TRAJECTORY MODELING IN GESTURE RECOGNITION USING CYBERGLOVES® AND MAGNETIC TRACKERS

Ng Yong Yi Kevin[1] *S. Ranganath*[1] *D. Ghosh*[2]

[1]Department of Electrical & Computer Engineering
National University of Singapore, Singapore

[2]Department of Electronics & Communication Engineering
Indian Institute of Technology Guwahati, India

ABSTRACT

The recognition of human gestures is important for several human-computer interaction applications. In this paper, we develop a gesture recognition system that uses the condensation-based trajectory matching/recognition algorithm. The gesture data are collected using a pair of CyberGloves® measuring hand-joint angles and three magnetic trackers that determine 3-D hand positions. The multi-dimensional gesture data are subsequently recognized by matching against trajectory models using probability measures. In our experiments, we evaluate the efficiency of our proposed gesture recognition system using three different gesture sets, viz., directional movements, static hand-shapes and American Sign Language (ASL) gestures. Experimental results show high recognition rate and signer-independence but less robustness to co-articulation effects.

1. INTRODUCTION

Gestures are physical positions or movements of a person's fingers, hands, arms or body used to convey information. Gestures are used in sign language communication which forms the most natural means of information exchange between deaf people. Gestures also find application in the field of virtual reality – they make human-computer interaction more natural, efficient and flexible for users, as compared to the simple point and click paradigm [1]. Therefore, a system capable of recognizing gestures efficiently is desired for reliable human-computer interaction as well as for sign language interpretation. Since most natural gestures involve hand movement, recognition of hand gestures has been an important subject of research in this field [2–6]. In this paper, we develop a hand gesture recognition system wherein the gesture data are collected using a pair of CyberGloves® and three magnetic trackers and are then analyzed using the condensation algorithm which is also used by Black and Jepson [7] for analyzing gestures acquired by a camera to control a whiteboard.

The 'Condensation'(Conditional density propagation) algorithm was first proposed by Isard and Blake [8] to track objects in clutter and has been extended to gesture recognition in [9]. The algorithm uses a probabilistic framework to incrementally match the trajectory models to the multi dimensional input data varying in time. The condensation-based trajectory modeling can be visualized as a generalization of Hidden Markov Models (HMMs) as it allows a discrete set of states with probabilistic transitions between states. However, the difference is that the recognition of each individual state involves the probabilistic matching of entire temporal trajectory model representing a portion of a gesture.

Condensation-based trajectory recognition in [7] allows recognition and motion tracking in the same probabilistic framework. This is advantageous as it solves the spatial and temporal issue of gesture recognition simultaneously. A brief discussion on this condensation-based trajectory modeling algorithm is given in Section 2. Our proposed feature extraction method using CyberGloves® and magnetic trackers is described in Section 3 followed by the recognition scheme in Section 4. Experimental results are presented in Section 5. Finally, conclusions are given in Section 6.

2. CONDENSATION-BASED TRAJECTORY MODELING

Essentially, the condensation algorithm consists of four distinct high-level building blocks as follows.

2.1. Initialization & Selection

Given that there are M gestures to be modeled, the training data are translated into M model trajectory vectors each consisting of N feature trajectories $m_i^{(\mu)}$, $i = 1,....,N$ and $\mu = 1,.....,M$. Each of these trajectories corresponds to a single gesture feature, e.g., the hand-joint angle for the thumb, at a particular instant of time. The search space is first initialized by choosing S sample states – each state is represented by the dynamic time warping (DTW) of one among the M model trajectory vectors. DTW of model vector μ is obtained by distorting each of its feature trajectories by phase, amplitude and rate adjustment parameters ϕ, α and ρ, given as $\alpha.m^{(\mu)}_{(\phi-\rho j),i}$. Thus, a state s_t at time t is defined by the 4-tuple $\{\mu, \phi, \alpha, \rho\}$.

0-7803-8560-8/04/$20.00©2004IEEE

Authorized licensed use limited to: COLORADO STATE UNIVERSITY. Downloaded on May 05,2022 at 20:49:24 UTC from IEEE Xplore. Restrictions apply.

The trajectory modeling algorithm seeks to find the most likely state s_t that gives the best match for the N-dimensional trajectory vector $\mathbf{Z}_t = (Z_{t,1},.....Z_{t,N})$ observed at time t. To find likelihood of each state, DTW is performed on the model data in accordance with the state parameters and the likelihood is calculated as the probability of $\mathbf{Z}_t$ for given state s_t, as follows:

$$P(\mathbf{Z}_t / s_t) = \prod_{i=0}^{N} P(Z_{t,i} / s_t) \quad (1\text{-a})$$

$$P(Z_{t,i} / s_t) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp \frac{-\sum_{j=0}^{w-1} \left(Z_{(t-j),i} - \alpha . m^{(\mu)}_{(\phi-\rho j),i}\right)^2}{2\sigma_i (w-1)} \quad (1\text{-b})$$

where w is the size of the matching temporal window from t backwards to $(t-w-1)$ and σ_i is the estimated standard deviation for i^{th} trajectory. The probability distribution of the whole search space at one time instant t is now obtained by calculating the likelihoods for all the S samples. Each conditional probability acts as a weighting factor for its corresponding state and with successive iterations, the distribution in the search space clusters around areas that represent the more likely gesture states.

2.2. Prediction

It is possible to predict the probability distribution over the search space at the next time instant by computing the normalized weighted probabilities at that instant as

$$\pi_t = \frac{P(\mathbf{Z}_t / s_t)}{\sum_{s_t \in \mathbf{S}} P(\mathbf{Z}_t / s_t)} \quad (2)$$

where **S** is the search space with S sample states.

Predicting the probability distribution at the next time instant is equivalent to propagating the more probable states over time in the observed gesture sequence. It is to be noted that more than one probable state can be propogated at each time instant. The parameters for each sample of the S sample set are initialized by uniform sampling within their ranges as follows

$$\mu = [0, \mu_{\max}] \quad (3\text{-a})$$

$$\phi = \frac{1-\sqrt{y}}{y} \quad \text{where} \quad y \in [0,1] \quad (3\text{-b})$$

$$\alpha = [\alpha_{\min}, \alpha_{\max}] \quad (3\text{-c})$$

$$\rho = [\rho_{\min}, \rho_{\max}] \quad (3\text{-d})$$

Once the samples are initialized, a cumulative probability distribution is constructed using the weighted probabilities obtained using eqn. 2 at time $(t-1)$, where $(t-1)$ represents the current time frame and t refers to the next time frame that is being predicted. A value of r is chosen uniformly and then the smallest value of the cumulative weight c_{t-1} is chosen such that $c_{t-1} > r$. The corresponding state s_t is then selected for propagation. This is illustrated in Fig. 1. With this method of selection, the larger weights are more likely to be chosen.

After the selection of propagation states, the parameters for these states at the next time instant are predicted as

$$\mu_t = \mu_{t-1} \quad (4\text{-a})$$

$$\phi_t = \phi_{t-1} + \rho + \mathrm{N}(\sigma_\phi) \quad (4\text{-b})$$

$$\alpha_t = \alpha_{t-1} + \mathrm{N}(\sigma_\alpha) \quad 4\text{-c})$$

$$\rho_t = \rho_{t-1} + \mathrm{N}(\sigma_\rho) \quad (4\text{-d})$$

where N(.) represents normal distribution and σ_* is the uncertainty in prediction.

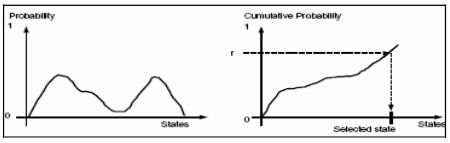

Fig. 1: Construction of cumulative probability table.

2.3. Updating

After the new state is predicted, the probability $P(\mathbf{Z}_t / s_t)$ is computed. If the conditional probability falls below 0, the state is predicted again and the probability is recalculated. If this keeps repeating for a pre-set number of times, the state is deemed unlikely and it is re-initialized. Once all the new states are generated, the weights π_t are recalculated for state selection and propagation at the subsequent time instants. This process of selection, prediction and propagation is repeated until the observed sequence ends or the criterion for recognition is met.

3. FEATURE EXTRACTION METHOD

The proposed feature extraction method utilizes a pair of Immersion Corp CyberGloves® and three Polhemus FASTRAK magnetic trackers. The gloves measure 18 hand-joint angles while the trackers provide the positional and orientation information. Two trackers are positioned on the wrist of each hand while another on the chest serves as the reference tracker.

A data extraction program captures both synchronized video and gesture data. Subsequently, a synchronizer program is used to accurately match each frame of gesture data to the video data. Using the synchronizer, as shown in Fig. 2, a search window can be accurately gauged allowing the recognition system to identify the individual gestures in a continuous stream of gestures. This facilitates the manual segmentation of the essential gesture data from the gesture sequence while eliminating co-articulation. Data from the trackers are preprocessed to obtain the relative positions of the hands. In our system, a total of 48 data values are obtained from the CyberGloves® and the magnetic trackers, thus, forming a 48-dimensional feature vector.

Authorized licensed use limited to: COLORADO STATE UNIVERSITY. Downloaded on May 05,2022 at 20:49:24 UTC from IEEE Xplore. Restrictions apply.

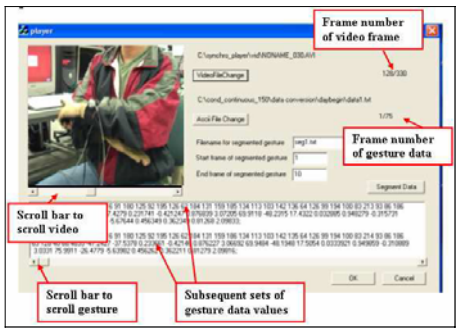

Fig 2: GUI interface of gesture and video data synchronizer

4. RECOGNITION METHOD

In our system, gestures are recognized using the configuration shown in Fig. 3 below.

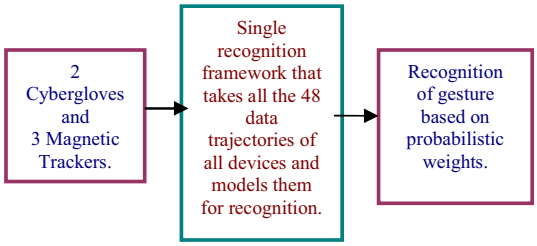

Fig. 3: Configuration of the recognition system

For each observation data vector, a set of probabilistic weights is generated for each time sample based on the set of gestures in the database of the recognition framework. In isolated gesture recognition, the trajectories of the different parameters are captured for certain duration of time so as to obtain a fixed number of time samples. The observed trajectory vector at each sampling instant is matched against the database gestures and the probabilistic weights are generated, as depicted in Fig. 4. The input is recognized as the gesture in the database that gives the maximum average weight.

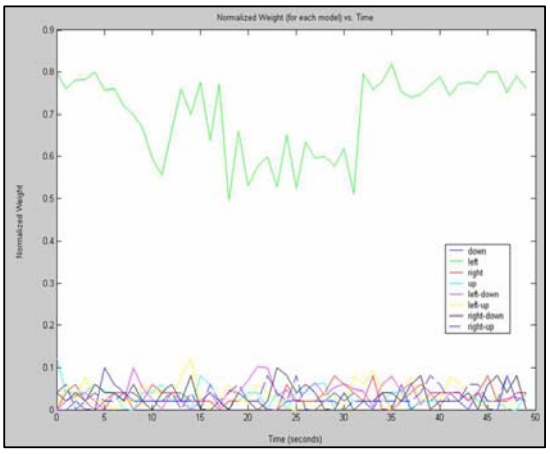

Fig.4: Graph of Normalized weights vs time samples for eight gestures

In continuous gesture recognition, a generalized search window as derived from the video and gesture data synchronizer is used to search for the individual gestures. These individual observed gestures are then matched against the database gestures. However, an additional check requires the weights for a particular gesture to be consecutively above a pre-defined threshold value for reasonable number of time samples within the continuous stream of gesture data. This is to eliminate the possibility of recognizing co-articulation effects as valid gestures.

5. EXPERIMENTAL RESULTS

In our experiments, positional recognition was investigated through a gesture set of 17 directional changes of the hand which revealed the ability to differentiate between directions of close proximity. With respect to hand-shape recognition, the subtleness of recognition was investigated for 21 hand shape gestures. The recognition ability of the system in response to practical sign languages was ascertained with a database of 20 isolated ASL gestures. Finally, 10 continuous gesture sequences were created by combining some of the above 20 ASL gestures to study the influence of co-articulation effects.

Two sets of experiments were carried out: One in which the test data are taken from the set of training data (TD) and the other in which the test data are unseen by the system (UD). Recognition rates in cases of isolated gestures are given in Table 1 (directional movement and hand shape gestures) and Table 2 (ASL). For directional movement gestures, the average accuracy is 89.6% and 86.1% for training and unseen data, respectively. For static hand shape, the accuracy obtained on training and unseen data is 80.6% and 77.4%, respectively. It is observed that as the proximity between directional movement gestures (in terms of 3-D direction with space) was decreased or the hand-shapes were very similar to each other, it became relatively more difficult to recognize the gestures. For isolated ASL gestures, the accuracy is 91.6% and 88.5% using training and unseen data, respectively. However, for continuous gestures the recognition result is variable as shown in Table 3. This is due to the co-articulation effects which affects system performance.

6. CONCLUSION

A gesture recognition framework is presented in this paper. Our work here is confined to hand gesture recognition. The system utilizes condensation-based algorithm for gesture modeling. The positional movements of the hands are tracked by using CyberGloves® and magnetic trackers. The information so obtained forms the set of feature trajectories for the gesture. Experimental results show high recognition rate in case of isolated gestures but the recognition is affected in continuous gestures due to co-articulation effect. Therefore, while the system in its present form is good for

Authorized licensed use limited to: COLORADO STATE UNIVERSITY. Downloaded on May 05,2022 at 20:49:24 UTC from IEEE Xplore. Restrictions apply.

isolated gesture recognition, it requires to be improved by making it robust to co-articulation effects for reliable recognition of continuous gestures.

TABLE 1: PERCENTAGE GESTURE RECOGNITION RATES

Directional Movement Gestures			*Static Hand Shape Gestures*		
Gesture	*TD*	*UD*	*Gesture*	*TD*	*UD*
Down	96.1	93.0	One	84.3	80.5
Left	94.7	92.5	Two	85.4	82.0
Right	95.3	91.7	Three	87.9	84.8
Up	94.2	91.7	Four	57.0	52.5
Front	95.7	91.2	Five	94.8	91.9
Left-Down	86.8	84.7	Six	72.9	70.8
Left-Up	90.0	86.9	Seven	91.8	88.3
Right-Down	89.3	85.8	Eight	92.6	90.6
Right-Up	89.6	85.1	Nine	48.8	47.4
Front-up	89.1	86.6	A	94.7	91.2
Front-down	89.3	85.3	B	55.8	50.0
Front-left	88.1	84.5	C	89.3	86.2
Front-right	90.9	86.3	D	82.8	80.3
Front-left-down	83.0	76.8	E	93.9	90.2
Front-right-down	83.4	80.3	F	58.9	54.0
Front-left-up	83.2	82.8	H	83.9	81.4
Front-right-up	84.3	77.8	I	90.8	88.5
Average	**89.6**	**86.1**	M	77.7	74.2
			R	88.5	86.1
			W	64.2	61.8
			Y	96.4	93.4
			Average	**80.6**	**77.4**

TABLE 2: PERCENTAGE RECOGNITION RATES FOR ISOLATED ASL GESTURES

Gesture	***TD***	***UD***
Sit	92.5	90.7
Agree	93.6	90.5
Day	91.1	90.2
Down	91.3	89.5
Interpret	92.3	89.7
Abortion	0	0
Book	93.7	90.2
Awake	90.1	88.5
Before	0	0
Begin	91.9	88.1
Buy	92.4	87.8
Read	88.7	84.2
Baby	93.2	89.2
Assume	84.2	80.2
Big large	5.9	0
Big tall	92.8	87.9
Beautiful	39.5	10.2
Good	91.6	87.8
Mother	93.3	90.1
Father	92.6	90.6
Average	**91.6**	**88.5**

TABLE 3: PERCENTAGE RECOGNITION RATES FOR ISOLATED ASL GESTURES

Gesture Sequence	*Gesture 1 of Sequence*		*Gesture 2 of Sequence*	
	TD	*UD*	*TD*	*UD*
baby-book	79.6	62.4	88.1	82.1
day-begin	68.9	51.4	32.0	12.5
good-day	34.5	11.7	89.2	81.0
Interpret-good	90.6	84.3	67.6	43.8
sit-down	46.9	26.7	37.9	12.9
read-book	56.1	43.1	91.3	83.6
mother-awake	38.7	23.8	84.9	67.9
bigtall-father	65.4	45.3	28.7	19.0
father-read	56.1	33.8	82.5	56.4
begin-assume	39.6	12.7	45.6	23.4

12. REFERENCES

[1] J. Weissmann and R. Salomon, "Gesture recognition for virtual reality applications using data glove and neural networks," *Proc. IEEE Int'l Joint Conf. Neural Net.* (*IJCNN'99*), vol. 3, pp. 2043 -2046, 1999.

[2] H. Hongo, M. Ohya, M. Yasumoto and K. Yamamoto, "Face and hand gesture recognition for human-computer interaction," *Proc. 15th IEEE Int'l Conf. Pattern Recog.* (*ICPR 2000*), vol. 2, pp. 921-924, 2000.

[3] C.R.P. Dionisio and R.M. Cesar, Jr., "A project for hand gesture recognition," *Proc. XIII Brazilian Symp. Computer Graphics and Image Process*., pg. 345, 2000.

[4] L.K. Lee, S. Kim, Y.-K. and M.H. Lee, "Recognition of hand gesture to human-computer interaction," *Proc. 26th Annual Conf. IEEE Industrial Electronics Society* (*IECON 2000*), vol. 3, pp. 2117-2122, 2000.

[5] T. Kapuscinski and M. Wysocki, "Hand gesture recognition for man-machine interaction," *Proc. 2nd IEEE Int'l Workshop Robot Motion and Control*, pp. 91-96, 2001.

[6] X. Yin and M. Xie, "Hand gesture segmentation, recognition and application," *Proc. IEEE Int'l Symp. Computational Intelligence in Robotics and Automation*, pp. 438-443, 2001.

[7] M.J. Black and A.D. Jepson, "Recognizing temporal trajectories using the condensation algorithm," *Proc. 3rd IEEE Int'l Conf. Automatic Face and Gesture Recog.*, pp. 16-21, 1998.

[8] M. Isard and A Blake, "Contour tracking by stochastic propagation of conditional density," *Proc. European Conf. Computer Vis.*, vol. 1, pp. 343-356, 1996.

[9] M. Isard and A. Blake, "A mixed-state condensation tracker with automatic model-switching," *Proc. 6th IEEE Int'l Conf. Computer Vis.*, pp. 107-112, 1998.

Authorized licensed use limited to: COLORADO STATE UNIVERSITY. Downloaded on May 05,2022 at 20:49:24 UTC from IEEE Xplore. Restrictions apply.