FAST AND RELIABLE RECOGNITION OF HUMAN MOTION FROM MOTION TRAJECTORIES USING WAVELET ANALYSIS

Shu-Fai WONG[1] and Kwan-Yee Kenneth WONG[1]

[1]*Department of Computer Science and Information Systems,*
The University of Hong Kong,
{*sfwong,kykwong*}*@csis.hku.hk*

Abstract

Recognition of human motion provides hints to understand human activities and gives opportunities to the development of new human-computer interface. Recent studies, however, are limited to extracting motion history image and recognizing gesture or locomotion of human body parts. Although the approach employed, i.e. the transformation of the 3D space-time (x-y-t) analysis to the 2D image analysis, is faster than analyzing 3D motion feature, it is less accurate and less robust in nature. In this paper, a fast trajectory-classification algorithm for interpreting movement of human body parts using wavelet analysis is proposed to increase the accuracy and robustness of human motion recognition. By tracking human body in real time, the motion trajectory (x-y-t) can be extracted. The motion trajectory is then broken down into wavelets that form a set of wavelet features. Classification based on the wavelet features can then be done to interpret the human motion. An online hand drawing digit recognition system was built using the proposed algorithm. Experiments show that the proposed algorithm is able to recognize digits from human movement accurately in real time.

1. Introduction

Recognition of human motion is one of the hot topics in computer vision and pattern recognition. It attracts attention from researchers because of its wide applications. If a computer can recognize human motion, it can make inference on the activities involved and can respond accordingly. This makes vision-based human-computer interface possible. Many vision-based applications, like virtual reality game interface, intelligent visual surveillance system, and automatic athletic performance analysis tools, can then be built.

To achieve the goal of building a bridge between computers and human beings, recognition of human motion must fulfill several criteria. To be fast, reliable and robust are the most important ones among the criteria. The recognition process should be fast so that there will be nearly no delay between the communication between human and computers. The reliability of the system should be high such that the chance of misunderstanding should be low. The algorithm used should be robust such that noise and changes in configuration will not affect the performance too much.

Recent research have been trying to fulfill the requirement of being fast by extracting useful motion features and reducing the dimension of such features. A comprehensive survey can be found in [Aggarwal and Cai, 1999], [Cedras and Shah, 1995], [Gavrila, 1999] and [Moeslund and Granum, 2001]. Researchers have tried to project high dimension ($4D$, x-y-z-t) motion features into $2D$ image patterns such that the recognition process can be fast. Motion history image, and discrete gesture analysis are the examples. However, there is often a trade off between efficiency and accuracy. Besides, high accuracy in the specific recognition task does not imply the ease in developing a robust and generic version. For instances, the use of motion history image [Bobick and Davis, 2001], [Davis and Bobick, 1997], [Essa and Pentland, 1995] is fast but is easily affected by noise and small changes in movement; the use of discrete gesture states as motion feature provides fast and accurate recognition results, but is difficult to be applied to continuous motion recognition.

This paper aims at proposing a fast and reliable motion recognition algorithm for recognizing continuous human motion using wavelet analysis. An tracking application have been developed by us using wavelet analysis in [Wong and Wong, 2003] and [Wong and Wong, 2004]. In our previous works, smooth motion trajectory can be captured by the wavelet-based tracker. In this paper, an algorithm is proposed to recognize continuous motion by using such motion trajectory. The experimental results show that the system built on this algorithm can recognize human motion with acceptable accuracy and speed.

2. Overview of the System

To achieve the requirement of being a good interface between human and computer, the motion recognition system should be fast and accurate in performance and can make inference from continuous motion. The proposed system aims at achieving this goal by extracting representative and sparse features and by adopting efficient classification algorithm.

The proposed system consists of three major modules, namely the object tracker, the feature extractor, and the pattern classifier. The object tracker will extract the motion trajectory from the image sequence. The feature extractor will extract motion features from the trajectory captured using wavelet anal-

ysis. During feature extraction, the motion trajectory in 2D is broken down into 2 constituent signals, which are along the x and y directions respectively. Wavelet analysis will be performed on the signal along these directions separately. The motion feature in wavelet form can be used in both tracking and pattern classification. By using the motion features and a wavelet-based estimation method, the tracking process can be facilitated. The motion features will also be directly used by the pattern classifier for motion recognition. Figure 1 summarizes the architecture of the system and the facilitation brought by wavelet-based motion analysis.

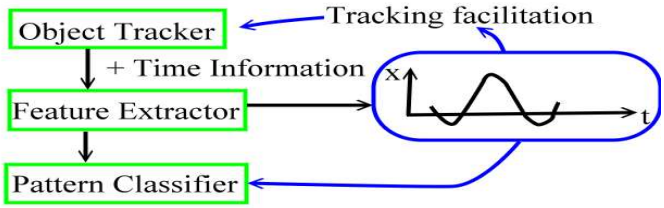


Figure 1. The facilitation in tracking and recognition through wavelet-based trajectory analysis.

3. Wavelet Theory

As described in the previous section, the motion trajectory will be converted to motion features using wavelet decomposition. In the system, the motion trajectories along x and y directions will be converted into wavelet-based representation separately. In this section, the method for decomposing an $1D$ signal into wavelets, the method for extracting representative wavelets and the method for facilitating the tracking will be presented.

Wavelet Analysis

Wavelet analysis was initially used in signal processing and financial forecasting [Chang et al., 1998], [Percival and Walden, 2000]. It is widely adopted in these fields because of its accuracy and efficiency in handling curve modeling and smoothening problem [Daubechies, 1992]. In these applications, observations, e.g. electrical signal, stock price, are usually noisy. In order to make reasonable prediction, the noise should be removed first. Wavelet analysis can decompose the input signal into wavelets for analysis. By removing those wavelets which represent noise, the output curve will be smoothened. Further analysis and prediction can be made on the smoothened signal.

Wavelet analysis consists of two steps, namely decomposition and reconstruction. In wavelet decomposition, the input signal is broken down into small components, called wavelets. The wavelets have internal parameters, such as

scale and transition. The mother wavelet is represented by a certain function, e.g. Harr function. The wavelets are differentiated from each other in terms of their scale and transition, but not the fundamental shape. By breaking down a signal into wavelets, a spectrum is formed. The spectrum is in two dimensions, the scale and transition. Every signal has its own spectrum. Analysis can be performed on this spectrum, e.g. finding out the major components of the signal. Given a wavelet spectrum, signal can be reconstructed from it. By performing wave superposition operation on the wavelets according to the wavelet coefficient (i.e. the intensity in wavelet spectrum), original signal can be reconstructed. The formula for wavelet decomposition and reconstruction is given by:

$$\phi_{j,k}(t) = 2^{-\frac{j}{2}}\phi(2^{-j}t - k), \psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t - k) \quad (1)$$

$$f(t) = \sum_{k \in Z} a_k^J \phi_{J,k}(t) + \sum_{j \leq J} \sum_{k \in Z} d_k^j \psi_{j,k}(t) \quad (2)$$

where $f(t)$ is the input signal at time t, ϕ is the scaling function, ψ is the mother wavelet, 2^{-j} and k represent the scaling and transition factor, a_k^J and d_k^j are the scaling and wavelet coefficients respectively.

The raw signal may contain noise and unnecessary details, wavelets decomposed from the raw signal may not be all representative. To obtain the best set of wavelets as motion feature, the signal reconstructed from such set of wavelets should be smooth. Thus, wavelet denoising is performed before the formation of wavelet-based feature vector.

Wavelet Denoising

Denoising can be done by wavelet decomposition and selective reconstruction. By decomposing the input trajectory (the input signal) into the constituent wavelets, the major wavelets can be identified from the wavelet spectrum. The mathematical details of wavelet decomposition and selective reconstruction can be found in [Donoho, 1994] and [Mallat, 1989]. Re-combination of these major wavelets forms the smoothened trajectory. The minor wavelets represent the noise and are removed.

After wavelet denosing, the major wavelets will be selected as motion features. Such major wavelets are treated as representative features and can be used in object tracking through wavelet estimation.

Wavelet Estimation

As described previously, object tracking can be facilitated by wavelet estimation. The location of the searching window can be predicted by using the information of the motion trajectory. Such prediction can be done by wavelets

superposition. Combining the constituent wavelets near to the current time frame and calculating the value of superposition will give the estimated location in the next time frame.

The estimation can be formulated as:

$$Est(t+1) = \sum_{i=1}^{N} w_i \psi_i(s_i, k_i, t+1) + w_0 Mean \tag{3}$$

where $Est(t)$ is the estimation made at time $t+1$, N is the number of major wavelets, ψ_i are the major wavelets, and w_i, s_i, k_i are the corresponding wavelet coefficient, scaling, and transition parameters respectively.

It represents the value of the superposition of the major wavelets at the prediction time frame. The result is based on the trend of the smoothened input trajectory.

During tracking, the object have to be located in the image from frame to frame. If the trend of the movement, and hence the location of the object can be predicted, searching can be performed within a small window instead of the whole image. By performing wavelet estimation and knowing the trend of the object movement, object tracking can be facilitated.

4. Blob Tracking

In order to capture the trajectory of the moving object, the object has to be tracked. Under the constraint of time complexity, fast and reliable tracking approach have to be adopted. In the proposed system, blob tracking with the use of wavelet estimation is used. Similar tracking system have been developed by us in [Wong and Wong, 2003].

Blob extraction

The blob extractor can extract heuristic features from an image. In the proposed system, optical field and color model are used as heuristic features.

Since the human body is moving in the scene, there should be intensity change (optical flow) within the moving region from image to image. Due to the variation of lighting, especially under fluorescent light, reference white adjustment have to be performed to reduce this variation. The change in intensity is then detected by background subtraction and thresholding. This region is selected as a candidate for further investigation.

Skin color ca be used to identify the human body in the image. Skin color is represented by a skin-color model [Hsu et al., 2001]. Under this scheme, the color intensity of every pixel within the searching region is transformed into the YC_bC_r coordinates. A pixel with color representation closes to the skin color representation in that model will be chosen as a potential region.

By using only one heuristic feature, the candidate region reported may not be accurate. By combining both the optical field and color heuristics, the region with optical flow and skin-color will be selected as the candidate region. The reported region will have a high chance of being a region representing the moving human body. The result is shown in figure 2.

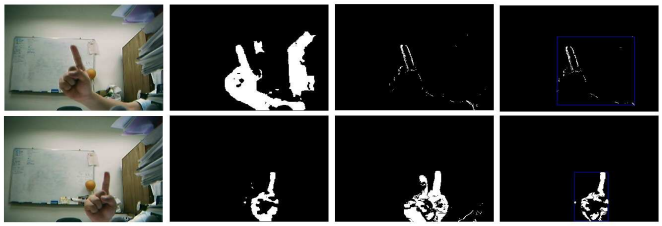


Figure 2. These two rows show the output images of applying different blob selection schemes. The left most images are the input images. The images in the second column are the output of using color detector only. The images in the third column are the output of applying optical flow extraction only. The right most images show the output by using both color and moving region detectors. It shows that using both color and moving region as heuristic features is much more reliable than using any one of them alone.

Wavelet tracking

It is impossible to search through the whole image in order to locate the object because there will be distractors and it is time consuming in doing so. To have fast tracking result, window searching will be used. The location of the searching window will be shifted according to the trajectory estimation. This kind of estimation can be done by using the wavelet-based motion features as described in the previous section.

The wavelet estimator can make reliable estimation based on noisy observations. The region reported by the blob extractor is rough and unreliable. The extractor may report false region, e.g. a box with skin-color and slight change in intensity. To avoid the negative effects due to noisy observation, wavelet estimator will ignore the outlying observations and make prediction based on the smoothened trajectory of the target.

Wavelet estimator performs estimation in two steps: trajectory smoothening and trend prediction. In the trajectory smoothening step, the input trajectory of the observations is smoothened using wavelet decomposition and selective reconstruction. The smoothened trajectory will be used to predict the trend using wavelet trend analysis in the prediction stage.

By performing trajectory smoothening, the outlying observations will be removed. The accuracy of the estimation will not be affected even if the observation is not highly reliable. Having high accuracy in estimating the location of the searching window implies that less effort is needed to locate the object.

The object tracking process can therefore be facilitated with the use of wavelet estimation.

5. Motion Classification

The raw input of the motion classifier is in wavelet format. Those wavelets are the representative constituents of the motion trajectory. However, the raw input should be preprocessed so that the final feature is sparse enough for fast classification.

Feature Dimension Reduction

The set of major wavelets of the motion trajectory will be first organized in spectrum form as described in Section 3. The spectrum of a motion trajectory has two dimensions, i.e. the scale and transition. It is in the form of a table of cells, and each cell stores the corresponding wavelet coefficient. The trajectories along the x and y direction have their own spectrum features. By converting each spectrum from a $2D$ matrix form into a $1D$ vector form and connecting the two vectors from the x and y direction into one, the resultant vector gives the intermediate motion feature vector.

The above motion feature vector is of extremely high dimension, and is thus not suitable to be fitted into the pattern classifier. Dimension reduction will be performed to transform the intermediate feature vector into a low dimension, discriminative and representative feature vector. In the system, principal component analysis (PCA)[Duda et al., 2000] is used for dimension reduction.

Competitive Network

Since the intermediate feature vectors have been transformed into low dimension and discriminative feature vectors, it is possible to use a fast but less reliable classifier. Competitive network was chosen as the pattern classifier in the system. It is simple in structure, fast in performance and accurate enough, provided that the features themselves are discriminative. The corresponding formula is described in formula (4) to (6). The mathematical details can be found in [Banzhaf and Haken, 1990] and [Hagan et al., 1995].

The classification of feature vectors can be done by:

$$\mathbf{a} = compet(\mathbf{W}\mathbf{y}) \tag{4}$$

$$a_i = compet_i(\mathbf{n}) = \begin{cases} 1, & \text{if } i = i^* \\ 0, & \text{if } i \neq i^* \end{cases} \tag{5}$$

where $n_{i^*} \geq n_i$, $\forall i$, and $i^* \leq i$,$\forall n_i = n_{i^*}$, $\mathbf{W}$ is the network weight, $\mathbf{y}$ is the transformed feature vector, $\mathbf{a}$ is the vector for classification result, $\mathbf{n}$ is the result of $\mathbf{W}\mathbf{y}$.

The training equation for competitive network is:

$$\mathbf{w_i}(t) = \mathbf{w_i}(t-1) + \alpha \mathbf{a_i}(\mathbf{t})(\mathbf{y}(t) - \mathbf{w_i}(t-1)) \quad (6)$$

where $\mathbf{W} = \{\mathbf{w_1 w_2}...\mathbf{w_L}\}^T$, L is the dimension of the output vector $\mathbf{a}$, weight coefficient $\mathbf{w_i}(t)$ at different time t is updated, and α is set as the learning rate.

6. Experiments and Results

Experimental Setup

The proposed recognition algorithm was implemented as an vision-based input device that can recognize digit from human motion. Users need to draw the digit in front of the web cam, the system will track the hand movement, determine the trajectory and then recognize the drawing movement. The corresponding digit will be reported after the motion pattern classification.

The proposed system was implemented using Visual C++ under Microsoft Windows. The experiments were done on a P4 2.26 G Hz computer with 512M ram running Microsoft Windows.

Reliability of Wavelet Analysis

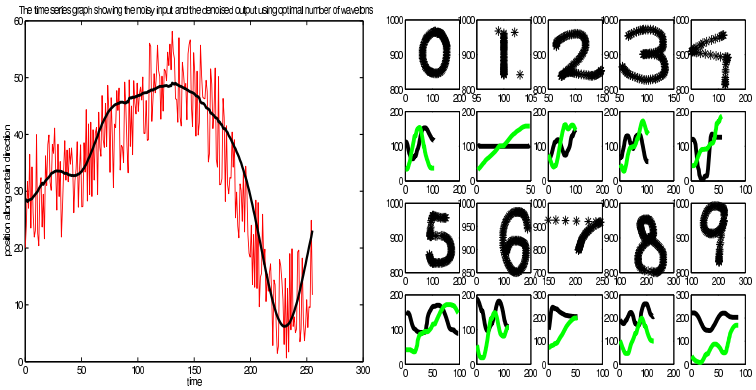
In this experiment, the smoothening performance of the wavelet estimator was tested. The input signal was the y-component of the $2D$ moving path of the hand. The result is shown in figure 3(a). The result shows that the wavelet estimator can handle noisy signal quite well. The smoothened signal is close to the actual movement.

Performance of Wavelet-based Tracking

The hand under clutter background was tracked in this experiment. The result is shown in figure 4. The hand can be tracked in the clutter background consisting of skin color. In addition, when there is locomotion of hand, the trajectory can still be smoothened. The average frame rate that the tracker can process is about 15 frames per second.

Accuracy and Speed of the Recognition System

The digit recognition system was tested for its accuracy and speed. The experiment shows that the accuracy of the recognizer is over 70% (out of 200 test cases) given that the size of training set per digit is 20 and the final feature vector with a dimension of 10. The time to perform classification is less than 1 second. The input digits and the corresponding trajectories along x and y direction is shown in figure 3(b).



(a) (b)

Figure 3. (a) The red fuzzy line represents the input signal (the y-component of the $2D$ moving path of the hand). The black solid line represents the smoothened signal using wavelet analysis. Even the input signal contains noise due to unreliability of blob extraction, the trajectory can be still smoothened using wavelet denoising. (b) The first and third rows show the handwritten digits. The second and fourth rows show the corresponding motion trajectory along the x and y directions. The black (dark) curve is the trajectory along the x direction while the green (light) curve is the trajectory along the y direction. It shows that the motion curves per each digit is discriminative and can be treated as motion feature.

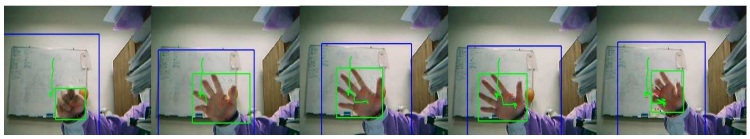


Figure 4. The green (lighter) block is the blob region detected. The blue (darker) block is the maximum bound of the searching window. The red fuzzy dots are the observations. The green continuous line is the predicted trajectory which describes the movement of the blob region. It shows that the hand can be tracked accurately under clutter background. In addition, even when there is locomotion of the hand, the predicted trajectory of the blob was not affected much.

7. Conclusions

Recognition of human motion has a wide range of applications in human-computer interface. However, there is a trade off between accuracy and computational complexity in developing such vision-based interface. Commonly used recognition approach seems to be inappropriate for such application. This paper addresses the issue by introducing a wavelet-based motion trajectory recognition algorithm for interpreting continuous motion sequences. The motion trajectory is first obtained from object tracking, and is transformed to wavelet-based motion features. By performing classification on the preprocessed wavelet-based features, the motion can be classified. The experiments

show that the proposed approach can extract low dimension and representative features, and it can recognize continuous movement of hand reliably and quickly. However, the proposed system cannot handle fast motion at this stage due to the limited speed of the object tracker. A better tracking techniques have to be introduced to boost the performance of the whole system. Besides, pattern classifier have to be improved to handle less discriminative data set such that the size of the dimension can be further reduced.

References

Aggarwal, J. K. and Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding: CVIU*, 73(3):428–440.

Banzhaf, W. and Haken, H. (1990). Learning in a competitive network. *Neural Networks*, 3:423–435.

Bobick, A. F. and Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257–267.

Cedras, C. and Shah, M. (1995). Motion-based recognition: A survey. *IVC*, 13(2):129–155.

Chang, C.S., Fu, W., and Yi, M. (1998). Short term load forecasting using wavelet networks. *Engineering Intelligent Systems for Electrical Engineering and Communications*, 6:217–223.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.

Davis, J. and Bobick, A. (1997). The representation and recognition of action using temporal templates. In *Proceedings Computer Vision and Pattern Recognition (CVPR'97)*, pages 928–934.

Donoho, D.L. (1994). Denoising by soft thresholding. *IEEE Trans. on Inform. Theory*, 41:613–617.

Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification*. John Wiley and Sons, Inc., 2nd edition.

Essa, I. A. and Pentland, A. (1995). Facial expression recognition using a dynamic model and motion energy. In *ICCV*, pages 360–367.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding: CVIU*, 73(1):82–98.

Hagan, M. T., Demuth, H. B., Beale, M. H., and Demuth, B. H. (1995). *Neural Network Design*. Brooks Cole.

Hsu, R.L., Abdel-Mottaleb, M., and Jain, A. (2001). Face detection in color images. In *Proc. IEEE ICIP*, pages 1046–1049.

Mallat, S. (1989). A theory of multiresolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11:674–693.

Moeslund, T. B. and Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding: CVIU*, 81(3):231–268.

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge.

Wong, S.-F. and Wong, K.-Y. K. (2003). Reliable and fast human body tracking under information deficiency. In *Proc. IEEE Intelligent Automation Conference*, pages 491–498, Hong Kong, China.

Wong, S.-F. and Wong, K.-Y. K. (2004). Real time human body tracking using wavenet. In *Proc. Asian Conference on Computer Vision*, pages 91–96, Jeju Island, Korea.