

Hand Gesture Recognition Based on Combined Features Extraction

Mahmoud Elmezain, Ayoub Al-Hamadi, and Bernd Michaelis
Institute for Electronics, Signal Processing and Communications (IESK)
Otto-von-Guericke-University Magdeburg
{Mahmoud.Elmezain, Ayoub.Al-Hamadi}@ovgu.de

Abstract—Hand gesture is an active area of research in the vision community, mainly for the purpose of sign language recognition and Human Computer Interaction. In this paper, we propose a system to recognize alphabet characters (A-Z) and numbers (0-9) in real-time from stereo color image sequences using Hidden Markov Models (HMMs). Our system is based on three main stages; automatic segmentation and preprocessing of the hand regions, feature extraction and classification. In automatic segmentation and preprocessing stage, color and 3D depth map are used to detect hands where the hand trajectory will take place in further step using Mean-shift algorithm and Kalman filter. In the feature extraction stage, 3D combined features of location, orientation and velocity with respected to Cartesian systems are used. And then, k-means clustering is employed for HMMs codeword. The final stage so-called classification, Baum-Welch algorithm is used to do a full train for HMMs parameters. The gesture of alphabets and numbers is recognized using Left-Right Banded model in conjunction with Viterbi algorithm. Experimental results demonstrate that, our system can successfully recognize hand gestures with 98.33% recognition rate.

Keywords—Gesture Recognition, Computer Vision & Image Processing, Pattern Recognition.

I. INTRODUCTION

Sign language recognition from hand motion or hand posture is an active area in gesture recognition research for Human Computer Interaction (HCI). A gesture is spatio-temporal pattern, which may be static or dynamic or both. Static morphs of the hands are called postures and hand movements are called gestures. The goal of gesture interpretation is to push the advanced human-machine communication to bring the performance of human-machine interaction close to human-human interaction. This is due to the existing complexities in hand tracking such as hand appearance, illumination variation, and inter-hands occlusion. These issues undermine the performance and efficiency of tracking algorithms. In the last decade, several methods of potential applications [1], [2], [3], [4], [5] in the advanced gesture interfaces for HCI have been suggested but these differ from one to another in their models. Some of these models are Neural Network [1], HMMs [2] and Fuzzy Systems [5].

Liu and Lovell [6] introduced a system for hand tracking in real-time based on the Camshift algorithm and the compound constant-acceleration Kalman filter algorithm. Whereas Nobuhiko *et al.* [7] used HSV color space to track hands and face in non complex background, where the overlapping between hands and face is solved by matching templates of the previous hands and face. Comanicu *et al.* [8] proposed a

technique to track the moving objects from a moving camera using Mean-shift algorithm and Kalman filter, where the implementation of this technique achieved real-time performance. Mostly, previous approaches have not been considered many points as the combination of accurate segmentation of both hands, robust tracking that containing overlap between hands and face, and the capability of the system to run in real-time on high resolution.

Vassilia *et al.* [9] developed a system that could recognize both isolated and continuous Greek Sign Language (GSL) sentences where the orientation vector is extracted from images and then used in sentences as input to HMMs. Ho-Sub *et al.* [10] introduced a hand gesture recognition method, which used the combined features of location, angle and velocity to determine the discrete vector that is used as input to HMMs. This method runs over the alphabets (A-Z), numbers (0-9), six edit commands and six drawing elements. Nianjun *et al.* [11] proposed a method to recognize the 26 letters from A to Z by using different HMMs topologies with different states number. But, these methods run off-line over a non complex background.

Nguyen *et al.* [12] proposed a system for hand gesture recognition, where the hand is tracked by Kalman filter and hand blobs analysis to obtain motion descriptors for hand region. This system is fairly robust to background cluster and uses skin color for hand gesture tracking and recognition. Also, the system was tested to a vocabulary of 36 gestures including the American Sign Language letter spelling alphabets and digits. But, this method [12] studies the posture of the hand, not the hand motion trajectory as it is in our system. One of such problems, which arise in real-time hand gesture recognition, is to caused by the fact that the same gesture varies in shape, trajectory and duration, even for the same person. So, HMMs is used in our system where it is capable of modeling spatio-temporal time series.

The main contribution of this paper is to examine the capabilities of combined features of location, orientation and velocity for gesture recognition, which are obtained from spatio-temporal hand gesture path. Additionally, it proposes a real-time capable system that recognize the alphabets characters (A-Z) and numbers (0-9) from stereo color image sequences by the motion trajectory of a single hand using HMMs. Color and 3D depth map are used to detect hands. Furthermore, the hand trajectory is estimated using Mean-shift algorithm [13] and Kalman filter [14] in conjunction with 3D

depth map. The blob segmentation of the hands and face with complex background takes place using 3D depth map from a passive stereo camera, Gaussian Mixture Models (GMM) and color information, which is more robust to the disadvantageous lighting and partial occlusion. The depth information solve the overlapping problem between hands and face. The system is tested on a different experiments with varying features that are extracted from Cartesian systems to decide which feature is the best in terms of result. Each alphabet and each number is based on 30 video (20 for training and 10 for testing). The recognition rate that achieved on testing gestures is 98.33%. The rest of this paper is organized as follow; Section II reviews the basic HMMs technique. Section III demonstrates the suggested system in three subsections. Experimental results are described in Section IV. Finally, the summary and conclusion are presented in Section V.

II. HIDDEN MARKOV MODELS

Markov model is a mathematical model of stochastic process where these processes generate a random sequence of outcomes according to certain probabilities [11], [15], [16], [17], [18]. An HMM is a triple $\lambda = (A, B, \Pi)$ as follows:

- The set of states $S = \{s_1, s_2, \dots, s_N\}$ where N is a number of states.
- An initial probability for each state $\Pi_i, i=1, 2, \dots, N$ such that $\Pi_i = P(s_i)$ at the initial step.
- An N -by- N transition matrix $A = \{a_{ij}\}$ where a_{ij} is the probability of a transition from state S_i to $S_j; 1 \leq i, j \leq N$ and the sum of the entries in each row of matrix A must be 1 because this is the sum of the probabilities of making a transition from a given state to each of the other states.
- The set of possible emission (an observation) $O = \{o_1, o_2, \dots, o_T\}$ where T is the length of gesture path.
- The set of discrete symbols $V = \{v_1, v_2, \dots, v_M\}$ where M represents the number of discrete symbols.
- An N -by- M observation matrix $B = \{b_{im}\}$ where b_{im} gives the probability of emitting symbol v_m from state s_i and the sum of the entries in each row of matrix B must be 1 for the same pervious reason.

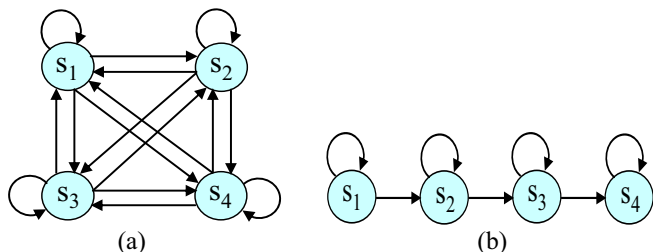


Fig. 1. HMMs topologies with 4 states (a) Ergodic topology (b) LRB topology.

There are three main problems for HMMs: Evaluation, Decoding and Training that can be solved by using Forward or Backward algorithm, Viterbi algorithm and Baum-Welch algorithm respectively [15]. Also, HMMs have a three topology: Fully Connected (Ergodic model) where any state in it can be reached from other states, Left-Right model such that each

state can go back to itself or to the following states and Left-Right Banded (LRB) model that also each state can go back to itself or the following state only (Fig. 1).

III. SUGGESTED SYSTEM

We propose an automatic system that recognizes isolated gesture for Alphabets (A-Z) and numbers (0-9) in real-time from stereo color image sequences by the motion trajectory of a single hand using HMMs. In particular, the proposed system consists of three main stages; an automatic hand segmentation and tracking, feature extraction and classification (Fig.2).

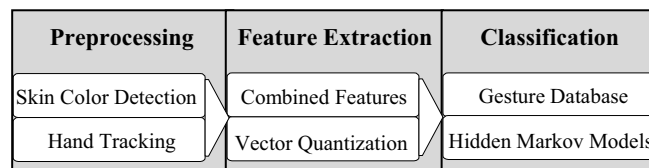


Fig. 2. Simplified structure showing the main computational modules for gesture recognition system.

- **Preprocessing**; localize and track the hand to generate its motion trajectory (gesture path).
- **Feature extraction**; Clustering extracted features to generate discrete vectors, which are used as input to HMMs recognizer.
- **Classification**; the gesture path is recognized using discrete vector and Left-Right Banded topology.

A. Automatic Hand Segmentation and Tracking

In this paper, a method for detection and segmentation of a hand in stereo color images with complex background is described where the hand segmentation takes place using 3D depth map and color information. Segmentation of skin colored regions becomes robust if only the chrominance is used in analysis. Therefore, YC_bC_r color space is used in our system where Y channel represents brightness and (C_b, C_r) channels refer to chrominance. We ignore Y channel to reduce the effect of brightness variation and use only the chrominance channels that fully represent the color information. A large database of skin and non-skin pixels is used to train the Gaussian model. The Gaussian Mixture Model begins with modeling of skin using skin database where a variant of k -means clustering algorithm [17], [19], [20] performs the model training to determine the initial configuration of GMM parameters. For the skin segmentation of hands and face in stereo color image sequences an algorithm is used, which calculates the depth value in addition to skin color information. The depth information (Fig. 3(c)) solves the overlapping problem between hands and face where it is obtained by passive stereo measuring based on cross correlation and the known calibration data of the cameras. Several clusters are composed of the resulting 3D-points. The clustering algorithm can be considered as kind of region growing in 3D that used two criteria; skin color and Euclidean distance. Furthermore, this method is more robust to the disadvantageous lighting and partial occlusion, which occur in real-time environment. In addition, blob analysis is used to derive the boundary area,

bounding box and hands centroid point. For more details, the reader can refer to [2], [21].

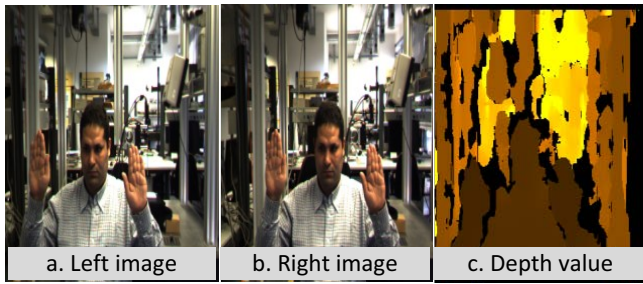


Fig. 3. (a) Left image frame of video stream. (b) Right image. (c) The depth value of left and right image via the Bumblebee stereo camera.

After localization of the hand's target from the segmentation step, we find its color histogram with Epanechnikov kernel [8]. This kernel assigns smaller weights to pixels father from the center to increases the robustness of the density estimation. To find the best match of our hand target in the sequential frames, the Bhattacharyya coefficient [22] is used to measure the similarity by maximizing Bayes error that arising from the comparison of the hand target and candidate. We take in our consideration the mean depth value that is computed from the previous frame for the hand region to solve the overlapping between hands and face. The mean-shift procedure is defined recursively and performs the optimization to compute the mean shift vector. After each mean-shift optimization that gives the measured location of the hand target, the uncertainty of the estimate can also be computed and then followed by the Kalman iteration, which drives the predicated position of the hand target. Thereby, the hand gesture path is obtained by taking the correspondences of detected hand between successive image frames (Fig. 5(d)). For more details, the reader can refer to [2], [8], [21].

B. Feature Extraction

There is no doubt that selecting good features to recognize the hand gesture path plays significant role in system performance. There are three basic features; location, orientation and velocity. We analyze the effectiveness of these features that are extracted from a hand trajectory and also combine them to test their recognition rate. A gesture path is spatio-temporal pattern that consists of hand centroid points (x_{hand}, y_{hand}) . The coordinates in the Cartesian space can be extracted from gesture frames directly. We consider two types of location features. The first location feature is L_c that measures the distance from the centroid point to all points of the gesture path because different location features are generated for the same gesture according to different starting points (Eq.1). The second location feature is L_{sc} , which is computed from the start point to the current point of hand gesture path (Eq.3).

$$L_{c_t} = \sqrt{(x_{t+1} - C_x)^2 + (y_{t+1} - C_y)^2} \quad (1)$$

$$(C_x, C_y) = \frac{1}{n} \left(\sum_{t=1}^n x_t, \sum_{t=1}^n y_t \right) \quad (2)$$

$$L_{sc_t} = \sqrt{(x_{t+1} - x_1)^2 + (y_{t+1} - y_1)^2} \quad (3)$$

where $t = 1, 2, \dots, T - 1$ and T represents the length of hand gesture path. (C_x, C_y) refers to the centroid of gravity at the n points. To verify the real-time implementation, the centroid point of gesture path is computed after each frame.

The second basic feature is the orientation, which gives the direction along that the hand when traverses in space during the gesture making process. As described above, the orientation feature is based on the calculation of the hand displacement vector at every point and is represented by the orientation according to the centroid of gesture path (θ_{1t}), the orientation between two consecutive points (θ_{2t}) and the orientation between start and current gesture point (θ_{3t}).

$$\theta_{1t} = \arctan \left(\frac{y_{t+1} - C_y}{x_{t+1} - C_x} \right) \quad (4)$$

$$\theta_{2t} = \arctan \left(\frac{y_{t+1} - y_t}{x_{t+1} - x_t} \right) \quad (5)$$

$$\theta_{3t} = \arctan \left(\frac{y_{t+1} - y_1}{x_{t+1} - x_1} \right) \quad (6)$$

The third basic feature is the velocity, which plays an important role during gesture recognition phase particularly at some critical situations. The velocity is based on the fact that each gesture is made at different speeds where the velocity of the hand decreases at the corner point of a gesture path. The velocity is calculated as the Euclidean distance between the two successive points divided by the time in terms of the number of video frames as follows;

$$V_t = \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2} \quad (7)$$

In the Cartesian coordinate system, we use different combination of features to obtain a variety feature vectors. For example, the feature vector at frame $t + 1$ is obtained by union of locations features (L_{c_t}, L_{sc_t}) , locations features with velocity feature (L_{c_t}, L_{sc_t}, V_t) , orientations features $(\theta_{1t}, \theta_{2t}, \theta_{3t})$, orientations features with velocity feature $(\theta_{1t}, \theta_{2t}, \theta_{3t}, V_t)$ and locations features with orientations features and velocity feature $(L_{c_t}, L_{sc_t}, \theta_{1t}, \theta_{2t}, \theta_{3t}, V_t)$.

Each frame contains a set of feature vectors at time t where the dimension of space is proportional to the size of feature vectors. In this manner, gesture is represented as an ordered sequence of feature vectors, which are projected and clustered in space dimension to obtain discrete codeword that are used as an input to HMMs. This is done using k -means clustering algorithm [19], [20], which classifies the gesture pattern into K clusters in the feature space.

1) *Vector Quantization*: The extracted features are quantized to obtain the discrete symbols. When the basic features such as locations and velocity are used separately, these features are normalized and multiplied by a different scalar ranging from 10 to 30. On the other side, the normalization of the orientation features is to divide by 10° , 20° , 30° and 40° to obtain its codeword. In addition to the combination features in the Cartesian system, we use k -mean clustering algorithm to classify the gesture feature into K clusters on the feature space. This algorithm is based on the minimum

distance between the center of each cluster and the feature point. We divide the set of feature vectors into set of clusters. This allows us to model the hand trajectory in the feature space by one cluster. The calculated cluster index is used as input (i.e. observation symbol) to the HMMs. Furthermore, we usually do not know the best number of clusters in the data set. In order to specify the number of clusters K for each execution of the k -means algorithm, we considered $K = 28, 29, \dots, 37$, which is based on the numbers of segmented parts in all alphabets character (A-Z) and numbers (0-9) where each straight-line segment is classified into single cluster.

Suppose we have n sample of trained feature vectors x_1, x_2, \dots, x_n all from the same class, and we know that they fall into k compact clusters, $k < n$. Let m_i be the mean of the vectors in cluster i . If the clusters are well separated, a minimum distance classifier is used to separate them. That is, we can say that x is in cluster i if $\|x - m_i\|$ is the minimum of all the k distances. The following procedure shows that how to find the k -means;

- Build up randomly an initial Vector Quantization Codebook for the means m_1, m_2, \dots, m_k
- Until there are no changes in any mean
 - Use the estimated means to classify each sample of train vectors into one of the clusters m_i
 for $i=1$ to k
 - Replace m_i with the mean of all of the samples of trained vector for cluster i
 - end (for)
- end (Until)

A general observation is that different gestures have different trajectories in the cluster space, while the same gesture show very similar trajectories.

C. Classification

The final stage in our system is classification. Throughout this stage, Baum-Welch algorithm [15] is used to do a full training for the initialized HMMs parameters to construct gestures database. Each reference pattern in the gesture database for alphabets (A-Z) and numbers (0-9) is modeled by Left-Right Banded model with varying number of states ranging from 3 to 6 states based on its complexity. As, the excessive number of states can generate the over-fitting problem if the number of training samples is insufficient compared to the model parameters. The hand gesture path is classified by selecting the maximal observation probability of gestures model. The maximal gesture model is the gesture whose observation probability is the largest among all 36 gestures (A-Z & 0-9). The type of observed gesture (O) is decided by Viterbi algorithm frame by frame (i.e. accumulatively until it receives the gesture end signal). The following steps show how the Viterbi algorithm works on gesture model $\lambda_g(a^g, b^g, \Pi^g)$ (Fig. 4):

1. Initialization: for $1 \leq i \leq N$,
 - $\delta_1^g(i) = \Pi_i^g \cdot b_i^g(o_1)$

2. Recursion (accumulative observation probability computation): for $2 \leq t \leq T$, $1 \leq j \leq N$,
 - $\delta_t^g(j) = \max_i [\delta_{t-1}^g(i) \cdot a_{ij}^g] \cdot b_j^g(o_t)$
3. Termination:
 - $P(O|\lambda_g) = \max_i [\delta_T^g(i)]$

where N is the number of states, Π_i^g represents the initial value for the state i , a_{ij}^g is the transition probability from state i to state j , $b_j^g(o_t)$ refers to the probability of emitting o at time t in state j , and $\delta_t^g(j)$ represents the maximum likelihood value in state j at time t .

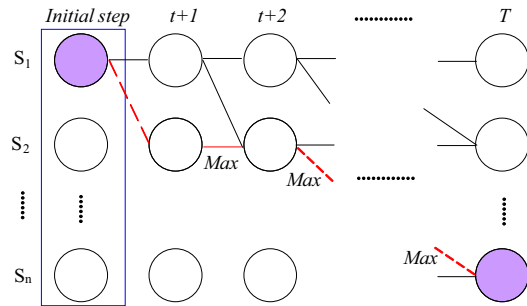


Fig. 4. The best path for LRB model with N states where it starts from S_1 to S_N , $N=3, 4, \dots, 6$ and $t=1$.

IV. EXPERIMENTAL RESULTS

Our proposed system has been tested on various video sequences with various hand shape as well as overlapping and partial occlusion. Our proposed system was capable for real-time implementation and showed good results to recognize alphabets character and numbers from stereo color image sequences via the motion trajectory of a single hand using HMMs. The input images were captured by Bumblebee stereo camera system that has 6 mm focal length at 15FPS with 240×320 pixels image resolution, Matlab implementation. In our experimental results, each isolated gesture was based on 30 video sequences, which 20 video samples for training by BW algorithm and 10 video samples for testing (Totally, our database contains 720 video sample for training and 360 video sample for testing). The gesture recognition module match the hand gesture path against the database of reference gestures, to classify which class it belongs to. The higher priority was computed by Viterbi algorithm to recognize the alphabets and numbers in real-time frame by frame over LRB topology with different number of states ranging from 3 to 6.

We test the importance of the three basic features (location, orientation, velocity) in the Cartesian coordinate. Moreover, the observation sequence for HMMs is quantified either by normalization in case of separated features or by the k -means clustering algorithm in case of combined features. From table 1, the recognition ratio of isolated gestures achieved best results with 98.33% using $(Lc, Lsc, \theta_1, \theta_2, \theta_3, V)$ feature. The recognition ratio is the number of correctly recognized gestures to the number of tested gestures (Eq. 8).

$$Reco. \text{ ratio} = \frac{\# \text{ recognized gestures}}{\# \text{ test gestures}} \times 100\% \quad (8)$$

According to the separated features in Fig. 5(a) & (b), the orientation features $(\theta_1, \theta_2, \theta_3)$ are better rather than the

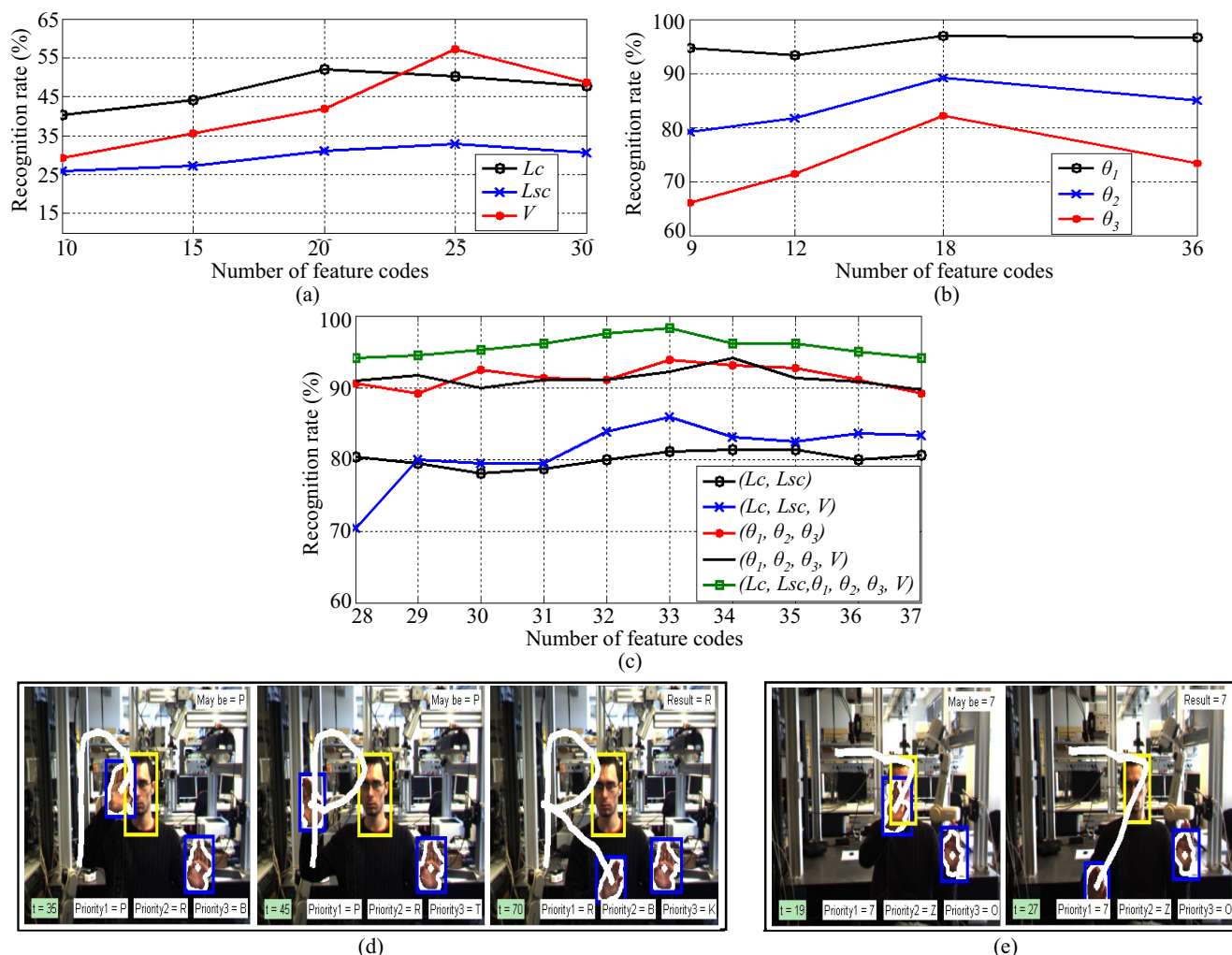


Fig. 5. (a) Recognition rate for the number of locations and velocity feature codes (10, 15, 20, 25, 30). (b) Results for three different orientations with varying feature codes number (9, 12, 18, 36). (c) Recognition rate according to a combined features in Cartesian system over feature codes number from 28 to 37. (d) The high priority is alphabet 'P' at $t=45$ and at $t=70$ the result is 'R'. (e) Solving overlap problem between hand and face at $t=19$ and the high priority is '7' at $t=27$.

recognition rate of the locations (Lc, Lsc) or the velocity (V) features. This in turn leads to the orientation feature ($\theta_1 = 96.94\%$) is the most effective among the three basic features. Furthermore, the velocity feature shows a lower discrimination power (57.22%) than that of the orientation features. Also, the Lsc feature result is the lowest recognition rate of 32.78% . Additionally, the (Lc, Lsc, V) , $(\theta_1, \theta_2, \theta_3, V)$ and $(Lc, Lsc, \theta_1, \theta_2, \theta_3, V)$ features, which include the velocity information show higher recognition than when using the velocity information (Fig. 5(c)). In short, Fig. 5 show the results of the experiments that were performed to determine the best feature codes number (the best number of feature code is 33 for the feature $(Lc, Lsc, \theta_1, \theta_2, \theta_3, V)$). Fig. 5(d) & (e) shows the output of the system for gesture alphabet 'R' and number '7' respectively, in addition to the solved overlapping problem between hand and face by 3D depth map.

V. CONCLUSION

This paper proposes a system to recognize the alphabets character (A - Z) and numbers (0 - 9) from stereo color image sequences by the motion trajectory of a single hand using HMMs. This system uses the combined features of

location, orientation and velocity for Cartesian systems. We have shown that the effective of these features can yield reasonable recognition rates. The database contains 720 video samples for training and 360 video sequences for testing the isolated gestures. The proposed system has shown good performance when applied on several video samples containing confusing situations such as partial occlusion and overlapping. The results show that; the proposed system is suitable for real-time application and can successfully recognize hand gestures with 98.33% recognition rate. The future research will address the hand gesture spotting and recognition for a sentence using fingertip instead of the hand centroid point in conjunction with multi-camera system.

ACKNOWLEDGMENT

This work is supported by BMBF Bernstein-Group (FKZ: 01GQ0702), LSA grants (C4-NIMITEK 2, FKZ: UC4-3704M), Forschungspraemie (BMBF-Frderung, FKZ: 03FPB00213) and Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by DFG.

TABLE I

RESULTS OF HAND GESTURES ACCORDING TO DIFFERENT FEATURES EXTRACTION IN CARTESIAN SYSTEM WITH THE BEST FEATURE CODE NUMBER.

Feature type	Feature space	# Feature code	Training data	Hand gestures results		
				Testing data	Correct data	Recognition (%)
Separated in Cartesian coordinates	Lc	20	720	360	187	51.94
	Lsc	25	720	360	118	32.78
	v	25	720	360	206	57.22
	θ_1	18 ; 36	720	360	349	96.94
	θ_2	18	720	360	321	89.17
	θ_3	18	720	360	236	82.22
Union in Cartesian coordinates	(Lc, Lsc)	35	720	360	293	81.39
	(Lc, Lsc, V)	33	720	360	309	85.83
	$(\theta_1, \theta_2, \theta_3)$	33	720	360	338	93.89
	$(\theta_1, \theta_2, \theta_3, V)$	34	720	360	339	94.16
	(Lc, Lsc, $\theta_1, \theta_2, \theta_3, V$)	33	720	360	354	98.33

REFERENCES

[1] X. Deyou, *A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG*, International Conference ICPR, pp. 519-522, 2006.

[2] M. Elmezain, A. Al-Hamadi, and B. Michaelis, *Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences*, The Journal of WSCG, Vol. 16, No. 1, pp. 65-72, 2008.

[3] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, *A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory*, International Conference on Pattern Recognition (ICPR) pp. 1-4, 2008.

[4] M. Elmezain, A. Al-Hamadi, and B. Michaelis, *A Novel System for Automatic Hand Gesture Spotting and Recognition in Stereo Color Image Sequences*, The Journal of WSCG, Vol. 17, No. 1, pp. 89-96, 2009.

[5] E. Holden, R. Owens, and G. Roy, *Hand Movement Classification Using Adaptive Fuzzy Expert System*, The Journal of Expert Systems, Vol. 9(4), pp. 465-480, 1996.

[6] N. Liu, and B. C. Lovell, *MMX-accelerated Real-Time Hand Tracking System*, In IVCNZ, pp. 381-385, 2001.

[7] T. Nobuhiko, S. Nobutaka, and S. Yoshiaki, *Extraction of Hand Features for Recognition of Sign Language Words*, In International Conference of VI, pp. 391-398, 2002.

[8] D. Comaniciu, V. Ramesh, and P. Meer, *Kernel-Based Object Tracking*, The IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 25, pp. 564-577, 2003.

[9] N. P. Vassilia and G. M. Konstantinos, *On Feature Extraction and Sign Recognition for Greek Sign Language*, International Conference on Artificial Intelligence and Soft Computer, pp. 93-98, 2003.

[10] Y. Ho-Sub, S. Jung, J. B. Young, and S. Y. Hyun, *Hand Gesture Recognition using Combined Features of Location, Angle and Velocity*, Journal of Pattern Recognition, Vol. 34(7), pp. 1491-1501, 2001.

[11] L. Nianjun, C. L. Brian, J. K. Peter, and A. D. Richard *Model Structure Selection & Training Algorithms for a HMM Gesture Recognition System*, International Workshop IWFHR, pp. 100-105, 2004.

[12] D. B. Nguyen, S. Enokida, and E. Toshiaki, *Real-Time Hand Tracking and Gesture Recognition System*, In GVIP Conf., pp. 362-368, 2005.

[13] D. Comaniciu, V. Ramesh, and P. Meer, *Real-Time Tracking of Non-Rigid Objects Using Mean Shift*, In Conference CVPR, pp. 1-8, 2000.

[14] G. Welch, and G. Bishop, *An Introduction to the Kalman Filter*, In Technical Report, University of North Carolina at Chapel Hill, pp. 95-041, 1995.

[15] R. R. Lawrence, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceeding of the IEEE, Vol. 77(2), pp. 257-286, 1989.

[16] S. Mitra, and T. Acharya, *Gesture Recognition: A Survey*, IEEE Transactions on Systems, MAN, and Cybernetics, pp. 311-324, 2007.

[17] M. Elmezain, A. Al-Hamadi, S. S. Pathan, and B. Michaelis, *Spatio-Temporal Feature Extraction-Based Hand Gesture Recognition for Isolated American Sign Language and Arabic Numbers*. IEEE Symposium on ISPA, pp. 254-259, 2009.

[18] M. Elmezain, A. Al-Hamadi, G. Krell, S. El-Etriby, and B. Michaelis, *Gesture Recognition for Alphabets from Hand Motion Trajectory Using Hidden Markov Models*, The IEEE International Symposium on Signal Processing and Information Technology, pp. 1209-1214, 2007.

[19] T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu, *An Efficient k-means Clustering Algorithm: Analysis and Implementation*, IEEE Transaction on PAMI, Vol. 24, pp. 881-892, 2002.

[20] C. Ding and X. He, *K-means Clustering via Principal Component Analysis*, International Conference on ML, pp. 225-232, 2004.

[21] R. Niese, A. Al-Hamadi, and B. Michaelis, *A Novel Method for 3D Face Detection and Normalization*, Journal of Multimedia, Vol. 2, pp. 1-12, 2007.

[22] S. Khalid, U. Ilyas, S. Sarfaraz, and A. Ajaz, *ABhattacharyya Coefficient in Correlation of Gary-Scale Objects*, The Journal of Multimedia, Vol. 1, pp. 56-61, 2006.



Mahmoud Elmezain was born in Egypt. He received his Masters Degree in Computer Science in 2004. Between 1997 and 2004 he worked as Demonstrator in Dept. of Statistic and Computer Science. Since 2004 he is Assistant lecturer in Dept. of Computer Science, Faculty of Science, Tanta University, Egypt. His current work on a Ph.D. thesis focuses on image processing, pattern recognition and human-computer interaction, at the Institute for Electronics, Signal Processing and Communications at Otto-von-Guericke University of Magdeburg, Germany.



Ayoub K. Al-Hamadi was born in Yemen in 1970. He received his Masters Degree in Electrical Engineering & Information Technology in 1997 and his Ph.D. in Technical Computer Science at the Otto-von-Guericke University of Magdeburg, Germany in 2001. Since 2002 he has been Assistant Professor and 2005 Post-Doc in KFST in Magdeburg. 2004 until 2005 he graduated Professional Training for Industrial Project Management and Start-Up of Business Establishment at University Magdeburg, Germany. Between 2006 and 2008 he has been Junior-Research-Group-Leader at the Institute for Electronics, Signal Processing and Communications at the Otto-von-Guericke University Magdeburg. In August 2008 he became Professor of Neuro-Information Technology at the Otto-von-Guericke University Magdeburg. His research work concentrates on the field of image processing, computer vision, pattern recognition, human-computer interaction, artificial intelligence and information technology. Prof. Dr.-Ing. Al-Hamadi is the author of more than 100 articles in peer-reviewed international journals and conferences.



Bernd Michaelis was born in Magdeburg, Germany in 1947. He received a Masters Degree in Electronic Engineering from the TH Magdeburg in 1971 and his first Ph.D. in 1974. Between 1974 and 1980 he worked at the TH Magdeburg and was granted a second doctoral degree in 1980. In 1993 he became Professor of Technical Computer Science at the Otto-von-Guericke University Magdeburg. His research work concentrates on the field of image processing, artificial neural networks, pattern recognition, processor architectures, and microcomputers. Professor Michaelis is the author of more than 200 articles.