

Human Activity Recognition by Radon Transform

1st Author

1st author's affiliation

1st line of address

2nd line of address

Telephone number, incl. country code

1st author's email address

2nd Author

2nd author's affiliation

1st line of address

2nd line of address

Telephone number, incl. country code

2nd E-mail

3rd Author

3rd author's affiliation

1st line of address

2nd line of address

Telephone number, incl. country code

3rd E-mail

ABSTRACT

A new feature description is used for human behaviour representation and recognition. The feature is based on Radon Transform of extracted silhouettes. Key postures are selected based on the transform feature. Key postures are summed up to represent each sequence. Linear Discriminant Analysis (LDA) is applied to the sum of key postures to obtain low dimensional feature vectors. Different classification methods are used to classify each sequence. Experiments are carried on a public human behaviour database and the result is exciting.

Categories and Subject Descriptors

I.5.3. [Pattern Recognition]: Clustering – Algorithms, similarity measures.

I.4.10. [Image Processing and Computer Vision]: Image representation – Statistical

General Terms

Algorithms, Experimentation, Human Factors.

Keywords

Behaviour recognition, Radon Transform, LDA

1. INTRODUCTION

Human action classification is a fundamental topic in computer vision and is an active research field as well in recent years. It has wide applications such as intelligent surveillance, caring of aged people and so on. Human action classification involves moving object tracking, feature extraction and representation, action model learning and classification. However, it remains a challenging task for computers to achieve robust action recognition.

Some interesting works can be found in action recognition and classification. Song and Perona [1] integrated multiple cues, such as velocities and positions of tracked feature points to model actions with triangulated graphs. Blank *et al.* [2] represented actions by using the shape of silhouettes. Their method relied much on moving object tracking. Efros *et al* [3] used optical flow to characterize motion in very low resolution video sequences. Their method relied on previous aligned video clips of human actions. This work is inspired by Boulgouris [4] who used Radon transform for gait recognition. Radon transforms of silhouettes are used for feature description for this work. Because Radon transform is not very sensitive to noise, this work can tolerant to noise to a certain degree. Radon transform is used to select key

postures in an action sequence. Key postures are selected as the representation of an action sequence because most of human actions contain only a few important postures which are significantly different from each other. Key postures can be used to represent the corresponding action and are sufficient to infer human actions by taking advantages of the contextual constraints imposed by actions[5].

For the purposed of sequence matching, most of researchers use Hidden Markov Models (HMM) [6] or Dynamic Time Warping (DTW) [7]. In this work, the sum of the key postures of a sequence is used for sequence matching. Linear Discriminant Analysis (LDA) [8] is applied to reduce the dimension of sum of key postures. After LDA, different classification methods are used for action recognition.

The remaining sections are organized as follows. Section 2 describes the algorithm on Radon Transform. Section 3 presents the key posture extraction based on Radon Transform. Section 4 illustrates the classification of two sequences. The experiments can be found in Section 5 and the conclusions are made in Section 6.

2. RADON TRANSFORM

Radon transform is well known for its wide range of imaging applications. It has been defined in many different forms and notations. The following is the most common used one is defined by [9]:

$$R_f(\rho, \theta) = \iint f(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy$$

where $R_f(\rho, \theta)$ is the line integral of a 2-D function $f(x, y)$ along a line from $-\infty$ to $+\infty$. $f(x, y)$ is defined as :

$$f(x, y) = \begin{cases} 1 & \text{if } (x, y) \in D \\ 0 & \text{other} \end{cases}$$

where D is the binary silhouette in this work. The position is determined by two parameters ρ and θ . $R_f(\rho, \theta)$ is the integral of f over the line $\rho = x \cos \theta + y \sin \theta$. Radon transform consists in the summation of pixel intensities along lines of different directions. The center point of the silhouette is selected as the reference point.

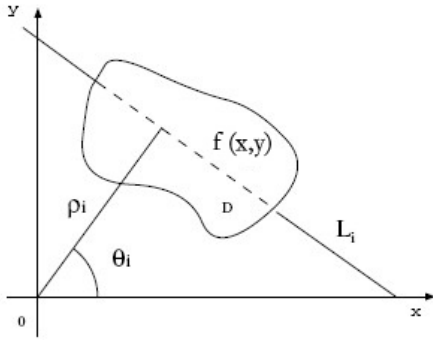


Figure 1 Definition of Radon Transform.

By using radon transform of silhouettes, a map between the coordinate system (x,y) and the radon system (ρ,θ) . Figure 1 shows the definition of the Radon Transform. Given a specific direction, the silhouette is projected to the ρ axis. In other words, pixels along a set of lines parallel to the line which is perpendicular to ρ axis are summer together. Figure 2 shows some silhouettes and their corresponding Radon transform of. The original silhouettes are at the upper line and their corresponding radon transforms are at the bottom line. The original silhouettes are a little different because of noise and shadow. However, their Radon transforms are not much different.

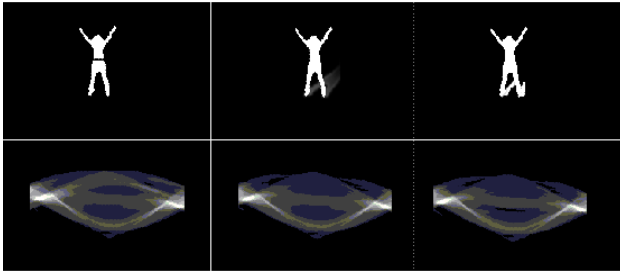


Figure 2 Silhouettes and their corresponding Radon transform.

Radon transform has several useful properties. Some of them are relevant for shape representation [9]. Radon transform taken with respect to center of the silhouette enjoys translation invariance and it is also easy to scale [9]. As for rotation invariant, Radon transform does not guarantee it. But in action recognition, rotation invariant is not required and sometimes rotation invariant is not good for action recognition.

3. KEY POSTURES SELECTION

Key postures are the representation of an action sequence because actions are continuous and key postures are the most significant postures in the action. Once the key postures are selected, the other postures in the sequences can be clustered into one of the key postures.

Clustering is often used in pattern recognition. It is a kind of unsupervised learning. Data with similarities are grouped together. APCLUSTER was proposed by Frey and Dueck [10]. It uses affinity propagation to identify data clusters, using a set of real-valued pair-wise data point similarities as input. Each cluster is represented by a data point called a cluster center, and the method searches for clusters so as to maximize a fitness function called net similarity. The method is iterative and stops after max its iterations.

Since Radon transform cannot guarantee the scale invariant, size normalization is needed before Radon transform. After normalization, the Radon transforms of the action sequence are calculated frame by frame. APCLUSTER is applied to the Radon transform and the output of APCLUSTER is the key postures of each action sequences.

4. ACTION RECOGNITION AND CLASSIFICATION

One of the challenges in action recognition and classification is to compact action information on templates which is suitable for recognition. In this section, how to use key postures to set up a template to achieve action recognition is described.

4.1 Learning Action Classes

Different people perform one action in different style; therefore, there are different action sequences for one action. However, from observation, each action has similar key postures although the key postures are not exactly same. The proposed method uses the sum of all the key postures as feature descriptors. The advantage using this method is that it does not need sequence alignment which is time consuming.

Supposing that there is an action sequence which has n frames and the frames are assigned to $F_1, F_2, F_3, \dots, F_n$. Their corresponding Radon transforms are $R_1, R_2, R_3, \dots, R_n$. The Radon transform $R_j (1 < j < n)$ is a matrix with large dimension.

The key postures selected by Section 3 are named $K_1, K_2, \dots, K_i, (1 < i < n)$. The corresponding Radon transforms of key postures are $R_{K1}, R_{K2}, \dots, R_{Ki}, (1 < i < n)$. For each action, the template is calculated as $T = R_{K1} + R_{K2} + \dots + R_{Ki}, (1 < i < n)$.

Since T is a matrix with very large dimension, LDA is applied to reduce the dimensionality of the template and make recognition more tractable.

LDA provides a methodology for the calculation of a matrix W such that the Fisher's criterion is maximized

$$J(W) = \frac{W \cdot S_B \cdot W^T}{W \cdot S_W \cdot W^T} \quad ()$$

where S_B is the between-class scatter matrix, and S_W is the within-class scatter matrix. While LDA has been widely applied to speech processing, fingerprint and face recognition, it is used in the proposed method in order to achieve dimensionality reduction; therefore, it will simplify the recognition task.

4.2 Classification Proccession

This section is concerned with the classification in the proposed method.

The classifier that used in the proposed method are the Bayesian-based classifier BayesNet, C4.5 or Decision Trees and the Sequential Minimal Optimization (SMO) algorithm[11]. These three classifiers are available as part of WEKA package, a publicly available toolbox for automatic classification[11].

BayesNet enables the use of a Bayesian Network learning using various search algorithms and quality measures. C4.5 is a class for generating a pruned or unpruned C4.5 decision tree. C4.5 is a supervised symbolic classifier based on the notion of entropy since its output - a decision tree - can be easily understood and interpreted by humans. Sequential Minimal Optimization (SMO) is an algorithm for training a support vector classifier.

4.3 System Evaluation

The performance of the proposed method is evaluated by the accuracy which is calculated by the proportion of the action correctly recognized against the total number of actions to be recognized.

Leave-one-out (LOO) cross-validation is used in this research. In LOO, the data set is divided into N subsets. N-1 subsets are used for training, and only the one omitted dataset is used for testing. LOO avoids the possible bias introduced by relying on any one particular division into test and train components. The experiments are conducted and the results are illustrated in Section 5.

5. EXPERIMENTS

The proposed method has been tested on Weizmann Institute of Science's human action database [12]. The database contains 9 subjects performing 10 natural actions such as walking, running, bending, gallop-sideways, one-hand waving, two-hand waving, jumping-forward-on-two-legs and skipping. The resolution of the video is 180 X 144 and video sequence was taken in the speed of 25fps. The average length of a sequence is 50 frames. Human silhouettes are provided by the database. Figure 3 is an example of 'jack' in the database. The quality of these silhouettes is generally good. There are also some defects. For example, some heads are missing in Figure 3.

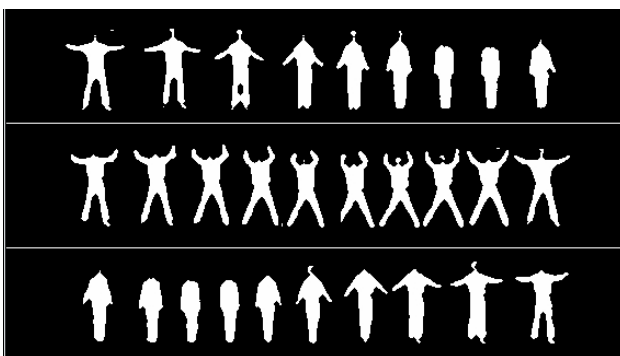


Figure 3 An example of database (Jack)

Since there are 9 subjects in the dataset, the actions performed by 8 subjects are selected for training, and the actions performed by

the left one subjected are used to testing. The experiment is stopped when all of these 9 subjects are used for test.

All of the action sequences will have their corresponding key postures according to Section 3. Figure 4 shows the key postures of running.



Figure 4 Key postures of running.

After that, the learning and testing process will be conducted according to Section 4. Figure 5 shows one hand waving's summation all nine subjects.

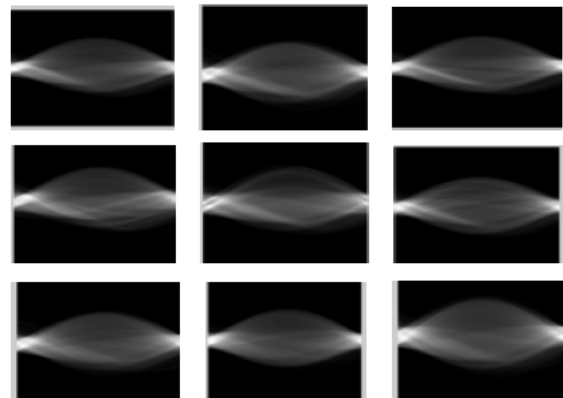


Figure 5 Summation of 'wave1' of nine subjects

The experiments results are shown at Table 1 to Table 3. Table 1 shows the result by using the SMO classifier. Each row represents that there are 9 actions to be classified and their actual classifying results. For example, at row 6, there are 9 run actions to be classified. The result for the running action is that 8 of them are correctly classified as running and one of them is classified as skip. At row 8, there are 9 skip testing samples to be classified. The result is that 7 out of the nine samples are correctly classified as skip, while one of them is classified as wave2 and the other one is classified as side. From Table 1, it can be seen that actions including bend, jack, side, walk and wave achieve are all correctly recognized. The recognition accuracy is calculated as:

$$Rate_{SMO} = \frac{9+9+6+6+8+9+7+9+7+9}{90} = 87.78\%$$

Table 2 shows the results by BayesNet classifier. The general recognition accuracy by BayesNet is 85.56%. The results obtained by C4.5 classifier is illustrated in Table 3. The general recognition rate achieved by C4.5 is 83.33%.

From the results, it can be concluded that the feature to represent human action is relative good. The SMO classifier achieve highest recognition rate at 87.78%.

6. CONCLUSIONS

A new feature representation and recognition system is presented for human behaviour recognition. The system is based on Radon transform of binary silhouettes. For human behaviour recognition and classification, the Radon transform is used for this target. Experiments results are exciting; however, in the future work, more human action databases will be studied. Algorithm optimization will also be studied because Radon transform itself is not fast enough to achieve real time recognition.

Table 1 Recognition Results by SMO classifier

bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2	
9	0	0	0	0	0	0	0	0	0	bend
0	9	0	0	0	0	0	0	0	0	jack
0	0	6	3	0	0	0	0	0	0	jump
0	0	3	6	0	0	0	0	0	0	pjump
0	0	0	0	8	0	1	0	0	0	run
0	0	0	0	0	9	0	0	0	0	side
0	0	0	0	0	1	7	0	0	1	skip
0	0	0	0	0	0	0	9	0	0	walk
0	0	0	1	0	0	0	0	7	1	wave1
0	0	0	0	0	0	0	0	0	9	wave2

Table 2 Recognition Results by ByersNet classifier

bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2	
9	0	0	0	0	0	0	0	0	0	bend
0	8	0	0	0	0	0	0	0	1	jack
0	0	7	1	1	0	0	0	0	0	jump
0	0	3	5	1	0	0	0	0	0	pjump
0	0	0	0	8	0	1	0	0	0	run
0	0	0	0	0	9	0	0	0	0	side
0	0	0	0	0	1	7	0	1	0	skip
0	0	0	0	1	0	0	8	0	0	walk
0	0	0	1	0	0	0	0	7	1	wave1
0	0	0	0	0	0	0	0	0	9	wave2

Table 3 Recognition Results by C.45 classifier

bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2	
8	0	0	0	0	1	0	0	0	0	bend
0	8	0	0	0	0	0	0	0	1	jack
1	0	7	1	0	0	0	0	0	0	jump
0	0	3	5	0	0	1	0	0	0	pjump
0	0	0	0	8	0	1	0	0	0	run
0	0	0	0	0	9	0	0	0	0	side
0	0	0	0	0	1	7	1	0	0	skip
0	0	0	1	0	0	0	8	0	0	walk
0	0	0	1	0	0	0	0	6	9	wave1
0	0	0	0	0	0	0	0	0	9	wave2

7. REFERENCES

- [1] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (Nov. 1993), 795-825. DOI=<http://doi.acm.org/10.1145/161468.161471>.

- [1] Y. Song, L. Goncalves, and P. Perona, "Unsupervised Learning of Human Motion," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2003, pp. 814-827.
- [2] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as Space-Time Shapes," in *The Tenth IEEE International Conference on Computer Vision*, 2005, pp. 1395-1402.
- [3] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in *Ninth IEEE Conference on Computer Vision*, 2003, pp. 726-733.
- [4] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, "Gait recognition: a challenging signal processing technology for biometric identification," *IEEE Signal Process*, vol. 22, pp. 78-90, 2005.
- [5] F. Lv and R. Nevatia, "Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching," in *IEEE Conference on Computer Vision and Pattern Recognition. CVPR '07*, 2007, pp. 1-8.
- [6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications inspeech recognition," in *Proceedings of the IEEE*, 1989, pp. 257-286.
- [7] C. Myers, L. Rabiner, and A. Rosenberg, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing* [see also *IEEE Signal Processing*], vol. 28, pp. 623-635, 1980.
- [8] J. Han and B. Bhanu, "Individual Recognition Using Gait Energy Image," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLEAGENT*, pp. 316-322, 2006.
- [9] S. R. Deans, "The Radon Transform and Some of Its Applications," New York: Wiley Interscience Publications, 1983.
- [10] B. J. Frey and D. Dueck, "Clustering by Passing Messages Between Data Points," *Science*, vol. 315, p. 972, 2007.
- [11] I. H. W. a. E. Frank, *Data mining: Practical machine learning tools with java implementations*,: Morgan Kaufmann, San Francisco, CA, 2000.

- [12] L. G. M. Blank, E. Shechtman, M. A. I. M. Irani, and R. A. B. R. Basri, "Actions as space-time shapes," in IEEE International Conference on Computer Vision. ICCV, 2005, 2005, pp. 1395-1402.

© [2008] IEEE. Reprinted, with permission, from [Chen Yan; Wu Qiang; He Xiangjian, Human Action Recognition by Radon Transform, *IEEE International Conference on Data Mining Workshops, 2008. ICDMW '08*]. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Technology, Sydney's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it