

Human Action Recognition Using Independent Component Analysis

Masaki Yamazaki, Yen-Wei Chen and Gang Xu

Department of Media Technology
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577, Japan

Abstract

Principal Component Analysis (PCA) is often used for reducing the dimensionality of input feature space. However, the eigenspace based on PCA is not always the best feature space for pattern recognition. In this paper, we use the feature space based on Independent Component Analysis (ICA) and show that the ICA representation is more effective than the PCA representation for human action recognition. The experimental results on the human action database show that the ICA approach produces more accurate recognition than the PCA approach.

1. Introduction

Recognition of human actions from video images is very important in various ways, such as surveillance, video retrieval and human computer interaction. The general method for human action recognition is to extract human motion information directly from video sequences and to compare it with a human action database. The important problem for human action recognition is how to learn and classify human actions efficiently. The approaches for human action recognition from images can be classified into two types. The first type [1, 8, 9] is to use 2D motion information from background subtraction, frame differencing, optical flow segmentation, and these methods are easy to calculate. However, 2D image information changes due to lighting conditions and camera view points. The second type [2, 10] is to use 3D motion information from a marker and marker-less motion capture system. It is difficult to get 3D human motions, because it needs especial devices or complex algorithms.

2D-based methods are popular methods for human action recognition. There is a famous method that uses an eigenspace method for learning [3]. The eigenspace method regards one image as one point in eigenspace,

so image sequences are point sequences of time series data. It is possible to recognize human actions by verifying the shape of locus constructed from the point sequences in eigenspace. Several successive methods using the eigenspace method for recognizing human motions have been proposed [4, 5, 15, 16, 17, 18]. The eigenspace transformation based on Principal Component Analysis (PCA) is applied to video images for reducing the dimensionality of input feature space. Several pattern classification techniques are finally performed in the lower-dimensional eigenspace for recognition. However, the eigenspace based on PCA is not always the best feature space for pattern recognition, because PCA can only reduce the dimensionality of input data. Recently, some applications of Independent Component Analysis (ICA) [14] that is a statistical data analysis method have been exploited in the field of image processing and computer vision. For example, face recognition, the blind deconvolution of a blurred image and separating reflections. The ICA generalizes the technique of PCA and has proven to be superior to PCA for face recognition [6, 7].

In this paper, we use the feature space based on ICA and show that the ICA representation is more effective than the PCA representation for recognizing human actions. We give a comparison result between ICA and PCA.

Section 2 describes the related work. Section 3 presents the ICA representation of human actions. Section 4 describes our experimental results. Section 5 summarizes our conclusion.

2. Related Work

Human action recognition consists of three major steps, human detection and tracking, feature extraction, and training or classification. The first step is to detect and track a human figure in an image sequence. Background subtraction is a simple method for human de-

tection and tracking. For each image sequence, a background subtraction algorithm and a simple correspondence procedure are firstly used for segmenting and tracking the moving silhouettes of a human figure. The second step is to extract human motion features from the each frame and to normalize the image size with the shape centroid. The third step is to apply PCA for reducing the dimensionality of input feature space and to recognize human actions using standard pattern classification techniques in the lower-dimensional eigenspace. The general human actions are named as follows: Walk, Run, Skip, Hand waving and so on. The experimental motion patterns of many researchers are approximately 5 ~ 10 patterns. Human action recognition is a very difficult problem, because human actions have a lot of motion patterns and there are the few open databases of human actions.

In this paper, we try to improve the PCA-based method by using ICA. The overview of the eigenspace method is shown in Fig. 1. Fig. 1 (a) shows the binary silhouettes extracted from a walking motion sequence using background subtraction and normalization. Fig. 1 (b) shows the PCA projected trajectory in the three-dimensional eigenspace. Recognition of human actions is done by measuring similarities between reference patterns and test samples in the lower-dimensional eigenspace. For example, there are several ways such as Spatiotemporal Correlation, Normalized Euclidean Distance and Canonical Angle. However, the eigenspace based on PCA is not the best feature space for human action recognition, because PCA can only reduce the dimensionality of input data and PCA basis vectors represent global features. For better recognizing human actions, we try to construct more effective feature space by using ICA than the feature space based on PCA, because ICA can get independent components and it represent local features efficiently.

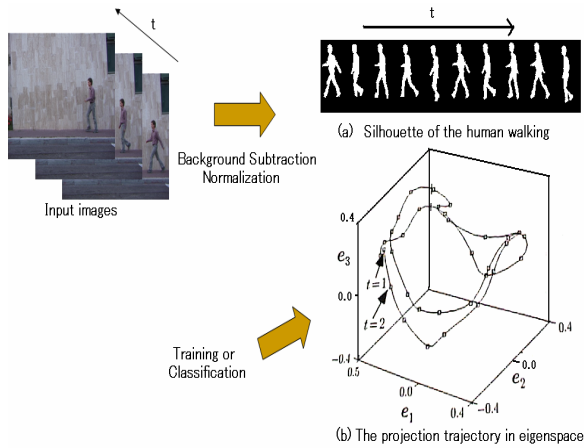


Fig. 1. Overview of the eigenspace method.

3. ICA representation of human actions

3.1. Brief Introduction of ICA

ICA generalizes the technique of PCA and has proven to be a good tool of feature extraction. When some mixtures of probabilistically independent source signals are observed, ICA recovers the original source signals from the observed mixtures without knowing how the sources are mixed. The general model can be described as follows:

We start with the assumption that the observation vectors $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)^T$ can be represented in terms of a linear superposition of unknown independent vectors $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N)^T$.

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

where \mathbf{A} is an unknown mixing matrix ($M \times N$). The goal of ICA is to find a matrix \mathbf{W} , so that the resulting vectors

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \quad (2)$$

recovers the independent vectors \mathbf{S} , probabilistically permuted and rescaled. \mathbf{W} is roughly the inverse matrix of \mathbf{A} .

Before performing ICA, the problem of estimating the matrix \mathbf{A} can be simplified by a prewhitening of the vectors \mathbf{X} . The observed vectors \mathbf{X} is first linearly transformed to other vectors

$$\mathbf{Z} = \mathbf{M}\mathbf{X} \quad (3)$$

whose correlation matrix equals unity: $E(\mathbf{Z} \cdot \mathbf{Z}^T) = \mathbf{I}$. This can be accomplished by PCA with

$$\mathbf{M} = \mathbf{D}^{-1/2}\mathbf{V} \quad (4)$$

where the matrix \mathbf{V} is the eigenvector matrix of the covariance matrix of \mathbf{X} and the matrix \mathbf{D} is the eigenvalue matrix of the covariance matrix of \mathbf{X} . At the same time, the dimensionality of the vectors is reduced. After this transformation we have

$$\mathbf{Z} = \mathbf{M}\mathbf{X} = \mathbf{M}\mathbf{A}\mathbf{S} = \mathbf{B}\mathbf{S} \quad (5)$$

where the matrix \mathbf{B} is the mixing matrix. ICA is performed on the sphered vectors \mathbf{Z} and the estimated mixing matrix \mathbf{B} is an orthogonal matrix,

since $E(\mathbf{Z} \cdot \mathbf{Z}^T) = \mathbf{B}E(\mathbf{S} \cdot \mathbf{S}^T)\mathbf{B}^T = \mathbf{B}\mathbf{B}^T = \mathbf{I}$.

After a prewhitening of the vectors \mathbf{X} , we can rewrite Eq.(2) to:

$$\mathbf{Y} = \mathbf{W}\mathbf{Z} \quad (6)$$

Several ICA algorithms have been proposed for solving \mathbf{W} . Here we use a neural learning algorithm proposed by Bell & Sejnowski [11]. The algorithm is to maximize the joint entropy by using a stochastic gradient ascent. The gradient update rule for the weight matrix \mathbf{W} is as follows:

$$\Delta \mathbf{W} = (\mathbf{I} + g(\mathbf{Y})\mathbf{Y}^T)\mathbf{W} \quad (7)$$

where $g(\mathbf{Y}) = 1 - 2/(1 + e^{-\mathbf{Y}})$ is calculated for the each component of \mathbf{Y} .

3.2. ICA basis images from human actions

From face recognition research reports, PCA basis images represent global information such as average faces and ICA basis images represent local information such as facial parts. The ICA representation is better performance than the PCA representation for face recognition. We show the ICA basis images and the PCA basis images from human walking sequences in Fig. 2 and Fig. 4. The 4 basis functions of the each algorithm are shown in Fig. 3 and Fig. 5. Fig. 3 (a) and Fig. 5 (a) are the 4 PCA eigenvectors with highest eigenvalues (left is first eigenvalue) and Fig. 3 (b) and Fig. 5 (b) are the 4 ICA basis images. From these basis images, we can see that the PCA basis images represent the frequency of body movements, which means that both arms and feet are the most frequent movement in a walking motion. In contrast, the ICA basis images represent the motion features of the each frame sequence, which means that a walking motion are repetitions such as opened feet and crossed feet at each time. In brief, ICA can get the local motion features of human motions.

3.3. Justification of ICA

ICA can be applied only if the independent components obey “non-gaussian” distribution. In order to measure “non-gaussianity” of the resulting independent components, we calculate the kurtosis of the independent components. Since the kurtosis of gaussian distribution is equal to 0, we can measure “non-gaussianity” by calculating the kurtosis. The normalized kurtosis [14] is defined as

$$\frac{\sum_i (x_i - \bar{x})^4}{\left(\sum_i (x_i - \bar{x})^2\right)^2} - 3 \quad (8)$$

The kurtosis of our resulting independent components is shown in Table 1.

TABLE 1. Kurtosis of ICA representation.

	Kurtosis
ICA	69.5

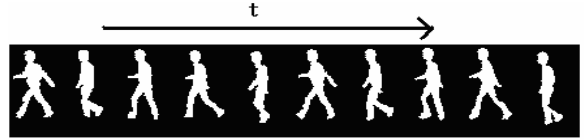


Fig. 2. Sample silhouette images of a human walking sequence.



(a) PCA



(b) ICA

Fig. 3. PCA and ICA basis images from the walking silhouette images.

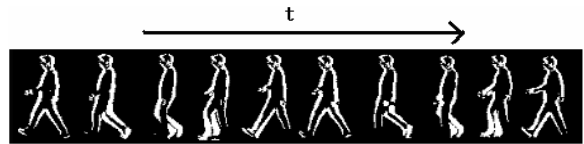


Fig. 4. Sample frame difference images of a human walking sequence.



(a) PCA



(b) ICA

Fig. 5. PCA and ICA basis images from the walking frame difference images.

4. Experimental results

To evaluate our proposed method, we used the human action database [12] containing the six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in the four different scenarios: outdoors s_1 , outdoors with scale variation s_2 , outdoors with different cloths s_3 and indoors s_4 . In our experiments, we used only the data of s_1 that is the easiest situation (see Fig. 6). Since appearance-based methods are depended on the image appearance of a camera angle, it is very difficult to learn all view patterns. We just tried to show the comparison results between PCA and ICA from the simple human action data. Since the database is limited, the recognition results were measured using a leave-one-out strategy which makes maximal use of the available data for training.

In order to select the number of independent components, we used PCA reduction before applying ICA. Recognition was done on the test sequences using a minimum Euclidean distance classifier and a Cosine similarity classifier that uses as similarity the angle between a test vector and a training one. The recognition rate was computed as the ratio of the number of samples classified correctly to the total number samples. Fig. 7 shows the recognition performance that was averaged over the recognition rate of the two motion features (silhouettes, frame difference) and the two classifiers (the minimum Euclidean distance, the Cosine similarity). It can be seen that the recognition rate of ICA was higher than that of PCA. The best recognition rate was about 75% for PCA and about 80% for ICA. Fig. 8 shows the confusion matrix of PCA and ICA. It turns out from these results that the ICA representation provides more discriminability power than the PCA representation for human action recognition.

In order to verify these results and gain some knowledge about the inter-class and intra-class dispersion, we measured the class separability by computing the within-class scatter matrix S_W , the between-class scatter matrix S_B and the mixture scatter matrix $S_M = S_W + S_B$ corresponding to the each class [13]. The criterion used for the class separability measurement was $J = \text{trace}(S_M) / \text{trace}(S_W)$. This number is large when S_B is dispersed or the scatter of S_W is small. Our experimental results are shown in Table 2. The results presented in Table 2 are consistent with the results shown in Fig. 8, since the smallest J are $J(\text{Jog})$ and $J(\text{Run})$ that correspond to the worst recognition

rates (“Jog” and “Run”) and $J(\text{Box})$ has a large value. The value of $J(\text{Jog})$ and $J(\text{Run})$ are small, since running style is very similar to jogging style and it is difficult to classify. As one can see, the ICA representation produces a better class separation than the PCA representation.

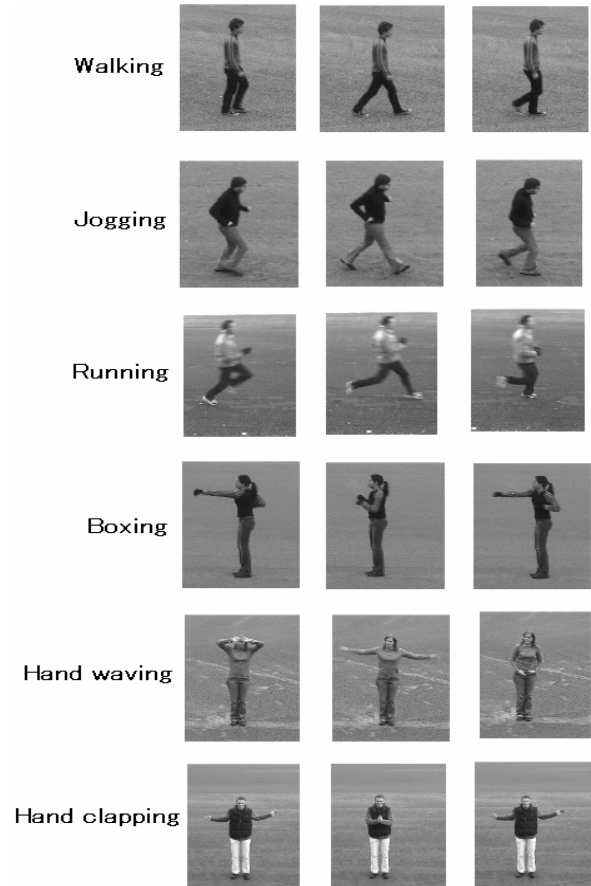


Fig. 6. Sample frames from the action database.

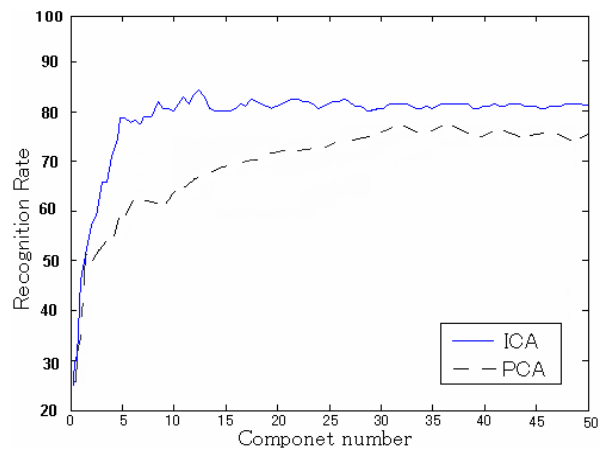


Fig. 7. Recognition performance of PCA and ICA.

PCA	Walk	Jog	Run	Box	Hwav	Hclp
Walk	80	20	0	0	0	0
Jog	20	70	10	0	0	0
Run	5	25	70	0	0	0
Box	0	0	0	85	5	10
Hwav	0	0	0	20	75	5
Hclp	0	0	0	25	5	70

(a) PCA

ICA	Walk	Jog	Run	Box	Hwav	Hclp
Walk	85	15	0	0	0	0
Jog	15	75	10	0	0	0
Run	5	20	75	0	0	0
Box	0	0	0	90	5	5
Hwav	0	0	0	15	80	5
Hclp	0	0	0	20	5	75

(b) ICA

Fig. 8. Confusion matrix of PCA and ICA (%).**TABLE 2. Class separability measurement of PCA and ICA.**

Case	PCA	ICA
J(Walk)	1.025	1.036
J(Jog)	1.014	1.019
J(Run)	1.015	1.021
J(Box)	1.026	1.039
J(Hwav)	1.023	1.031
J(Hclp)	1.017	1.023

5. Conclusions

We have shown that the ICA approach is more effective than the PCA approach for human action recognition. The ICA representation produces a better class separation than the PCA representation, since ICA can get the local motion features of human actions. Further investigations will be tried to compare the other methods with more motion patterns.

6. References

- [1] Y. Yacoob and M. J. Black, "Parameterized modeling and recognition of activities," *Journal of Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 232-247, 1999.
- [2] D. M. Gavrila and L. S. Davis, "3-D model-based tracking of humans in action: a multi-view approach," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 73-80, 1996.
- [3] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: Gait analysis and lip reading," *Pattern Recognition Letters*, vol. 17, pp. 155-162, 1996.
- [4] O. Masoud, and N. Papanikolopoulos, "A Method for Human Action Recognition," *Image and Vision Computing*, vol. 21, no. 8, pp.729-743, Aug. 2003
- [5] L. Wang, T. Tan, H. Ning and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505-1518, 2003.
- [6] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Networks*, pp. 1450-1464, 2002.
- [7] J. Yang, D. Zhang and J. Y. Yang, "Is ICA Significantly Better than PCA for Face Recognition?," *Proc. IEEE Int. Conf. Computer Vision*, pp. 198-203, 2005.
- [8] C. Cedras and M. Shah, "Motion-based recognition: a survey," *Image and Vision Computing*, vol. 13, no. 2, pp. 129-155, 1989.
- [9] D. M. Gavrila, "The visual analysis of human movement: a survey," *Journal of Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82-98, January, 1999.
- [10] L. Campbell and A. Bobick, "Recognition of human body motion using phase space constraints," *Proc. IEEE Int. Conf. Computer Vision*, pp. 624-630, 1995
- [11] A. Bell, and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation* 7:1129-1159, 1995.
- [12] C. Schuldt, I. Laptev and B. Caputo, "Recognizing Human Actions: A Local SVM Approach" *Proc. Int'l Conf. Pattern Recognition*, pp. 32-36, 2004.
- [13] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990.
- [14] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley, New York, 2001.
- [15] M. Black and A. Jepson, "Eigenttracking: Robust matching and tracking of articulated objects using view-based representation," *Int'l Journal of Computer Vision*, vol. 26, pp. 63-84, 1998.
- [16] C. BenAbdelkader, R. Culter, H. Nanda, and L. Davis, "EigenGait: Motion-Based Recognition of People Using Image Self-Similarity," *Proc. Int'l Conf. Audio-and Video-Based Biometric Person Authentication*, pp. 284-294, 2001.
- [17] J. Lu, E. Zhang, Z. Zhang, and Y. Xue, "Gait Recognition Using Independent Component Analysis," *Proc. Advances in Neural Networks*, pp. 183-188, 2005.
- [18] K. K. Lee and Y. Xu, "Modeling Human Actions from Learning," *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, pp. 2787-2792, 2004.