International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:5, 2008

A Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition

Mahmoud Elmezain, Ayoub Al-Hamadi, Jörg Appenrodt, and Bernd Michaelis
Institute for Electronics, Signal Processing and Communications (IESK)
Otto-von-Guericke-University Magdeburg
D-39106 Magdeburg, Germany
{Mahmoud.Elmezain, Ayoub.Al-Hamadi}@ovgu.de

***Abstract*—**Gesture recognition is a challenging task for extracting meaningful gesture from continuous hand motion. In this paper, we propose an automatic system that recognizes isolated gesture, in addition meaningful gesture from continuous hand motion for Arabic numbers from 0 to 9 in real-time based on Hidden Markov Models (HMM). In order to handle isolated gesture, HMM using Ergodic, Left-Right (LR) and Left-Right Banded (LRB) topologies is applied over the discrete vector feature that is extracted from stereo color image sequences. These topologies are considered to different number of states ranging from 3 to 10. A new system is developed to recognize the meaningful gesture based on zero-codeword detection with static velocity motion for continuous gesture. Therefore, the LRB topology in conjunction with Baum-Welch (BW) algorithm for training and forward algorithm with Viterbi path for testing presents the best performance. Experimental results show that the proposed system can successfully recognize isolated and meaningful gesture and achieve average rate recognition 98.6% and 94.29% respectively.

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

I. Introduction

The hand gesture recognition is an active area of research in the vision community, mainly for the purpose of sign language recognition and 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. In the last decade, several methods of potential applications [1], [2], [3], [4] in the advanced gesture interfaces for HCI have been suggested but these differ from one another in their models. Some of these models are Neural Network [1], Fuzzy Systems [5] and HMM [4]. HMM is a statistical model where the distributed initial points work well and the output distributions are automatically learned by the training process. In addition, HMM is widely used in hand writing, speech and character recognition [4]. Another advantage of HMM is that it is capable of modeling spatio-temporal time series where the same gesture can differ in shape and duration. There are two major problems arising in real-time gesture recognition system for continuous gesture to extract meaningful gesture. The first problem is the segmentation that means how to determine when a gesture starts and when it ends from hand motion trajectory, which is due to the intermediate movement of the hand between two gestures. The second problem is that the same gesture varies in shape and duration. Vassilia *et al.* [3] developed a system that could recognize both isolated and continuous Greek Sign Language (GSL) sentences for hand postures where the orientation vector is extracted from images, which is used in sentences as input to HMM. Nianjun *et al.* [6] proposed a method to recognize the 26 letters from A to Z by using different HMM topologies with different states. Lee *et al.* [7] proposed an ergodic model based on adaptive threshold to classify the meaningful gestures by combining all states from all trained gesture models using HMM. But, all these methods run off-line over a non complex background. Elmezain *et al.* [4] developed a system to recognize the alphabets letter from A to Z and Arabic number from 0 to 9 in real-time from stereo color image sequences using LRB topology of HMM with 9 states. Nguyen *et al.* [2] introduced a hand gesture recognition system to recognize real-time gesture in unconstrained environments where the system was tested to a vocabulary of 36 gestures including the American Sign Language (ASL) letter spelling alphabet and digits. The previous method [2] runs in real-time over a complex background, and it studies the hand posture, not the hand motion trajectory as it is in our system.

In this paper, the focus is to designing of HMM topologies with different states from 3 to 10 in order to decide which topology is the best in terms of results for isolated gesture. So, we develop a system to recognize the isolated and continuous Arabic numbers from 0 to 9 in real-time from color image sequences by the motion trajectory of a single hand using HMM. The proposed system depends upon the following main steps; using Gaussian Mixture Model (GMM) for skin color detection, the orientation between two consecutive points is extracted as basic feature, zero-codeword detection with static velocity motion for meaningful gesture, BW algorithm for training and forward algorithm in conjunction with Viterbi path for testing. Moreover, each isolated gesture number is based on 30 video sequences (20 for training and 10 for testing) and the continuous gestures are also based on 70 video sequences for testing where the input images are captured by a Bumblebee stereo camera that has 6mm focal length for about 2 to 5 second at 15 frames per second with 240×320 pixels image resolution on each frame. The achievement recognition rates on isolated and meaningful gestures are 98.6% and

94.29% respectively. The organization of this paper is as follows; Section II demonstrates the suggested system in three subsections. The experimental results are described in Section III. Finally, Section IV ends with summary and conclusion.

II. Gesture Recognition System

We propose a system that recognizes both isolated and meaningful gesture for Arabic numbers from 0 to 9 in real-time from stereo color image sequences by the motion trajectory of a single hand using HMM. Our main motivation is to improve gesture recognition in natural conversation. This requires techniques for skin segmentation and handling occlusion between hands and faces to overcome the difficulties of overlapping regions. In particular, the gesture recognition system consists of three main stages (Fig. 1, Fig. 8).

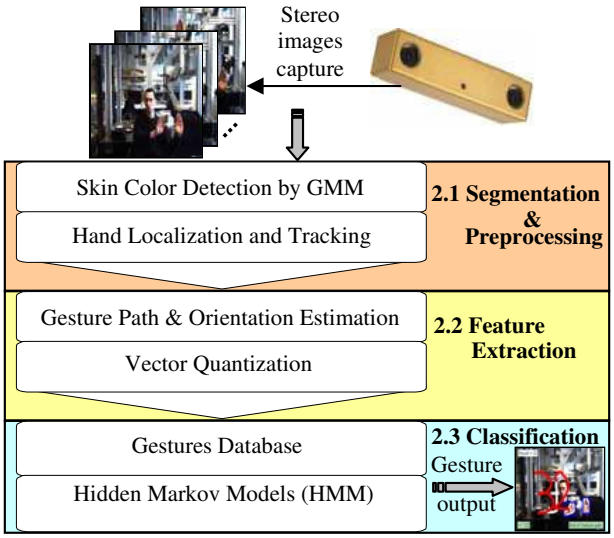

Fig. 1. Gesture recognition system using Hidden Markov Models (HMM).

- Automatic segmentation and preprocessing; the hand is segmented, localized and tracked to generate its motion trajectory (gesture path) by using GMM for skin color detection.
- Feature extraction; determine the discrete vector feature from gesture path by the orientation quantization.
- Classification; the hand motion trajectory is recognized by discrete vector feature, HMM topologies and forward algorithm in conjunction with Viterbi path.

A. Automatic segmentation and preprocessing

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. Moreover, morphological operations are used as a preprocessing to remove the remaining errors. This stage contains two steps; skin segmentation by using GMM over YC_bC_r color space [8] and hand tracking by blob analysis.

1) Skin segmentation via a GMM: 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 [9], [10]. We ignore Y channel in order to reduce the effect of brightness variation and use only the chrominance channels which fully represent the color information. A large database of skin and non-skin pixel is used to train the Gaussian model (Fig. 2, Fig. 3). In the training set, 18972 skin pixels from 36 different races persons and 88320 non-skin pixels from 84 different images are used.

Fig. 2. Database of skin pixel where these cropped images were collected from the World Wide Web for different races.

Fig. 3. Database of non skin pixel where these cropped images were collected from the World Wide Web for different background.

The GMM technique begins with modeling of skin by using skin database where a variant of K-means clustering algorithm [10] performs the model training to determine the initial configuration of mean vector μ covariance matrix Σ and mixture weight (Table I).

TABLE I
Gaussian mixture model for skin color database that contains the mean vector, covariance matrix and mixture weight for $K = 4$ clusters in our system.

K	Mean μ	Covariance Σ	Weight
1	$\begin{pmatrix} 119.5; 144.1 \end{pmatrix}$	$\begin{pmatrix} 35.81 & -13.5 \\ -13.5 & 14.88 \end{pmatrix}$	0.2422
2	$\begin{pmatrix} 110.3; 153.2 \end{pmatrix}$	$\begin{pmatrix} 13.34 & 2.124 \\ 2.124 & 5.73 \end{pmatrix}$	0.2612
3	$\begin{pmatrix} 98.6; 165.9 \end{pmatrix}$	$\begin{pmatrix} 46.09 & -21.6 \\ -21.6 & 46.82 \end{pmatrix}$	0.1668
4	$\begin{pmatrix} 103.1; 157.3 \end{pmatrix}$	$\begin{pmatrix} 16.83 & -1.26 \\ -1.26 & 16.94 \end{pmatrix}$	0.3298

Suppose that $x = [C_b; C_r]^T$ represents the chrominance

vector of an input pixel. The probability of skin pixel over vector x for mixture model is a linear combination of its probabilities and is calculated as follows:

$$p(x|skin) = \sum_{i=1}^{K} p(x|i).p(i) \quad (1)$$

where K is the number of Gaussian components (K has a value 4 in our experiment) and is estimated by a constructive algorithm that is used the criteria of maximizing likelihood function [11]. *p(i)* is the mixture weight and $p(x|i)$ is the Gaussian density model for the i^{th} component.

$$p(x|i) = \frac{e^{-1/2(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}}{(2\pi)^{f/2}\sqrt{|\Sigma_i|}} \quad (2)$$

where μ_i and Σ_i represent the mean vector and the covariance matrix of i^{th} component respectively and f is the dimension of feature space, $x \in R^f$.

$$\sum_{i=1}^{K} p(i) = 1 \; ; \; 0 \leq p(i) \leq 1 \quad (3)$$

The Expectation Maximization (EM) algorithm [8], [12] is used to estimate the maximum likelihood of parameters (mean vector, covariance matrix and mixture weight), which run on the training database of skin pixels. For the probability $p(x|non-skin)$, the non skin color pixels are modeled as a unimodel Gaussian in order to reduce the computational complexity of skin probability calculation (Table II). For more details, the reader can refer to [4].

TABLE II
UNIMODEL GAUSSIAN FOR NON SKIN COLOR.

Mean μ	Covariance Σ
$\begin{pmatrix} 85.65; 84.24 \end{pmatrix}$	$\begin{pmatrix} 13.22 & -8.73 \\ -8.73 & 19.01 \end{pmatrix}$

For the skin segmentation of hands and face in stereo color image sequences an algorithm is used, which calculates the depth information in addition to skin color information [13]. The depth information can be gathered 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 [13]. The clustering algorithm can be considered as kind of region growing in 3D which 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 (for instance, in case of gesture recognition). For more details, the reader can refer to [14]. By the given depth information from the camera set-up system (Fig. 4 (b)), the overlapping problem between hands and face is solved since the hand regions are closer to the camera rather than the face region. For removing the outliers (noise, spurious components) from the skin probability image, we use morphological operation (median filter, erosion and dilation) since there are small regions that are close to skin but does not belong to the human skin. Thereby, the skin color regions are detected (hands and face). Fig. 4(a) shows the first frame of the stereo color image sequences.

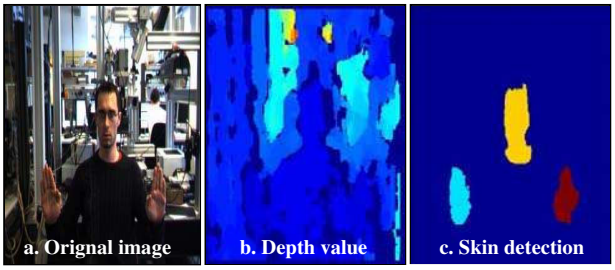

Fig. 4. Skin segmentation. (a) First frame of stereo color image sequences (b) Depth information of the original image from a Bumblebee stereo camera. (c) Labeled skin color detection after using morphological operation.

2) Hand localization & tracking: After the labeled skin image is determined (Fig. 4(c)), the localization of two hands is found by selecting the two small areas (Fig. 5(a)) where the face represents the bigger area and the furthest away from the camera. In addition, we use a blob analysis to determine the boundary area, the bounding box and the centroid point of each hand region. Our attention concentrates on the motion of a single hand in order to detect the hand motion trajectory for a specific number. Consequently, we select a search area in the next frame (Fig.5(b)) around the bounding box that is determined from the last frame in order to track the hand and to reduce the gesture region of interest. Thereby, the new bounding box is calculated and the centroid point is determined. By iteration of this process, the motion trajectory of the hand so-called gesture path is generated from connecting hand centroid points (Fig. 6).

Fig. 5. Hand localization and search area (a) Hand localization with a boundary area, bounding box and centroid point (b) Search area around the hands in the next frame.

B. Feature extraction

There is no doubt that selecting good features to recognize the hand gesture path play significant role in system performance. There are three basic features; location, orientation and velocity. The previous research [4], [6] showed that the orientation feature is the best in terms of accuracy results.

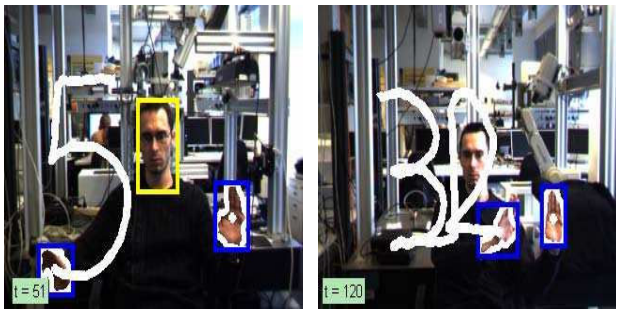

Fig. 6. Gesture path for isolated number 5 and continuous number 32 from connected centroid points of hand regions.

Therefore, we will rely upon it as a main feature in our system. A gesture path is spatio-temporal pattern which consists of centroid points (x_{hand}, y_{hand}). So, the orientation is determined between two consecutive points from hand gesture path by Eq. 4.

$$\theta_t = \arctan\left(\frac{y_{t+1} - y_t}{x_{t-1} - x_t}\right) \quad ; t = 1, 2, ..., T-1 \qquad (4)$$

where T represents the length of gesture path. The orientation θ_t is quantized by dividing it by 20° in order to generate the codewords from 1 to 18 (Fig. 7). Also the codewords contain zero codeword notably in case of continuous gesture. Thereby, the discrete vector is determined and then is used as input to HMM.

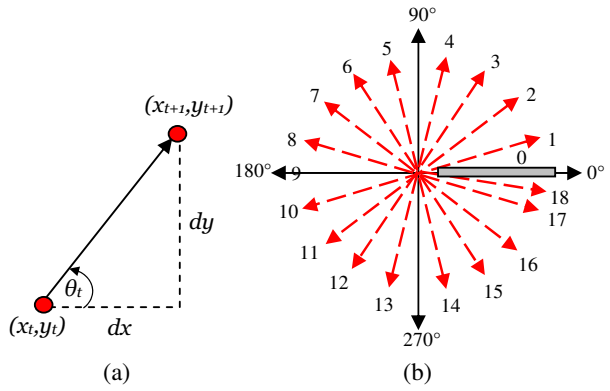

Fig. 7. The orientation and its codewords (a) orientation between two consecutive points (b) directional codewords from 1 to 18 including also zero codeword.

C. Gesture path classification

The final stage in our system is classification. Throughout this stage, the isolated and continuous gestures paths are recognized by HMM forward algorithm in conjunction with Viterbi path [15] over its discrete vector. Moreover, Baum-Welch algorithm [15] is used to do a full training for the initialized HMM parameters by the discrete vector to construct gestures database. The gestures database contain 70 video sequences for continuous gestures and 30 video sequences for each isolated gesture number from 0 to 9. The following subsections describe this stage in some details.

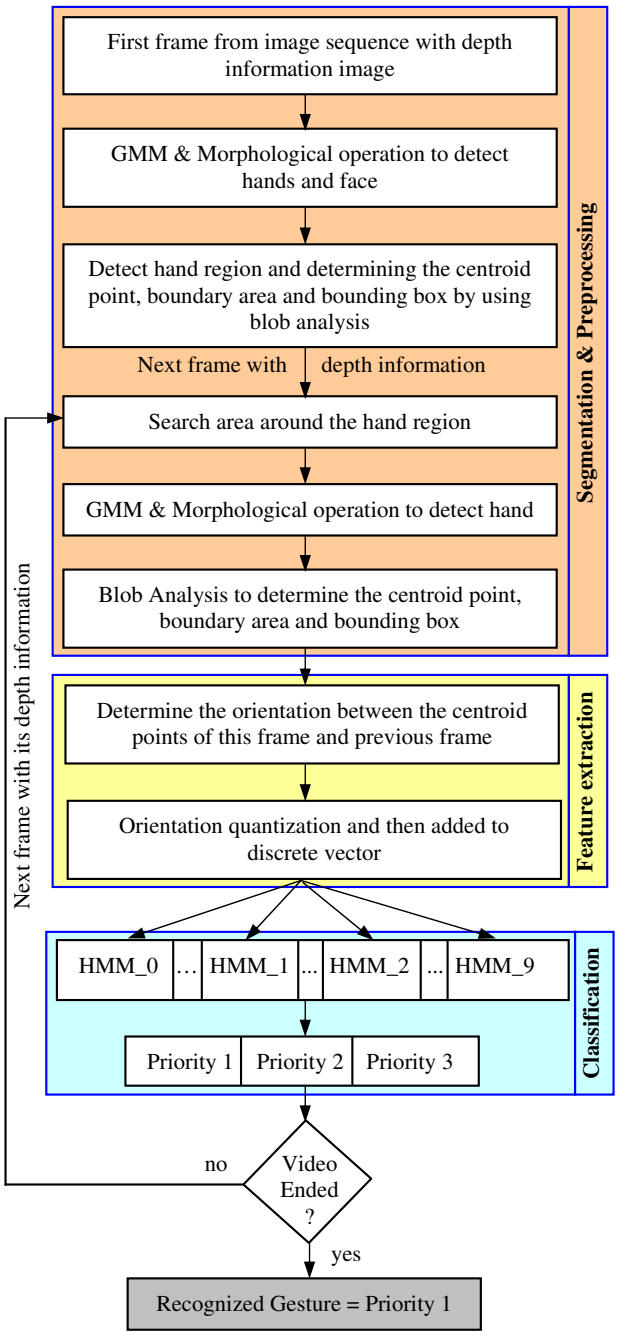

Fig. 8. Real-time system for isolated Arabic numbers gesture recognition.

1) 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 [6], [15], [16], [17], [18]. An HMM is a triple $\lambda = (A, B, \Pi)$ as follows:

- The set of states $S = \{s_1, s_2, ..., s_N\}$ where N is a number of states.
- An initial probability for each state Π_i, i=1, 2, ..., 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) $O = \{o_1, o_2, ..., o_T\}$ where T is the length of gesture path.
- The set of discrete symbols $V = \{v_1, v_2, ..., v_M\}$ 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.

There are three main problems for HMM: Evaluation, Decoding and Training that can be solved by using Forward or Backward algorithm [15], Viterbi algorithm and Baum-Welch algorithm respectively. Also, HMM has a three topology: Fully Connected (Ergodic model) where any state in it can be reached from any 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. 9).

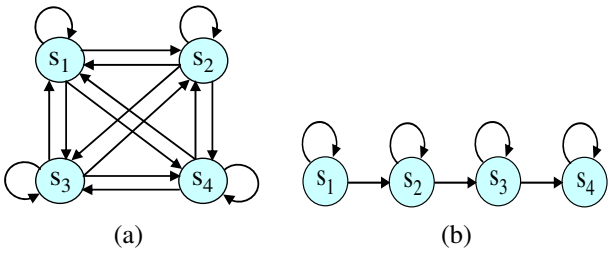

Fig. 9. HMM topologies with 4 states (a) Ergodic topology (b) LRB topology.

2) Line segment of HMM states: The number of states is an important parameter because the excessive number of states can generate the over-fitting problem if the number of training samples is insufficient compared to the model parameters. When there are insufficient number of states, the discrimination power of the HMM is reduced, since more than one single should be modeled on one state. Moreover, the number of states in our gesture recognition system is based on the complexity of each gesture number (Fig. 10) and is determined by mapping each straight-line segment into a single HMM state (Fig. 11). In practice, we considered the LRB model with 5 states for the following reasons (Fig. 10, Table III). Since each state in Ergodic topology has many transitions rather than LR and LRB topologies, the structure data can be lost easily. On the other hand, LRB topology has no backward transition where the state index either increases or stays the same as time increases. In addition, LRB topology

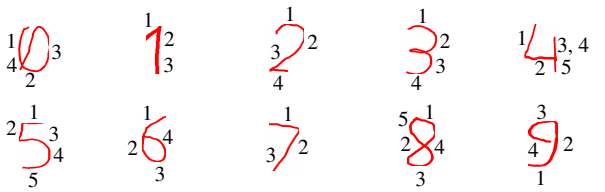

Fig. 10. Gesture path from hand motion trajectory for Arabic numbers with its segmented parts.

is more restricted rather than LR topology and simple for training data that will be able to match the data to the model. Also, the gesture paths for 4 and 5 contain the largest number of segmented part and to ensure that all these parts are used, we consider using 5 states.

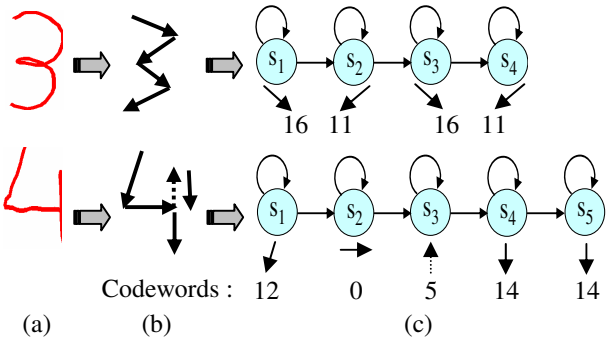

Fig. 11. Straight-line segment for HMM topologies (a) Gesture number from hand motion trajectory (b) Line segment of gesture number (c) LRB model with line segmented codewords.

3) Initializing parameters for LRB model: No one can deny that, a good parameters initialization for HMM (A, B, Π) achieves better results. Matrix A is the first parameter and is determined by Eq. 6. It depends on the duration time d of states for each number such that d is defined as;

$$d = \frac{T}{N} \tag{5}$$

where T is the length of gesture path and N represents the number of states that has a value 5 in our system.

$$A = \begin{pmatrix} a_{ii} & 1 - a_{ii} & 0 & 0 & 0 \\ 0 & a_{ii} & 1 - a_{ii} & 0 & 0 \\ 0 & 0 & a_{ii} & 1 - a_{ii} & 0 \\ 0 & 0 & 0 & a_{ii} & 1 - a_{ii} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{6}$$

Such that;

$$a_{ii} = 1 - \frac{1}{d} \tag{7}$$

The second important parameter is matrix B that is determined by Eq. 8. Since HMM states are discrete, all elements of matrix B can be initialized with the same value for all different states.

$$B = \{b_{im}\}; \qquad b_{im} = \frac{1}{M} \tag{8}$$

where *i*, *m* run over the number of states and the number of discrete symbols respectively. The third HMM parameter is the initial vector Π which takes value;

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}^T \tag{9}$$

That is because we use 5 states as the maximum numbers of the segmented part and in order to ensure that it begin from the first state as shown in Fig. 13.

For the continuous gesture, our system is designed to recognize the meaningful gesture by zero-codeword detection (Fig. 7(b), Fig. 12). Each gesture ends by line segment, which is assigned a 0-codeword. There are many gestures (i.e. 2, 4 and 7) which contain zero-codeword in some segments where these cause separation problems in the same gesture. To overcome this problem, we assign static velocity as a threshold. Furthermore, between the two gestures, there are links that must be ignored and is done by neglecting some frames adaptively after detecting the end point of gesture.

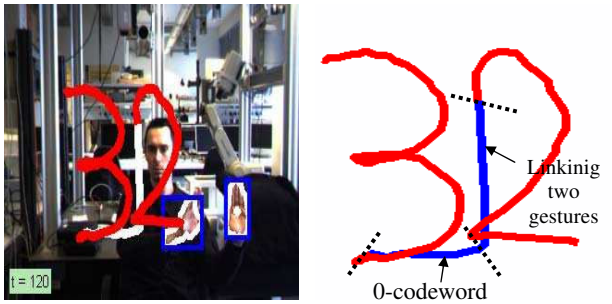

Fig. 12. Continuous gesture path for number 32 where 0-codeword detection refers to the end of first gesture, followed by adaptive link between two gestures.

4) Baum-Welch and Viterbi path: Our database contains 30 videos for each isolated gesture number from 0 to 9 (20 video sequences for training and 10 video sequences for testing) and 70 videos for continuous gestures. Baum-Welch algorithm plays very significant role in our system where it is used to do a full training for the initialized HMM parameters $\lambda = (\Pi, A, B)$. Our system is trained on 20 sequences of discrete vector for each isolated gesture number by using Ergodic, LR and LRB topologies with the number of states ranging from 3 to 10. After finished from the training process by computing the HMM parameters for each video sequence, the value of matrix *A* and matrix *B* for them is averaged. According to the forward algorithm with Viterbi path, the other 10 video sequences for each isolated gesture number are tested where this algorithm is built on discrete vector, matrix *A*, matrix *B* and vector Π as inputs for it. The forward algorithm computes the probability of the discrete vector sequences for all HMM topologies with different states. Thereby, the gesture path is recognized corresponding to the maximal likelihood of ten Gestures HMM models from 0 to 9 over the best path that is determined by Viterbi algorithm. The following steps demonstrate how the Viterbi algorithm works on LRB topology (Fig. 13).

1. Initialization: $for\ 1 \le i \le N$,
   a) $\delta_1(i) = \Pi_i.b_i(o_1)$
   b) $\phi_1(i) = 0$
2. Recursion: $for\ 2 \le t \le T,\ 1 \le j \le N$,
   a) $\delta_t(j) = \max_i[\delta_{t-1}(i).a_{ij}].b_j(o_t)$
   b) $\phi_t(j) = arg\max_i[\delta_{t-1}(i).a_{ij}]$
3. Termination:
   a) $p^* = \max_i[\delta_T(i)]$
   b) $q_T^* = arg\max_i[\delta_T(i)]$
4. Reconstruction: $for\ t = T-1, T-2, ..., 1$
   $q_t^* = \phi_{t+1}(q_{t+1}^*)$

The resulting trajectory (optimal states sequence) is $q_1^*, q_2^*, ..., q_T^*$ where a_{ij} is the transition probability from state S_i to state S_j, $b_j(o_t)$ is the probability of emitting o at time t in state S_j, $\delta_t(j)$ represents the maximum value of S_j at time *t*, $\phi_t(j)$ is the index of S_j at time *t* and p^* is the state optimized likelihood function.

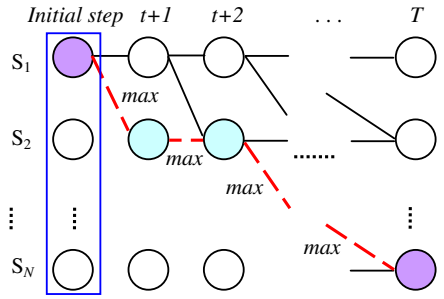

Fig. 13. The best path for LRB model with *N* states where it starts from S_1 to S_N, *N*=3, 4, ..., 10 and *t*=1.

III. Experimental Results

Our proposed system showed good results to recognize numbers in real-time from stereo color image sequences via the motion trajectory of a single hand using HMM. In our experimental results, each isolated gesture number from 0 to 9 was based on 30 video sequences, which 20 video samples for training and nearly 10 video samples for testing. In other words, our database contains 200 video sequences for training and 98 video sequences for testing isolated gestures. It also contains 70 video sequences for testing continuous gestures. The higher priority was computed by forward algorithm in conjunction with Viterbi path to recognize the numbers in real-time frame by frame. The system was implemented in Matlab language and the input images were captured by Bumblebee stereo camera system that has 6 mm focal length for about 2 to 5 second at 15 frames per second with 240 × 320 pixels image resolution.

A. Isolated gesture recognition

In our experiment, we designed a different HMM topologies with different states ranging from 3 to 10. From table III, the

average ratio of LRB topology from 3 to 10 states was 98.6%. Also, LR and LRB topologies with 3 and 4 states achieved the best recognition. In addition, LRB topology was always better than LR and Ergodic topologies. In general, LRB topology with 5 states is the best in terms of results empirically and this confirms what we have said theoretically and presented in subsection II.C.2. Fig. 14 shows the output of our system for isolated gesture number 3 where the higher priority was computed by forward algorithm in conjunction with Viterbi path. The following criteria evaluated our result as follows: The testing data is considered as, $\tau = 98$, for isolated Arabic gesture numbers from 0 to 9 with a specific HMM topology where these test data include valid gestures ν and also invalid gestures $\bar{\nu}$ such that;

$$\tau = \nu_j + \bar{\nu_j} \; ; \; j = 3, 4, ..., 10 \tag{10}$$

where j represents the number of states. The valid ratio for each specific HMM topology is calculated by Eq.11 and the average ratio for specific HMM topology with number of states ranging from 3 to 10 is determined by Eq. 12.

$$\eta_j = \frac{\nu_j}{\tau} \cdot 100 \tag{11}$$

$$\Re = \frac{1}{8} \sum_{i=3}^{10} \eta_j \tag{12}$$

where η_j is the result of HMM topology with a specific state number and $\Re$ represents the average ratio of HMM topology for all states number from 3 to 10.

TABLE III
ISOLATED GESTURE RECOGNITION RESULTS FOR HMM TOPOLOGIES WITH THE NUMBER OF STATES RANGING FROM 3 TO 10.

Number of states	Data		Topologies Recognition (%)		
	Train	Test	Ergodic	LR	LRB
3	200	98	70.40	100	100
4	200	98	58.16	100	100
5	200	98	65.31	98.98	100
6	200	98	59.18	96.94	96.94
7	200	98	45.92	96.94	97.96
8	200	98	46.94	94.90	96.94
9	200	98	53.06	94.90	97.96
10	200	98	44.90	96.94	98.98
Average	200	98	55.48	97.45	98.6

B. Meaningful gesture recognition

The continuous gestures include the meaningful gestures where these gestures are recognized by our idea zero-codeword detection with static velocity as a threshold. The zero-codeword detection refers to the end of isolated gesture and static velocity is used to overcome the problem that is related to isolated gesture with 0-codeword line segment (for example; 2, 4 and 7 gesture numbers). The threshold of static velocity that was used in our system is smaller than 54 pixels per second. Our system tested on 70 video sequences for continuous gestures where each video sequence contains two meaningful gestures within itself. The recognition was achieved on continuous gestures 94.29% (Fig. 15).

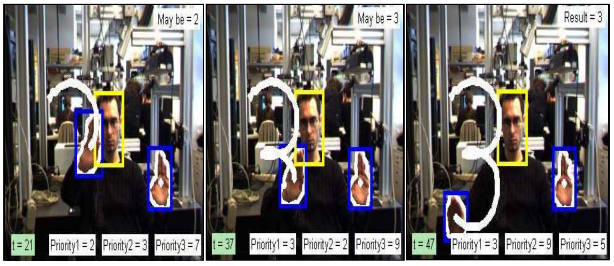

Fig. 14. System output for isolated gesture number 3, where at t=21 the high priority is number 2, at t=37 the high priority is number 3 and at t=47 the result is 3.

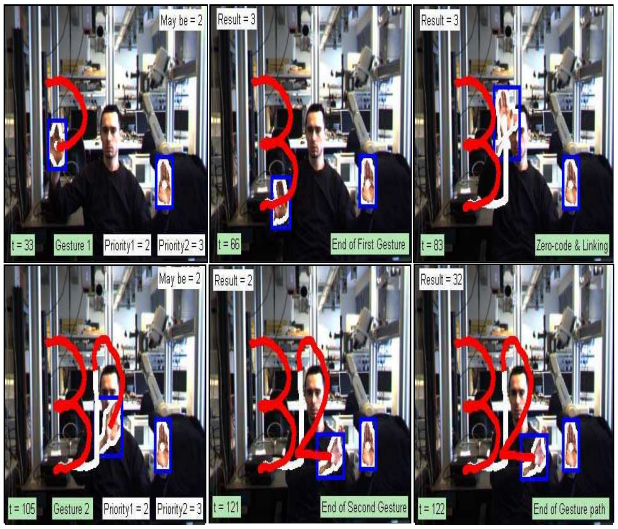

Fig. 15. System output for continuous gesture 32, where at t=66 the first gesture is ended with result 3, at t=83 the linking between two gestures, at t=121 the second gesture is ended with result 2 and at t=122 the final result is 32.

IV. SUMMARY AND CONCLUSION

This paper proposes a system to recognize both isolated and meaningful gestures for Arabic numbers from stereo color image sequences by the motion trajectory of a single hand using HMM which is suitable for real-time application. The system consists of three main stages; automatic segmentation and preprocessing, feature extraction, and classification. The system depends on our idea of zero-codeword detection with static velocity motion to recognize the meaningful gestures from continuous gestures. Our database contains 30 video sequences for each isolated gesture number from 0 to 9 (20 video sequences for training and 10 video sequences for testing) and 70 video sequences for continuous gestures. For isolated gestures HMM using Ergodic, LR, and LRB topologies with the number of states ranging from 3 to 10 are applied and tested. The LRB topology with 5 states in conjunction with

BW algorithm for training and forward algorithm with Viterbi path for testing presents the best performance. Our results show that; an average recognition rate is 98.6% and 94.29% for isolated and meaningful gestures respectively. In future, our research focuses on the motion trajectory, which will be determined by a fingertip instead of the centroid point for the hand region. Also, the research will be carried out to recognize the American Sign Language words over combined features.

ACKNOWLEDGMENT

This work was supported by Bernstein-Group (BMBF: 01GQ0702) and NIMITEK grants (LSA: XN3621E/1005M).

REFERENCES

[1] X. Deyou, *A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG*, In International Conference on Pattern Recognition (ICPR 06), pp. 519-522, 2006.
[2] D. B. Nguyen, S. Enokida, and E. Toshiaki, *Real-Time Hand Tracking and Gesture Recognition System*, IGVIP05 Conference, CICC, pp. 362-368, 2005.
[3] N. P. Vassilia, and G. M. Konstantinos, *On Feature Extraction and Sign Recognition for Greek Sign Language*, Proceedings of the 7th IASTED International Conference Artificial Intelligence and Soft Computer, pp. 93-98, 2003.
[4] 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 W S C G'08, Vol. 16(1), pp. 65-72, 2008.
[5] E. Holden, R. Owens, and G. Roy, *Hand Movement Classification Using Adaptive Fuzzy Expert System*, Journal of expert Systems Research and Application, Vol. 9(4), pp. 465-480, 1996.
[6] L. Nianjun, C. L. Brian, J. K. Peter, and A. D. Richard, *Model Structure Selection & Training Algorithms for a HMM Gesture Recognition System*, In International Workshop in Frontiers of Handwriting Recognition, pp. 100-106, 2004.
[7] H. Lee, and J. Kim, *An HMM-Based Threshold Model Approach for Gesture Recognition*, IEEE Trans.on Pattern Analysis and Machine Intelligence, Vol. 21(10), pp. 961-973, 1999.
[8] Y. Ming-Hsuan, and A. Narendra, *Gaussian Mixture Modeling of Human Skin Color and Its Applications in Image and Video Databases*, In the SPIE/EI&T Storage and Retrieval for Image and Video Databases, pp. 458-466, 1999.
[9] S. Askar, Y. Kondratyuk, K. Elazouzi, P. Kauff, and O. Schreer, *Vision-Based Skin-Colour Segmentation of Moving Hands for Real-Time Application*, 1st European CVMP, pp. 79-85, 2004.
[10] S. L. Phung, A. Bouzerdoum, and D. Chai, *A Novel Skin Color Model in YC_bC_r Color Space and its Application to Human Face Detection*, In IEEE International Conference on Image Processing (ICIP), pp. 289-292, 2002.
[11] Y. Raja, S. J. Mckenna, and S. Gong, *Colour Model Selection and Adaptation in Dynamic Scenes*, In Proceedings European Conference on Computer Vision, pp. 460-474, 1998.
[12] A. R. Richard, and F. W. Homer, *Mixture Densities, Maximum Likelihood and the EM Algorithm*, SIAM Review, Vol. 26(2), pp. 195-239, 1984.
[13] R. Niese, A. Al-Hamadi, and B. Michaelis, *A Stereo and Color-based Method for Face Pose Estimation and Facial Feature Extraction*, The IEEE 18th International Conference on Pattern Recognition, Vol. 1, pp. 299- 302, 2006.
[14] R. Niese, A. Al-Hamadi, and B. Michaelis, *A Novel Method for 3D Face Detection and Normalization*, Journal of Multimedia, Vol. 2(5), pp. 1-12, 2007.
[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] H. Yoon, J. Soh, Y. J. Bae, and H. S. Yang, *Hand Gesture Recognition using Combined Features of Location, Angle and Velocity*, Pattern Recognition 34(7), pp. 1491-1501, 2001.
[18] M. Elmezain, A. Al-Hamadi, 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.

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. Since 2006 he has been Junior-Research-Group-Leader at the Institute for Electronics, Signal Processing and Communications 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. Dr. Al-Hamadi is the author of more than 70 articles.

Jörg Appenrodt was born in Magdeburg, Germany in 1982. He received his Masters Degree in Electronic Engineering at the Otto-von-Guericke University of Magdeburg in 2008. His current work concentrates on the field of image processing, pattern recognition and human-computer interaction, at the Otto-von-Guericke University of Magdeburg, Germany.

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.