Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 986134, 11 pages
doi:10.1155/2012/986134

Research Article

Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition

Xiaoyan Wang,[1] Ming Xia,[1] Huiwen Cai,[2] Yong Gao,[3] and Carlo Cattani[4]

[1] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[2] *Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[3] *Zhejiang Jieshang Vision Science and Technology Cooperation, Hangzhou 310013, China*
[4] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

Correspondence should be addressed to Xiaoyan Wang, xiaoyanwang@zjut.edu.cn

Received 12 January 2012; Accepted 3 February 2012

Academic Editor: Ming Li

This paper is concerned with the recognition of dynamic hand gestures. A method based on Hidden Markov Models (HMMs) is presented for dynamic gesture trajectory modeling and recognition. Adaboost algorithm is used to detect the user's hand and a contour-based hand tracker is formed combining condensation and partitioned sampling. Cubic B-spline is adopted to approximately fit the trajectory points into a curve. Invariant curve moments as global features and orientation as local features are computed to represent the trajectory of hand gesture. The proposed method can achieve automatic hand gesture online recognition and can successfully reject atypical gestures. The experimental results show that the proposed algorithm can reach better recognition results than the traditional hand recognition method.

1. Introduction

The goal of Human Computer Interaction (HCI) is to bring the performance of human machine interaction similar to human-human interaction [1]. Gestures play an important part in our daily life, and they can help people convey information and express their feelings. Among different body parts, the hand is the most effective, general-purpose interaction tool. Therefore, hand gesture tracking and recognition becomes an active area of research in human computer interaction and digital entertainment industry [2–4]. A gesture can be static or dynamic or both. According to this, there are three types of gesture recognition: static hand posture recognition, dynamic hand gesture recognition, and complicated hand gesture recognition. Our work in this paper concentrates on dynamic gesture recognition, which

characterizes the hand movements. Tracking frameworks have been used to handle dynamic gestures. Isard and Blake [5] established a hand tracking approach based on 2D deformable contour model and Kalman filter [6]. However, it is inefficient to track an articulated object which has a high dimension state space using condensation alone. MacCormick and Blake [7] introduced a partition sampling method to track more than one object. MacCormick and Isard [8] implemented a vision-based articulated hand tracker using this technique after that. Their tracker is able to track position, rotation, and scale of the user's hand while maintaining a pointing gesture. Based on Blake's work, Tosas [9] makes some technique extensions and implements a full articulated hand tracker.

Several methods on hand gesture recognition have been proposed [10–13], which differ from one another in their models, just like Neural Network, Fuzzy Systems and Hidden Markov Models (HMMs) [14]. The most challenging problem of dynamic gesture recognition is its spatial-temporal variability, when the same gesture can differ in velocity, shape, duration, and integrality. These characteristics make it more difficult to recognize dynamic hand gestures than to recognize static ones. HMM is a statistical model widely used in hand writing, speech, and character recognition [13, 15] because of its capability of modeling spatial-temporal time series. HMM has also been successfully used in hand gesture recognition [13, 16–18], in respect that it can preserve the spatial-temporal identity of hand gesture and have an ability to do the segmentation automatically. Motion features of each time point have been modeled in most of the dynamic hand gesture recognition methods using HMM, nevertheless, the whole trajectory shape characters are not considered at the same time. The recognition based on local features is very sensitive to sampling period and velocity, and the continuous local process of gesture will cause false recognition.

Researches on psychology indicate that human brains lean to perceive object from a whole, and then apprehend its details, which illustrates that an object can only be described perfectly when the local and global information are integrated. In this paper, we propose a dynamic gesture trajectory modeling and recognition method based on HMM. Cubic B-spline is adopted to approximately fit the trajectory points into a curve, and invariant curve moments as global features and orientation as local features are computed to represent the trajectory of hand gesture. Threshold model is used to model all the atypical gesture patterns, and automatically segment and recognize the dynamic gesture trajectory. The proposed method can achieve automatic hand gesture online recognition and can successfully reject atypical gestures. Meanwhile, the experiment results show that the recognition performance of the proposed algorithm can be greatly improved by combining the global invariant curve features with local orientation features.

The rest of the paper is organized as follows: Section 2 describes the dynamic gesture representation and the global and local features we used. Section 3 gives the continuous hand gesture recognition procedure, which contains hand detection, tracking, and gesture recognition based HMM. The experimental results are shown in Section 4. Finally, Section 5 and ends the paper with a summary.

2. Dynamic Gesture Representation

A dynamic hand gesture is a spatial-temporal pattern and has four basic features: velocity, shape, location, and orientation. The motion of the hand can be described as a temporal sequence of points with respect to the hand centroid of the person performing the gesture.

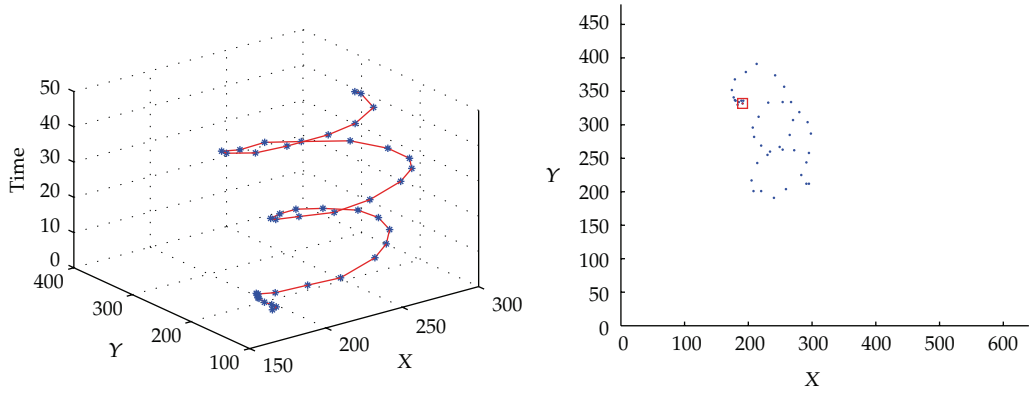


Figure 1: A dynamic hand gesture instance.

In this paper, the hand shape is not considered and each dynamic hand gesture instance is represented by a time series of the hand's location:

$$p_t = (x_t, y_t), \quad (t = 1, 2, \ldots, T), \tag{2.1}$$

where T represents the length of gesture path and varies across different gesture instances. Consequently, a gesture containing an ordered set of points can be regarded as a mapping from time to location. Figure 1 shows a dynamic hand gesture instance and gives its projection along the time axis onto the image plane.

2.1. Local Feature Representation

There is no doubt that selecting good features plays significant role in hand gesture recognition performance. The orientation feature is proved to be the best local representation in terms of accuracy results [19–21] and it is considered as the most important feature in dynamic gesture recognition using HMM [22, 23]. Therefore, we will rely upon it as a main local feature in our system. The orientation of hand movement is computed between two consecutive points of the hand gesture trajectory:

$$\theta_t = \arctan\left(\frac{y_{t+1} - y_t}{x_{t+1} - x_t}\right), \quad (t = 1, 2, \ldots, T). \tag{2.2}$$

A feature vector will be determined by converting the orientation to directional codewords by a vector quantizer. For example, in Figure 2 the orientation is quantized to generate the codewords from 1 to 20 by dividing it by 20 degree. Thereby, the discrete feature vector will be used as an input to discrete HMM.

2.2. Global Feature Presentation

The human brain is inclined to sense object from a whole, and people also try to understand a gesture as integrity. Accordingly, we try to connect all the discrete points of gesture using

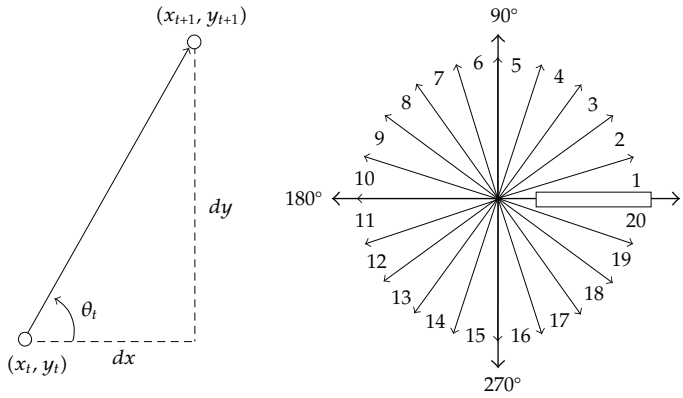


Figure 2: The orientation and its codewords.

a slippery line. Cubic B-spline function is adopted to approximately fit the trajectory points into a curve:

$$p(t) = \sum_{m=0}^{3} B_m(t)\mathrm{CP}_m, \tag{2.3}$$

where $B_0(t) = (1-t)^3$, $B_1(t) = 3t(1-t)^2$, $B_2(t) = 3t^2(1-t)$, $B_3(t) = t^3$, CP_m are control points. After the curve is shaped, an issue to be addressed is the variation of speed of the same gesture. To overcome this problem, all curves are scaled such that they lie within the same range. Those curves for faster moves are relatively expanded by interpolation and those of slower moves are contracted.

The trajectories of a same gesture vary in size and shape. We use invariant curve moments as global features to represent the trajectory [24]. The advantage of moment methods is that they are mathematically concise and invariant to translation, rotation, and scale. Furthermore, they reflect not only the shape but also the density distribution within the curve.

The $(p+q)$th-order moments of plane curve l are defined as

$$m_{pq} = \int x^p y^q \, ds, \quad (p, q = 0, 1, 2, \ldots), \tag{2.4}$$

where ds is the arc differentiation of curve l. The $(p+q)$th-order central moments are defined as:

$$\mu_{pq} = \int (x - \overline{x})^p (y - \overline{y})^q ds, \quad (p, q = 0, 1, 2, \ldots), \tag{2.5}$$

where $\overline{x} = m_{10}/m_{00}$, $\overline{y} = m_{01}/m_{00}$.

For a digital image $f(x,y)$,

$$m_{pq} = \sum_{x,y} x^p y^q f(x,y),$$
$$\mu_{pq} = \sum_{x,y} (x-\overline{x})^p (y-\overline{y})^q f(x,y). \tag{2.6}$$

This paper defines $f(x,y)$ as

$$f(x,y) = \begin{cases} 1, & (x,y) \in l, \\ 0, & (x,y) \notin l. \end{cases} \tag{2.7}$$

Thus, the global descriptors of hand gestures have been calculated using the central moments of the curve. As we use discrete HMM, all the features extracted need to be represented as an integer. The statistical distributions of the central moments are calculated and then a feature is denoted as one or two digits.

3. The Continuous Hand Gesture Recognition Scheme

In this paper, we consider online-continuous-handed dynamic gestures based on discrete HMM. The hand gesture recognition system consists of three major parts: palm detection, hand tracking, and trajectory recognition. Figure 3 shows the whole process. The hand tracking function is trigged when the system detects an opened hand before the camera; the hand gesture classification based on HMM is activated when the user finishing the gesture. The basic algorithmic framework for our recognition process is the following.

(1) Detect the palm from video and initialize the tracker with the template of hand shape.

(2) Track the hand motion using a contour-based tracker and record the trajectory of palm center.

(3) Extract the discrete vector feature from gesture path by the global and local feature quantization.

(4) Classify the gesture using HMM which gives maximum probability of occurrence of observation sequence.

3.1. Hand Detection and Tracking

We use Adaboost algorithm with (histograms of gradient) HOG feature to detect the user's hand. The shape information of an opening hand is relatively unique in the scene. We calculate the HOG features of a new observed image to detect the opened hand at different scales and location. When the hand is detected, we update the hand color model which will be used in hand tracking. The system requires user to keep his palm opened vertically and statically before the palm is captured by the detection algorithm. In this paper, we have considered single handed dynamic gestures. A gesture is composed of a sequence of epochs. Each epoch is characterized by the motion of distinct hand shapes.

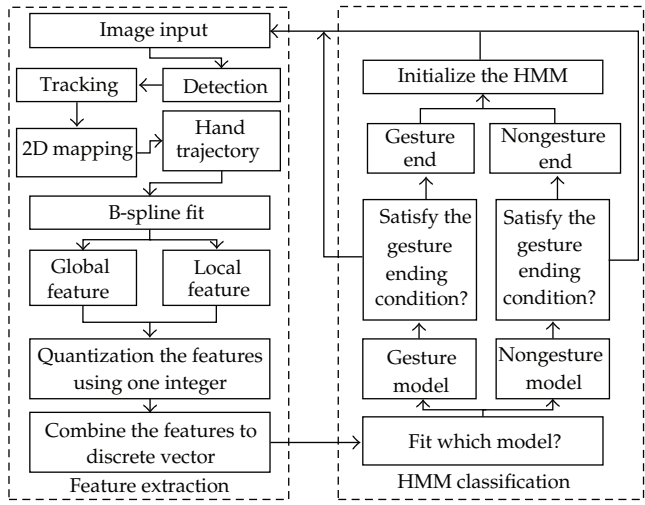


Figure 3: Overview of the hand gesture recognition process.

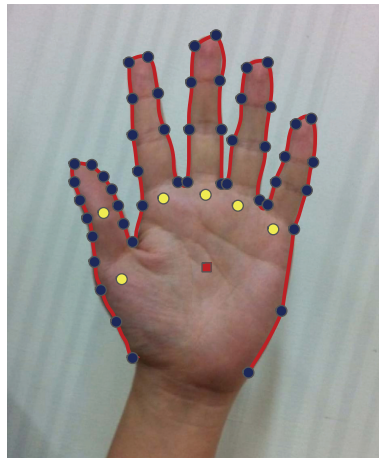


Figure 4: Hand contour.

We have implemented a contour-based hand tracker, which combines two techniques called condensation and partitioned sampling. During tracking, we record the trajectory of the hand which will be used in the hand recognition stage. The hand contour is represented with B-Splines, as shown in Figure 4. A fourteen-dimension state vector is used to describe the dynamics of the hand contour:

$$\chi = \left(t_x, t_y, \alpha, s, \theta_L, l_L, \theta_R, l_R, \theta_M, l_M, \theta_I, l_I, \theta_{\mathrm{Th1}}, \theta_{\mathrm{Th2}}\right), \tag{3.1}$$

where the subvector (t_x, t_y, α, s) is a nonlinear representation of a Euclidean similarity transform applied to the whole hand contour template, (t_x, t_y) is the palm center. (θ_L, l_L) represents the nonrigid movement of the little finger, θ_L means the little finger's angle with respect to the palm, and l_L means the little finger's length relative to its original length in

the hand template. (θ_R, l_R), (θ_M, l_M), and (θ_I, l_I) have the same meaning as the subvector (θ_L, l_L), but for different fingers. θ_{Th1} represents the angle of the first segment of the thumb with respect to the palm, and the last part θ_{Th2} represents the angle of the second segment of the thumb with respect to the first segment of the thumb.

We use a second-order autoregressive processes to predict the motion of the hand contour:

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + B\omega_t, \tag{3.2}$$

where A_1 and A_2 are fixed matrices representing the deterministic components of the dynamics, B is another fixed matrix representing the stochastic component of the dynamics, and ω_t is a vector of independent random normal $N(0, 1)$ variants.

In prediction, lots of candidate contours will be produced. We choose the one which matches the image feature (edges, boundaries of regions in skin color) best. Usually, more dimensions of the state space are required to make the condensation filter achieve considerable performance. However, this will increase computation complexity. In order to alleviate the problem, partitioned sampling is used, which divides the hand contour tracking into two steps: first, track the rigid movement of the whole hand, which is represented by (t_x, t_y, α, s); second, track the nonrigid movement of the each finger, which is represented by angle and length of each finger. The above operations can reduce the amount of candidate contours and improve the efficiency of tracking.

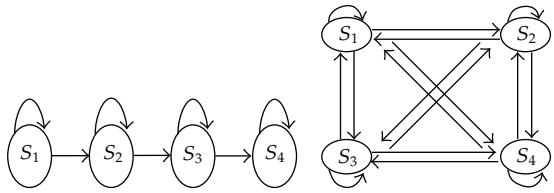
3.2. Recognition Based on HMM

After the trajectory is obtained from the tracking algorithm, features are abstracted and used to compute the probability of each gesture type with HMM. We use a vector to describe those features and as the input of the HMM.

There are three main problems for HMM: evaluation, decoding, and training, which are solved by using Forward algorithm, Viterbi algorithm, and Baum-Welch algorithm, respectively [25]. The gesture models are trained using BW re-estimation algorithm and the numbers of states are set depending on the complexity of the gesture shape.

We choose left-right banded model (Figure 5(a)) as the HMM topology, because the left-right banded model is good for modeling-order-constrained time-series whose properties sequentially change over time [26]. Since the model has no backward path, the state index either increases or stays unchanged as time increases. After finishing the training process by computing the HMM parameters for each type of gesture, a given gesture is recognized corresponding to the maximal likelihood of seven HMM models by using viterbi algorithm.

Although the HMM recognizer chooses a model with the best likelihood, we cannot guarantee that the pattern is really similar to the reference gesture unless the likelihood is high enough. A simple threshold for the likelihood often does not work well. Therefore, we produce a threshold model [22] that yields the likelihood value to be used as a threshold. The threshold model is a weak model for all trained gestures in the sense that its likelihood is smaller than that of the dedicated gesture model for a given gesture and is constructed by collecting the states of all gesture models in the system using an ergodic topology shown in Figure 5(b). A gesture is then recognized only if the likelihood of the best gesture model is higher than that of the threshold model; otherwise, it is recognized as nongesture type. Therefore, we can segment the online gestures using the threshold model.



(a) Left-right banded topology (b) Ergodic topolgy

Figure 5: HMM topologies.

4. Experiments

For experimentation, we develop a human machine interaction interface based on hand gesture. It can work with regular webcams that is connected to PC, which is used to capture live images of the users' hand movement. The minimum requirements of webcams are (1) frame rate up to 25 frames per second and (2) capture capability up to 640 × 480 pixels. The interface can be deployed in indoors environment, which generally has static background and less light changes. Hand gestures are those articulated with poses and movement with hands. The interface is able to track and recognize the following predefined hand gestures:

(1) user drawing three circles continuously in a line horizontally with hand movement in the air,

(2) user drawing a question mark (?) with hand movement in the air,

(3) user drawing three circles continuously in a line vertically with hand movement in the air,

(4) hand being vertically lifted upwards,

(5) hand waving from left to right,

(6) hand waving from right to left,

(7) user drawing an exclamation mark (!) with hand movement in the air.

For the quantification of local oriental features, we pick 18 as the codeword number from experience. Figure 6 shows the distribution histogram of central moment μ_{11} of the seven gestures as our global feature, where all the sample amounts are 450. We can set the number of the vector quantizer of global features to 20 according to the distribution. It can also be seen that the central moment feature can express the shape characteristic of trajectories. For example, gesture 1 and gesture 3, gesture 4 and gesture 7, gesture 5 and gesture 6 are close in their integral form, respectively, and it can be separated easily using the global feature.

We choose the state number of HMM for each gesture according to the experiment results and find that the recognition rate cannot be promoted when the state numbers of gestures 1 and gesture 3 are 10, and the other state numbers are set to 8. Therefore, we use this setting in the following experiments.

We collected more than 800 trajectory samples of each isolated gesture from seven people for training and more than 330 trajectory samples of each isolated gesture from eight different users for testing. The recognition results are listed in Table 1. It can be seen that the proposed method can greatly improve the recognition process, especially for those relatively complicated gestures such as predefined gesture 1 and gesture 3. It is difficult to separate

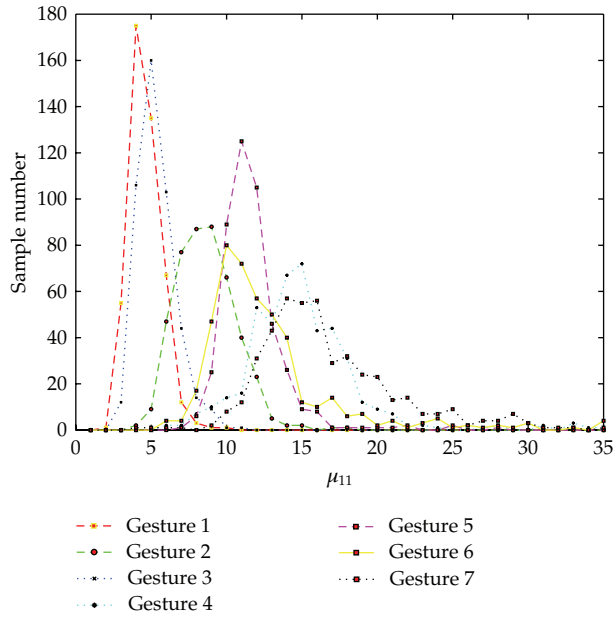


Figure 6: The distribution of μ_{11}.

Table 1: Recognition results comparison.

Gestures	Test sets' numbers	Our method (%)	Traditional method (%)
1	339	84.1	94.7
2	407	95.1	98.2
3	372	73.4	89.7
4	454	95.9	98
5	424	98.1	100
6	476	95.8	99.8
7	474	98.9	99.6

gesture 1 and gesture 3 only using local features, because their motions resemble temporally. Our algorithm can resolve this problem effectively.

5. Conlusion

We have implemented an automatic dynamic hand gesture recognition system in this paper. The user's hand is detected using Adaboost algorithm with HOG features and tracked using condensation and partitioned sampling. The trajectory of hand gesture is represented by both local and global features. Then, we take a discrete HMM method to recognize the gestures. The experimental results show that the proposed algorithm can reach better recognition results than the traditional hand recognition method. However, the tracking algorithm is still very sensitive to light and the system can only report the detection until a gesture reaches its end. Therefore, our future work will focus on improving the tracking algorithm and making the recognition more natural.

Acknowledgments

This work was supported by the Research Project of Department of Education of Zhejiang Province (Y201018160), and the Natural Science Foundation of Zhejiang Province (Y1110649).

References

[1] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: a survey of recent developments," *International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.

[2] T. Gu, L. Wang, Z. Wu, X. Tao, and J. Lu, "A pattern mining approach to sensor-based human activity recognition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 9, pp. 1359–1372, 2011.

[3] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Transactions on Systems, Man, and Cybernetics Part A*, vol. 41, no. 6, pp. 1064–1076, 2011.

[4] I. N. Junejo, E. Dexter, I. Laptev, and P. Pérez, "View-independent action recognition from temporal self-similarities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 172–185, 2011.

[5] M. Isard and A. Blake, "Condensation—conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–8, 1998.

[6] S. Chen, "Kalman filter for robot vision: a survey," *IEEE Transactions on Industrial Electronics*, vol. 59, Article ID 814356, 18 pages, 2012.

[7] J. MacCormick and A. Blake, "Probabilistic exclusion principle for tracking multiple objects," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 572–578, September 1999.

[8] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking," in *Proceedings of the European Conferene Computer Vision*, 2000.

[9] M. Tosas, *Visual articulated hand tracking for interactive surfaces*, Ph.D. thesis, University of Nottingham, 2006.

[10] X. Deyou, "A neural network approach for hand gesture recognition in virtual reality driving training system of SPG," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, pp. 519–522, August 2006.

[11] D. B. Nguyen, S. Enokida, and E. Toshiaki, "Real-time hand tracking and gesturerecognition system," in *Proceedings of the International Conference on on Graphics, Vision and Image Processing (IGVIP '05)*, pp. 362–368, CICC, 2005.

[12] E. Holden, R. Owens, and G. Roy, "Hand movement classification using an adaptive fuzzy expert system," *International Journal of Expert Systems*, vol. 9, no. 4, pp. 465–480, 1996.

[13] M. Elmezain, A. Al-Hamadi, and B. Michaelis, "Real-time capable system for handgesture recognition using hidden markov models in stereo color image sequences," *Journal of WSCG*, vol. 16, pp. 65–72, 2008.

[14] G. Saon and J. T. Chien, "Bayesian sensing hidden markov models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 43–54, 2012.

[15] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 2011.

[16] S. Eickeler, A. Kosmala, and G. Rigoll, "Hidden markov model based continuous online gesture recognition," in *Proceedings of 14th International Conference on Pattern Recognition*, vol. 2, pp. 1206–1208, 1998.

[17] N. D. Binh and T. Ejima, "Real-time hand gesture recognition using pseudo 3-d Hidden Markov Model," in *Proceedings of the 5th IEEE International Conference on Cognitive Informatics (ICCI '06)*, pp. 820–824, July 2006.

[18] L. Shi, Y. Wang, and J. Li, "A real time vision-based hand gestures recognition system," in *Proceedings of the 5th International Symposium on Advances in Computation and Intelligence (ISICA '10)*, vol. 6382, no. M4D, pp. 349–358, 2010.

[19] M. Elmezain, A. Al-Hamadi, and B. Michaelis, "Real-time capable system for handgesture recognition using hidden markov models in stereo color image sequences," *The Journal of WSCG*, vol. 16, pp. 65–72, 2008.

[20] N. Liu, B. C. Lovell, P. J. Kootsookos, and R. I. A. Davis, "Model structure selection & training algorithms for an HMM gesture recognition system," in *Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR-9 '04)*, pp. 100–105, October 2004.
[21] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, pp. 389–390, 2011.
[22] H. K. Lee and J. H. Kim, "An HMM-Based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 961–973, 1999.
[23] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 302786, 13 pages, 2012.
[24] S. Chen, J. Zhang, Q. Guan, and S. Liu, "Detection and amendment of shape distor-tions based on moment invariants for active shape models," *IET Image Processing*, vol. 5, no. 3, pp. 273–285, 2011.
[25] S. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalizedbelief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.
[26] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden markovmodel-based isolated and meaningful hand gesture recognition," *Proceedings of World Academy of Science, Engineering and Technology*, vol. 31, pp. 1307–6884, 2008.