Dynamic Gesture Recognition for Human Robot Interaction

Javier Ruiz-del-Solar, Rodrigo Verschae, Jong Lee-Ferng, Mauricio Correa

Abstract— **In this article a robust and real-time dynamic hand gesture recognition system meant to allow a natural interaction with a service robot, in dynamic environments, is proposed. The main novelty of the proposed approach is the use of temporal statistics about the hand's positions and velocities as basic information to recognize the gestures. The use of these features allows carrying out the final recognition using a standard Bayes classifier, instead of the traditional Hidden Markov Models. The gesture segmentation and recognition is achieved simultaneously by finding gesture's candidate subsequences that give high scores when matched to a gesture. The system uses boosted classifiers to detect hands, and the mean-shift algorithm for their tracking. The system performance is validated in a digit recognition system database and real-world video sequences.**

Index Terms— **dynamic hand gesture recognition, human robot interaction, RoboCup @Home.**

I. Introduction

Hand gestures are extensively employed in human non-verbal communication. They allow to express orders (e.g. "stop"), mood state (e.g. "victory" gesture), or to transmit some basic cardinal information (e.g. "two"). In addition, in some special situations they can be the only way of communicating, as in the cases of deaf people (sign language) and police's traffic coordination in the absence of traffic lights.

Thus, it seems convenient that human-robot interfaces incorporate hand gesture recognition capabilities. For instance, we would like to have the possibility of transmitting simple orders to personal robots using hand gestures. The recognition of hand gestures requires both hand's detection and gesture's recognition. Both tasks are very challenging, mainly due to the variability of the possible hand gestures (signs), and because hands are complex, deformable objects (a hand has more than 25 degrees of freedom, considering fingers, wrist and elbow joints) that are very difficult to detect in dynamic environments with cluttered backgrounds and variable illumination.

Several hand detection and hand gesture recognition systems have been proposed. Early systems usually require markers or colored gloves to make the recognition easier. Second generation methods use low-level features as color (skin detection) [4][5], shape [8] or depth information [2] for detecting the hands. However, those systems are not robust enough for dealing with dynamic conditions; they usually require uniform background, uniform illumination, a single person in the camera view [2], and/or a single, large and centered hand in the camera view [5]. Boosted classifiers allow the robust and fast detection of hands [3][6][7]. In addition, the same kind of classifiers can be employed for detecting static gestures [7]; dynamic gestures are normally analyzed using Hidden Markov Models [4][16]. 3D hand model-based approaches allow the accurate modeling of hand movement and shapes, but they are time-consuming and computationally expensive [6][7].

In this context, we are proposing a robust and real-time hand gesture recognition system to be used in the interaction with personal robots. We are especially interested in dynamic environments such as the ones defined in the *RoboCup @Home league* [15], with the following characteristics: variable illumination, cluttered backgrounds, (near) real-time operation, large variability of hands' pose and scale, and limited number of gestures (they are used for giving the robot some basic orders). It is important to mention that in the new RoboCup @Home league' rules gesture recognition is emphasized: *An aim of the competition is to foster natural interaction with the robot using speech and gesture commands* (2009's Rules book, pp. 7, available in [15]). We would like to build a system that fulfills the basic requirement of the league, which basically consists of recognition of static gestures, but that also allows a richer interaction with the robots using dynamic gestures. Using such a system it will be possible to give the robot basic cardinal information as well as complex orders.

Manuscript received July 15, 2009. This work was supported in part by CONICYT (Chile) under Grant FONDECYT 1090250. Full names of authors are preferred in the author field, but are not required. Put a space between authors' initials.

J. Ruiz-del-Solar is with the Elec. Eng. Dept. - Universidad de Chile, and with Mining Technology Center - Universidad de Chile (e-mail: jruizd@ing.uchile.cl).

R. Verschae is with the Elec. Eng. Dept. - Universidad de Chile (e-mail: rodrigo@verschae.org).

J. Lee-Ferng is with the Elec. Eng. Dept. - Universidad de Chile (e-mail: jongbor@gmail.com).

M. Correa is with the Elec. Eng. Dept. - Universidad de Chile (e-mail: mauricio.knight@gmail.com).

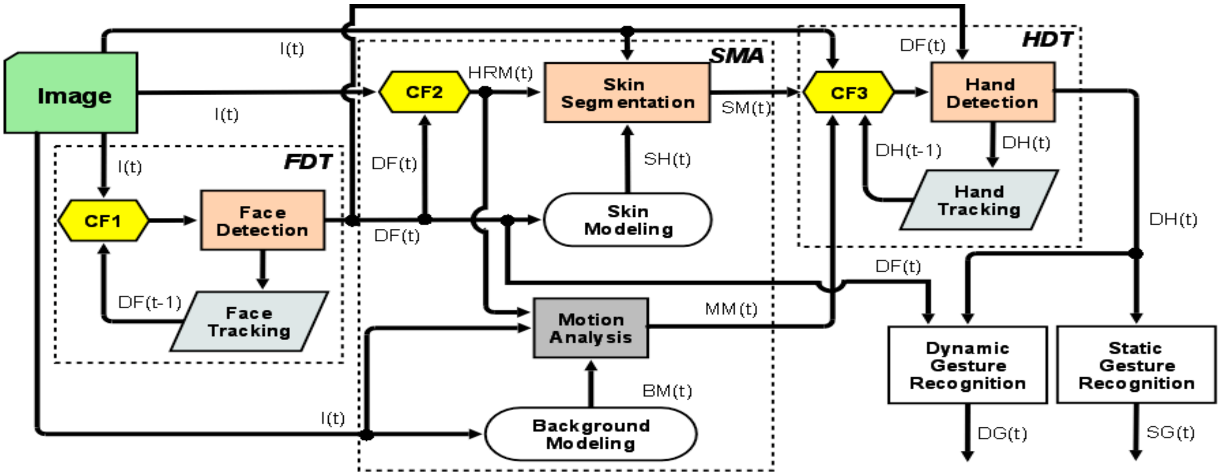

Fig. 1. Proposed hand gesture recognition system. CT*i*: *Context Filter i*. I: Image. DF: *Detected Face*. HRM: *Hand Region Mask*. SH: *Skin Histogram*. SM: *Skin Mask*: BM: *Background Model*. MM: *Motion Mask*. DH: *Detected Hand*. DG: *Dynamic Gesture*. SG: *Static Gesture*. **t**: *Frame index*. See main text for a detailed explanation.

The proposed system is able to recognize static and dynamic gestures, and its most innovative features include:

- The use of context information to achieve, at the same time, robustness and real-time operation, even when using a low-end processing unit (standard notebook), as in the case of humanoid robots. The use of context allows adapting continuously the skin model used in the detection of hand candidates, to restrict the image's regions that need to be analyzed, and to cut down the number of scales that need to be considered in the hand-searching and gesture recognition processes.
- The employment of boosted classifiers for the detection of faces and hands, as well as the recognition of static gestures. The main novelty is in the use of innovative training techniques - active learning and bootstrap -, which allow obtaining a much better performance than similar boosting-based systems, in terms of detection rate, number of false positives and processing time.
- The use of temporal statistics about the hand's positions and velocities and a Bayes classifier to recognize dynamic gestures. This approach is different from the traditional ones, based on Hidden Markov Models, which are not able to achieve real-time operation.

This article is focused on the description of the dynamic gesture recognition approach. The recognition of static gestures and the use of context to assist the gesture recognition processes are described in [10] and [11], respectively. In sections 2 some related work to dynamic gesture recognition is presented. In section 3 an overview of the whole gesture recognition system is presented. The dynamic gesture recognition approach is described in section 4. Results of the application of this approach in real video sequences are presented and analyzed in section 5. Finally, some conclusions of this work are given in section 6.

II. Related Work

The proposed approach relies on accurate hand detection and tracking as well as face detection and tracking. Face's position and size are needed in order to gain invariance to translation and scale. The hand and face detection module is based on Viola and Jones' cascade classifiers [13], while tracking is achieved with the mean-shift technique. Details can be found in a previous work [10].

The identification of the hand shape may or may not be relevant to the purposes of a dynamic gesture recognition system. In *Sign Language* recognition systems, both hand trajectory and hand shape are relevant. In this work, we restrict the problem to identifying gestures related only to the hand trajectory.

Following [16], we use simple features such as hand position and hand velocity in order to represent the hands detected in each frame. The main difference among dynamic gesture recognition systems lies in the way frame subsequences are treated. Many of the techniques rely upon matching observed sequences with known patterns, which are composed of various "stages". Observed frames may be matched more than once with a stage, or a stage may be skipped, in order to handle the fact that the same gesture may be performed with different durations. Dynamic Time Warping [20], Hidden Markov Models (HMM) [27], Continuous Dynamic Programming (CDP) [16] and Conditional Random Fields [29] are known instances of this approach.

HMMs in particular have become the predominant approach in dynamic gesture recognition systems. This technique was first used in the speech recognition community, where it has attained good results. The key advantages of HMMs are a rich mathematical structure, coupled with well known algorithms for training and evaluation of models.

Our approach is based on a quite different strategy, which involves computing overall geometric and kinematics information that is independent of the length of the performed gesture. Our approach can be contrasted with several other methods. Appearance-based methods [19] represent each detected hand directly as a raw set of pixels, normalized to a certain size, and a sequence of the last such hands is fed to a classifier, which outputs the recognized gesture. This kind of approach intends to statistically derive the most discriminant features, instead of relying on human defined ones, while our approach explicitly defines these features. Other strategies to deal with time variability include Time Delay Neural Networks [31], Motion History Images [18][28], which avoid explicit temporal analysis by transforming spatiotemporal motion into images, and bag of words approaches [32], which detect local salient features of movement, regardless of the spatiotemporal characteristics.

In online recognition, the start and end points of a gesture are unknown and determining them is an important problem called gesture segmentation. The simplest approach to gesture segmentation involves using low-level information such as hand velocity, acceleration and angle variations [30][3][17][33]. Some works propose the use of explicit gesture completion indicators, like putting the hand out of the camera range or performing a predefined gesture [26]. Some HMM-based methods find candidate boundary points by applying a threshold on the probability of a gesture [21][24] or by explicitly modeling a "garbage model" or "filling model", which matches the meaningless hand movement occurring between gestures [23][25][22]. In our system, gesture segmentation and recognition are achieved simultaneously by finding candidate subsequences that give high scores when matched to a gesture, while low level information --such as hand velocity and undetected (probably out of range) hands-- is also used to detect boundaries.

III. Hand Gesture Recognition System: System Overview

The whole hand gesture recognition system consists of five main modules Face Detection and Tracking (FDT), Skin Segmentation and Motion Analysis (SMA), Hand Detection and Tracking (HDT), Static Gesture Recognition, and Dynamic Gesture Recognition (see figure 1).

The FDT module is in charge of detecting and tracking faces. These functionalities are implemented using boosted statistical classifiers [12], and the *mean shift* algorithm [1], respectively. The information about the detected face (DF) is used as context in the SMA and HDT modules. Internally the CF1 (Context Filter 1) module determines the image area that need to be analyzed in the current frame for face detection, using the information about the detected faces in the past frame.

The SMA module determines candidate hand regions to be analyzed by the HDT module. The Skin Segmentation module uses a skin model that is adapted using information about the face-area's pixels (skin pixels). The module is implemented using the *skindiff* algorithm [9]. The Motion Analysis module is based on the well-known background subtraction technique. CF2 (Context Filter 2) uses information about the detected face and the human-body dimensions to determine the image area (HRM: Hand Region Mask) where a hand can be present in the image. Only this area is analyzed by the *Skin Segmentation* and *Motion Analysis* modules.

The HDT module is in charge of detecting and tracking hands. These functionalities are implemented using boosted statistical classifiers and the *mean shift* algorithm, respectively. CF3 (Context Filter 3) determines the image area where a hand can be detected in the image, using the following information sources: (i) skin mask (SM) which corresponds to a skin probability mask, (ii) motion mask (MM) that contains the motion pixels, and (iii) information about the hands detected in the last frame (DH: Detected Hand).

The Static Gesture Recognition module is in charge of recognizing static gestures. The module is implemented using statistical classifiers: a boosted classifier for each gesture class, and a multi-class classifier (C4.5 pruned tree [14]) for taking the final decision. The Dynamic Gesture Recognition module spots and recognizes dynamic gestures. This module computes temporal statistics about the hand's positions and velocities, features that are feed a Bayes classifier that recognizes the gesture.

IV. Dynamic Gesture Recognition

Multiple dynamic gestures are recognized (classified) using standard statistical classifiers. Considering that a given dynamic gesture is composed by a sequence of hand's positions and its corresponding dynamics, feature vectors that characterize both, positions and dynamics are defined. Gesture segmentation (i.e., determination of the gesture start and end) and classifcation is carried out simultanaously, by analyzing a so called gestures-table, that keeps the scores of each gesture's classifier in the last *k* frames.

A. Representation

Each detected hand is represented as a vector (x, y, v_x, v_y, t), with (x, y) the hand's position, (v_x, v_y) the hand's velocity, and *t* the frame's timestamp. In order to achieve translation and scale invariance, coordinates (x, y) are measured with respect to the face, and normalized by the size of the face. Using this hand's vector, statistics (features) that characterize the subsequence of detections (a list of (x, y, v_x, v_y, t) vectors) are calculated. The components of the feature vector are:

- DELTA_X: difference between maximal and minimal position in the *x* axis.
- DELTA_Y: difference between maximal and minimal position in the *y* axis.
- AVE_X: mean hand's position in the *x* axis.
- AVE_Y: mean hand's position in the *y* axis.
- AVE_VX: mean hand's speed in the *x* axis.
- AVE_VY: mean hand's speed in the *y* axis.
- AVE_R: average distance between (x, y) and (AVE_X,AVE_Y).
- AVE_ANGLE: average polar angle of the hand, measured with respect to the x-axis of a coordinate system centered in (AVE_X,AVE_Y).
- COV00, COV01, COV02, COV03, COV11, COV12, COV13, COV22, COV23, COV33: Components of the covariance matrix of vectors (x, y, v_x, v_y). (COV00=cov(X,X), COV23=cov(VX,VY), etc.).

- CH_AEXT: Area of the convex hull of the (x,y) positions.
- CH_PEXT: Perimeter of the convex hull of the (x,y) positions.
- SKEW_X: Skewness of *x* positions of the observed hands.
- SKEW_Y: Skewness of *y* positions of the observed hands.
- HISTX00, HISTX01, …, HISTX*k*: This histogram is computed by dividing the x-axis into *k*+1 columns (*k*=9 was used) and counting the number of frames that fall into each column. The part of the *x*-axis that is divided is the one between the minimal and maximal *x* coordinates observed in the stored frames.
- HISTY00, HISTY01, …, HISTYk: Analogous to above, but computed in the *y*-axis.
- HIST2D00, ..., HIST2D*nm*: An *n*x*m* grid that spans the tightest bounding box that includes all hands detected in the observed frames is defined. The value of HIST2D*xy* equals the percentage of hands that have been detected in cell (x,y) of this grid.
- WX: **x**-axis hand frequency. This frequency is measured by observing the instants in which the hand reaches minima and maxima in the *x*-axis, and registering the time elapsed between these extrema. Frequency is then estimated as the inverse of such time. WX is the average of such inverses.
- WY: Analogous to above, but computed in the *y*-axis.

B. *Classification of Segmented Gestures*

Segmented gestures are characterized using the feature vector defined in the former section, and classified using standard statistical classifiers. To accomplish this Naïve Bayes, SVM and C4.5 classifiers are analyzed and compared using a dataset comprised by 300 gestures, 30 for each digit from 0 to 9 [16]. In this dataset colored gloves are used in order to make labeling easier; correct gestures can be easily extracted since the initial and final frame of each gesture is known. Gesture classification results are shown in tables 1-3. It can be seen that all classifiers get very high recognition rates. However, a Naïve Bayes classifier is selected, because besides being the most accurate one (99%), it is very fast and the probability score can be easily thresholded for gesture spotting.

Table 1. Confusion matrix of the dynamic gesture classification. 10-fold Cross validation results. Naive Bayes classifier: 99% correct classification. Rows/Columns: real/predicted gestures.

Real\Predicted	0	1	2	3	4	5	6	7	8	9
0	29						1			
1		30								
2			30							
3				30						
4					30					
5				1		28	1			
6							30			
7								30		
8									30	
9										30

Table 2. Confusion matrix of the dynamic gesture classification. 10-fold Cross validation results. SVM classifier: 98.67% correct classification. Rows/Columns: real/predicted gestures.

Real\Predicted	0	1	2	3	4	5	6	7	8	9
0	30									
1		30								
2			30							
3				29					1	
4					30					
5						29			1	
6	1						29			
7								30		
8									30	
9									1	29

Table 3. Confusion matrix of the dynamic gesture classification. 10-fold Cross validation results. C4.5 classifier: 95.67% correct classification. Rows/Columns: real/predicted gestures.

Real\Predicted	0	1	2	3	4	5	6	7	8	9
0	29									1
1		28			1	1				
2			27		1			2		
3				28		1				1
4			1		29					
5						28				2
6							30			
7					1			29		
8									30	
9						1				29

C. Gesture Segmentation and Classification

The described Naïve Bayes classifier achieves good performance on segmented gestures. However, in a online gesture recognition scenario, the start and end frames are not known. Thus, using this classifier alone is not enough to recognize a gesture in an arbitrary sequence of frames in which only a subsequence may correspond to a true gesture. A solution to this problem is to identify start and end frames by means of thresholding hand velocity. However, we have discovered that this approach is not robust enough, since the hand can have low velocity even during the execution of a gesture, and one should instead consider several start and end frame candidates. A gesture will be declared as recognized only when it achieves a high probability score in several subsequences (delimited by several different start, end frame pairs). The Dynamic Gesture Recognition Module (DGRM) is responsible of accomplishing this task.

In practice, the DGRM analyzes all available subsequences of "adequate" length (not too long or too short, measured in frames). Every time a new detected hand arrives, it is represented as described in section IV.A, and it is added to a global frame sequence. Then, all "adequate" length subsequences that end with this new frame are reduced to a feature vector (as in section IV.A.) and fed to the Naïve Bayes classifier. For each feature vector, the classifier outputs a list of gestures and their probabilistic scores. Each score represents the likelihood of each gesture in the given subsequence. To sum up this large amount of information, the scores may be grouped by gesture, regardless of the subsequence they came from. Then, for each gesture, the maximum of these scores is taken to be the best likelihood of that gesture having occurred, given the new frame.

These maxima are stored, and the procedure is repeated when a new frame arrives. Getting the highest score after one particular frame is not enough to declare the gesture as recognized: the score should be consistently high during several frames. Thus, to account for this fact and to have robustness against noise and outliers, the scores are low-pass filtered, and a moving-average of k maximum scores for each gesture is computed. To avoid the problem of storing all moving-averages of a given gesture, a *best moving-average* (*bma*) is computed as the highest moving average that has been computed for that gesture. In the current frame the computed moving-average is compared to the *bma*, and the highest value is kept as *bma*. The gesture with the highest *bma* is the recognized gesture in this moment.

Since not every frame is a real-end of a gesture, gesture segmentation is still a problem. Following HMM or CDP based dynamic gesture recognition frameworks, thresholding the *bma* is a possible approach for gesture spotting. In addition, the current *bma* can be decremented in each round as a penalty for the subsequence from which it was extracted becoming older. Also, stored frames are discarded when an inactivity condition is detected (still hand, hand out of camera range).

To manage all these computations we use a so-called "gestures-table" data structure (see figure 2). In this table, all information of a given gesture is stored in a row: the last k maximum scores, as well as the *bma*. In the table $ms_{i,n-j}$ represents the maximum score of gesture i in frame n-j. In addition, the associated sequences, including starting and ending times are stored. In summary, after calculating the *bma* for each of the m gestures (bma_i), the largest one is determined $j = \arg\max_{i=1,\ldots m} bma_i$, and afterwards bma_j is thresholded for gesture spotting.

Gesture	List of last k maximum scores (ms)	Best moving average of ms
G_1	$ms_{1,n-k}, \ldots, ms_{1,n-1}$	bma_1
$\vdots$	$\vdots$	$\vdots$
G_m	$ms_{m,n-k}, \ldots, ms_{m,n-1}$	bma_m

Fig. 2. Gestures-table used to gesture discrimination.

V. Results

The whole gesture recognition system was tested in real video sequences obtained in office environments with dynamic conditions of illumination and background. In all these sequences the service robot interact with the human user at a variable distance of one to two meters. The size of the video frames is 320x240 pixels, and the robot main computer where the gesture recognition system runs is a standard notebook (Tablet HP 2710p, Windows Tablet SO, 1.2 GHz, 2 GB in RAM). Under these conditions, once the system detects the user's face, it is able to run at a variable speed of 4-8 frames per second, which is enough to allow an adequate interaction between the robot and the human user. The system's speed is variable because it depends on the convergence time of the *mean shift* algorithm and the face and hands statistical classifiers. See figure 3 for an example of the output of the tracking system in an office environment. In figure 4 are shown some examples of the trajectories generated by the hand tracking process. Notice the stability of the hand detections and tracking, which includes the translation and scale normalization done using the output of the face detection and tracking.

We have also evaluated the proposed dynamic gesture recognition framework in the *10 Palm Graffiti Digits* database [16], where users perform gestures corresponding to the 10 digits (see example in figure 3). In the experiments the users and signers can wear short sleeved shirts, the background may be arbitrary (e.g, an office environment) and even contain other moving objects, and hand-over-face occlusions are allowed. We use the easy test set, which contains 30 short sleeve sequences, three from each of 10 users (altogether 300 sequences). The training set is the same one presented in the previous section.

The system was able to detect and track hands in 266 of the 300 sequences (89%). In these 266 sequences, the dynamic gestures (i.e. digits) were correctly recognized in 84% of the cases. This corresponds to a 75% recognition rate (225 from 300 cases). It can be seen that this recognition rate is very similar to the one obtained in state of the art systems (e.g. [16], based on Hidden Markov Models, which are not able to operate in real-time or near real-time.

Table 4 shows the confusion matrix of the dynamic gesture recognition. It can be observed that the recognition rate of six digits is very high ("0"-"4", "8" and "9"). Two digits are recognized in most of the cases ("6" and "7"), and just the "5" digit has recognition problems. The "5" is confused, most of the time with the "3".

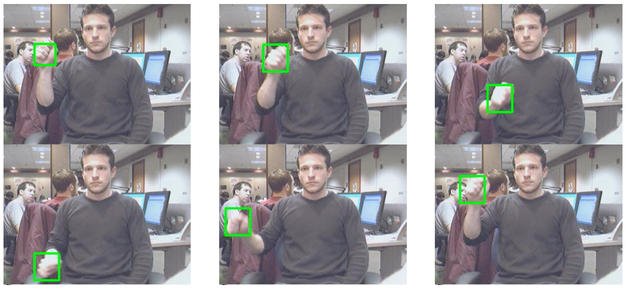

Fig. 3. Example of tracked hands in the *10 Palm Graffiti Digits* database [16].

Table 4. Confusion matrix of the dynamic gesture recognition module (rows: real gesture, columns: predicted gesture). TP: True Positives. FP: False Positives. RR: Recognition Rate.

	0	1	2	3	4	5	6	7	8	9	TP	FP	RR (%)
0	20	0	0	0	0	0	0	0	1	0	20	1	95
1	0	30	0	0	0	0	0	0	0	0	30	0	100
2	0	0	22	0	0	0	0	0	0	0	22	0	100
3	0	0	0	26	0	0	0	0	0	0	26	0	100
4	0	0	0	0	30	0	0	0	0	0	30	0	100
5	0	0	0	22	0	3	2	0	0	0	3	24	11
6	4	0	0	0	0	0	23	0	0	0	23	4	85
7	0	9	0	0	0	0	0	18	0	1	18	10	64
8	0	0	0	0	0	0	0	0	28	0	28	0	100
9	0	0	0	2	0	0	0	0	0	25	25	2	93

VI. Conclusions

In this article a hand gesture recognition system that allows interacting with a service robot, in dynamic environments and in real-time, was described. The system detects hands and static gestures using a cascade of boosted classifiers, and recognizes dynamic gestures by computing temporal statistics of the hand's positions and velocities, and classifying these features using a Bayes classifier. The approach makes use of temporal statistics about the hand's positions and velocities as basic information to recognize the gestures. The use of these features allows carrying out the final recognition using a standard Bayes classifier, instead of the traditional Hidden Markov Models. The gesture segmentation and recognition is achieved simultaneously by finding candidate subsequences that give high scores when matched to a gesture. The system performance is validated in real video sequences. The size of the video frames is 320x240 pixels, and the robot computer where the gesture recognition system runs is a standard notebook (Tablet HP 2710p, Windows Tablet SO, 1.2 GHz, 2 GB in RAM). Under these conditions, once the system detects the user's face, it is able to run at a variable speed of 4-8 frames per second. In average the system correctly detects and tracks the hands in 89% of the cases, the dynamic gestures are correctly classified in 84% of the cases when the hand is correctly tracked, which gives a 75% recognition rate for the whole system.

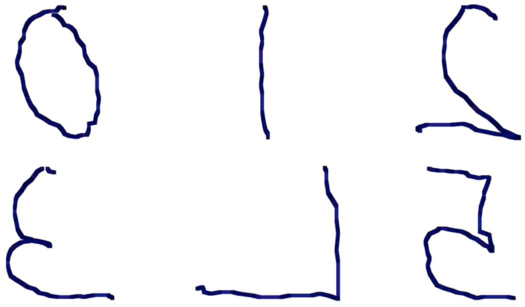

Fig. 4. Examples of the input to the dynamic gesture recognition module when the tracking module works correctly.

REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, Kernel-Based Object Tracking, *IEEE Trans. on Pattern Anal. Machine Intell.*, vol 25, no. 5, (2003) pp. 564 – 575.

[2] X. Liu and K. Fujimura, Hand gesture recognition using depth data, *Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition*, (2004) pp. 529 – 534, Seoul, Korea.

[3] M. Kolsch, M.Turk, Robust hand detection, *Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition*, (2004) pp. 614 – 619, Seoul, Korea.

[4] N. Dang Binh, E. Shuichi, T. Ejima, Real-Time Hand Tracking and Gesture Recognition System, *Proc. GVIP 05*, (2005) pp. 19-21 Cairo, Egypt.

[5] C. Manresa, J. Varona, R. Mas, F. Perales, Hand Tracking and Gesture Recognition for Human-Computer Interaction, *Electronic letters on computer vision and image analysis*, Vol. 5, Nº. 3, (2005) pp. 96-104.

[6] Y. Fang, K. Wang, J. Cheng, H. Lu, A Real-Time Hand Gesture Recognition Method, *Proc. 2007 IEEE Int. Conf. on Multimedia and Expo*, (2007) pp. 995-998

[7] Q. Chen, N.D. Georganas, E.M. Petriu, Real-time Vision-based Hand Gesture Recognition Using Haar-like Features, *Proc. Instrumentation and Measurement Technology Conf. – IMTC 2007*, (2007)Warsaw, Poland

[8] A. Angelopoulou, J. García-Rodriguez, A. Psarrou, Learning 2D Hand Shapes using the Topology Preserving model GNG, *Lecture Notes in Computer Science* 3951 (*Proc. ECCV 2006*), pp. 313-324

[9] J. Ruiz-del-Solar, and R. Verschae, Skin Detection using Neighborhood Information. *6th Int. Conf. on Face and Gesture Recognition* – FG 2004, pp. 463 – 468, Seoul, Korea, May 2004.

[10] H. Francke, J. Ruiz-del-Solar, R. Verschae, Real-time Hand Gesture Detection and Recognition using Boosted Classifiers and Active Learning, *Lecture Notes in Computer Science* 4872 (*Proc. PSIVT 2007*), Springer, pp. 533-547, 2007.

[11] J. Ruiz-del-Solar, R. Verschae, M. Correa, J. Lee-Ferng, N. Castillo, Real-Time Hand Gesture Recognition for Human Robot Interaction, *RoboCup Symposium 2009* (in press).

[12] R. Verschae, J. Ruiz-del-Solar, M. Correa, A Unified Learning Framework for object Detection and Classification using Nested Cascades of Boosted Classifiers, *Machine Vision and Applications*, Vol. 19, No. 2, pp. 85-103, 2008.

[13] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, (2001) pp. 511 – 518.

[14] I.H. Witten and E. Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[15] RoboCup @Home Official website. Available in March 2009 in http://www.robocupathome.org/

[16] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff, A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation, *IEEE Trans. on Pattern Anal. Machine Intell.* (in press, electrically available on July 28, 2008).

[17] M.K. Bhuyan, D. Ghosh, and P.K. Bora. Continuous hand gesture segmentation and co-articulation detection, *Lecture Notes in Computer Science 4338*, pp 564-575, 2006.

[18] Bobick, A. F. and Davis J. W., The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(3), pp 257-267, 2001.

[19] Cui Y. and Weng J., Appearance-based hand sign recognition from intensity image sequences. *Comput. Vis. Image Underst.*, 78(2), pp. 157-176, 2000.

[20] Darrell T. J., Essa I. A., and Pentland A. P,. Task-specific gesture analysis in real-time using interpolated views. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(12), pp. 1236-1242, 1996.

[21] Deng J. W. and Tsui. H. T., An hmm-based approach for gesture segmentation and recognition. *In ICPR '00: Proceedings of the International Conference on Pattern Recognition*, page 3683, Washington, DC, USA, 2000. IEEE Computer Society.

[22] Eickeler S., Kosmala A., and Rigoll G., Hidden Markov model based continuous online gesture recognition, *Proceedings of the Fourteenth International Conference on Pattern Recognition*, vol.2, pp.1206-1208, 1998

[23] Kim D., Jinyoung Song, and Kim D., Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMs. *Pattern Recogn.*, 40(11), pp. 3012-3026, 2007.

[24] Lee H-K, and Kim J-H., An HMM-based threshold model approach for gesture recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(10):961-973, 1999.

[25] Lee S-W. Automatic gesture recognition for intelligent human-robot interaction. *In FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pp 645-650, Washington, DC, USA, 2006. IEEE Computer Society.

[26] Malassiotis S. and Strintzis M. G., Real-time hand posture recognition using range data. *Image Vision Comput.*, 26(7), pp. 1027-1037, 2008.

[27] Rabiner L. R., A tutorial on hidden markov models and selected applications in speech recognition, *Proceedings of the IEEE* , vol.77, no.2, pp.257-286, 1989.

[28] Shan C., Wei Y., Qiu X., and Tan T., Gesture recognition using temporal template based trajectories.*International Conference on Pattern Recognition*, 3, pp. 954-957, 2004.

[29] Wang S.B., Quattoni A., Morency, L-P, Demirdjian D., and Darrell T., Hidden conditional random fields for gesture recognition. *In CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1521-1527, Washington, DC, USA, 2006. IEEE Computer Society.

[30] Wang T-S, Shum H-Y, Xu Y-Q, and Zheng N-N. Unsupervised analysis of human gestures. *In PCM '01: Proceedings of the Second IEEE Pacific Rim Conference on Multimedia*, pages 174-181, London, UK, 2001. Springer-Verlag.

[31] Yang M-H, Ahuja N., and Tabb M., Extraction of 2d motion trajectories and its application to hand gesture recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8), pp. 1061-1074, 2002.

[32] Zhao Z. and Elgammal A., Spatiotemporal pyramid representation for recognition of facial expressions and hand gestures. *In 8th IEEE Int'l Conf. on Automatic Face and Gesture Rec.*, Washington, DC, USA, 2008. IEEE Computer Society.

[33] Kong W.W., Ranganath S, Automatic hand trajectory segmentation and phoneme transcription for sign language. In *8th IEEE Int'l Conf. on Automatic Face and Gesture Rec.*, Washington, DC, USA, 2008. IEEE Computer Society.