from 2 to 25) to select an appropriate learning model. Fifteen (15) neurons are then used in the hidden layer as leading to best results. The MLP training parameters used in this study are given in Table I.

V. EXPERIMENTAL RESULTS

A. Experimental Evaluation

1) Dataset

Our experimental data consist of 4000 images extracted from "fall detection dataset". This dataset contains 191 Video sequences. The frame rate is 25 frames/s and the resolution is 320×240 pixels. The videos are recorded from different environments (home, coffee room, office, and lecture room) and contained variable illumination as well as shadows and reflections that can be detected as moving objects [15]. The actors performed various activities (walking, sitting, crouching down, housekeeping, moving a chair, bending, falling, etc.). The selected images should contain one of the four postures previously defined as shown in Fig. 5.

TABLE II: OVERALL CONFUSION MATRIX

| | Reference data | | | | Total Recognition Rate |
|----------|----------------|---------|---------|-------|------------------------|
| | standing | Sitting | Bending | Lying | |
| standing | 314 | 6 | 5 | 0 | 96.61 % |
| sitting | 0 | 246 | 79 | 0 | 75.69 % |
| Bending | 0 | 34 | 291 | 0 | 89.54 % |
| Lying | 0 | 0 | 0 | 325 | 100 % |

In order to calculate the accuracy of a model, approximately one-third of the data are to be randomly assigned to the test set (1300), with 325 images for each of the four postures. The rest will be used for building the model;

it will be split into training set and validation set. The first one is used to obtain the network parameters (1400) and the second one is used to determine the optimal number of neurons in the hidden layer of the MLP networks (1300).

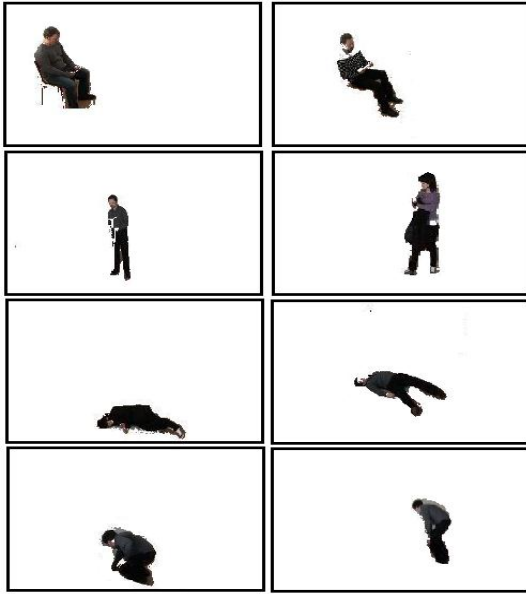


Fig. 5. The four postures of standing, bending, sitting, and lying after segmentation.

The human body is segmented automatically from an input image by the approach described previously; for the three cases, such as training, validation and test data.

2) Evaluation

A confusion matrix is useful when assessing the classification accuracy of one or all categories. The confusion matrix is derived from a comparison of reference postures to the classified postures and is organized as a two dimensional table. This matrix takes the form of the columns representing the reference data by category and rows representing the classification by category.

3) Result

The Table II presents the total confusion matrix of the experiments performed with this setting. One can notice the absence of errors for the lying posture. And confusion mainly occurs between sitting and bending postures which are relatively close.

The Table III summarizes the percentage of correct classification of human body posture, where the total average recognition rate is 90.46%.

The majority of postures are well separated in the classification phase; this confirms the effectiveness in the choice of the parameters of the human posture model and the ability of the classifier to correctly separate postures.

Note that the highest classification accuracy was obtained for the lying and standing postures at (100% and 96.61%), on the other hand, the sitting posture is characterized by the lowest classification accuracy as presented in Table II (75.69%). The level of confusion between sitting and bending postures motivates more investigation to improve the overall classification accuracy by making the approach more robust for these particular postures.

Table III illustrates the comparison of the recognition rate with competing methods.

Zou *et al.* proposed a behavior classification method which classifies three different behaviors from video images [18]. Gkalelis *et al.* realized a model for human movement recognition, using a multi-view video database [20]. The work presented by Girondel *et al.* proposes a recognition system for four different static human body postures in video sequences, acquired with a Sony DFW – VL500 camera [21].

TABLE III: CLASSIFICATION RATES OF HUMAN BODY POSTURE

| Methods | Classifier | Recognition Rate |
|-----------------|-------------------------------|------------------|
| Zou [18] | Bayesian Networks | 76.6 % |
| Girondel [21] | plausibility-based classifier | 88.1 % |
| Gkalelis [20] | LDA | 90 % |
| Proposed method | ANN | 90.46 % |

We can see that the average accuracy rate of the proposed method is slightly better than the results reported in [18], [20] and [21]. Also, the ANN classifier provides significant performance in terms of the effectiveness and the accuracy rate than the methods based on Bayesian networks classifier [18], Linear Discriminant Analysis (LDA) [20], and plausibility-based classifier [21].

B. Results Discussion

By analyzing these results (Table II); we can notice a good accuracy of posture classification (1176 postures over 1300) and the false posture classifications are mainly due to the fact that sitting postures are slightly confused with the bending postures. This confusion relates to the degree of similarity between sitting and bending postures in some positions.

The second problem is tied to the fact that we need more precision in the segmentation step in order to correctly identify more postures. By visually inspecting the final results we observe that most of the missed posture classifications refer to human body parts partially hidden by environment objects. Segmentation errors may also be induced by shadows.

VI. CONCLUSION

A posture classification method is proposed in this paper. The feature vector used consists of truncated SVD coefficients and the height-width body's ratio. These features are invariant under variations of location and size of the person's silhouette. The proposed features are the inputs of a neural network classifier. A high classification accuracy is achieved even in cases where the segmentation method is not perfect or when the postures admit a large variability.

REFERENCES

- [1] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "Two-Dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 131-137, Jan. 2004.
- [2] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 65-81, Jan. 2007.
- [3] M. Panwar, "Hand gesture recognition based on shape parameters," in *Proc. International Conference on Computing, Communication and Applications (ICCCA)*, India, 2012, pp. 1-6.
- [4] R. Poppe, D. Heylen, A. Nijholt, and M. Poel, "Towards real-time body pose estimation for presenters in meeting environments," in *Proc. 13th*

International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2005, pp. 41-44.

- [5] L. S. Hooi, G. Sainarayanan, and L. C. Fan, "Human pose modeling and body tracking from monocular video sequences," in *Proc. IEEE International Conference on Intelligent and Advanced Systems*, 2007, pp. 571-576.
- [6] A. J. Capo, J. Varona, M. G. Hidalgo, R. Mas, and F. J. Perales, "Automatic human body modelling for vision-based motion capture," in *Proc. WSCG Communication Proceedings*, 2006, pp. 259-268.
- [7] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 65-81, Jan. 2007.
- [8] I. Haritaoglu, D. Harwood, and L. S. Davis, "Ghost: A human body part labeling system using silhouettes," in *Proc. 14th International Conference on Pattern Recognition*, Australia, 1998, pp. 16-20.
- [9] L. Goldmann, M. Karaman, and T. Sikora, "Human body posture recognition using mpeg-7 descriptors," in *Proc. Visual Communications and Image Processing*, 2004, pp. 18-22.
- [10] C. Juang and C. Chang, "Human body posture classification by a neural fuzzy network and home care system application," *IEEE Trans. On Systems, Man, And Cybernetics—Part A: Systems And Humans*, vol. 37, pp. 984-994, Nov. 2007.
- [11] J. Yin and Yan Meng, "Abnormal behavior recognition using self-adaptive hidden markov models," in *Proc. Springer ICIAR, LNCS 5627*, 2009, pp. 337-346.
- [12] I. Cohen, H. Li, and M. W. Lee, "Human body posture interface for immersive interaction," in *Proc. International Workshop on Immersive Telepresence (ITP) in conjunction ACM Multimedia*, 2002, pp. 640-644.
- [13] N. Werghi and Y. Xiao, "Recognition of human body posture from a cloud of 3d points using wavelet transform coefficients," in *Proc. of the fifth IEEE International Conference on Automatic face and gesture recognition*, 2002, pp. 70-75.
- [14] H. Su and F. G. Huang, "Human gait recognition based on motion analysis," in *Proc. Int. Conf. Mach. Learn. and Cybern*, 2005, pp. 464-468.
- [15] I. Charfi, J. Mit éan, J. Dubois, M. Atri, and R. Tourki, "Definition and performance evaluation of a robust svm based fall detection solution," in *Proc. 8th International Conference on Signal Image Technology and Internet Based Systems (SITIS)*, 2012, pp. 218-224.
- [16] S. Y. Chien, S. Y. Ma, and L. G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. Circuits and Systems for Video Technique*, vol. 12, pp. 577-586, July 2002.
- [17] S. O. Aase, J. H. Hus øy, and P. Waldemar, "A critique of SVD-based image coding systems," in *Proc. IEEE International Symposium on Circuits and Systems*, 1999, pp. 13-16.
- [18] C. Zou and Z. Liu, "Behavior Classification method based on skeleton model from video images," in *Proc. IEEE international Conference on Computer Science and Information Technology*, 2008, pp. 309-312.
- [19] N. Benoudjit, K. Ferroudji, M. Bahaz, and A. Bouakaz, "In vitro microemboli classification using neural network models and RF signals," *Ultrasonics*, vol. 51, pp. 247-252, Apr. 2011.
- [20] N. Gkalelis, N. Nikolaidis, and I. Pitas, "View independent human movement recognition from multi-view video exploiting a circular invariant posture representation," in *Proc. IEEE international conference on Multimedia and Expo*, 2009, pp. 394-397.
- [21] V. Girondel, L. Bonnaud, A. Caplier, and M. Rombaut, "Belief theory-based classifiers comparison for static human body postures recognition in video," in *Proc. WEC*, 2005, pp. 237-240.



Nabil Zerrouki was born in Algiers, Algeria. He received the engineer degrees and Master in Electrical Engineering from the University of Sciences and Technology Houari Boum édiennne, Algiers, Algeria. His current research interests are computer vision, image processing and their applications to scene analysis and human-machine interaction.



Amrane Houacine was born in Algeria. He is a professor at the Faculty of Electronics and Informatics, University of Sciences and Technology Houari Boum édiennne of Algiers, Algeria. His current research interests include signal and image processing and their applications to human-machine interaction.