COMPARISON OF FOURIER DESCRIPTORS AND HU MOMENTS FOR HAND POSTURE RECOGNITION

S. Conseil [1], *S. Bourennane* [1] *and L. Martin* [2]

[1] Institut Fresnel (CNRS UMR6133), École Centrale de Marseille
DU de Saint Jérôme, F-13013 Marseille Cedex 20, France
phone: +33 4 91 28 82 02, fax: +33 4 91 28 88 13
email: simon.conseil@fresnel.fr, salah.bourennane@fresnel.fr

[2] ST Microelectronics
ZI Rousset BP 2, F-13106 Rousset, France
phone: +33 4 42 68 88 95, fax: +33 4 42 68 87 00
email: lionel.martin@st.com

ABSTRACT

In this paper, we propose to use Fourier descriptors (FD) for hand posture recognition in a vision-based approach. FD are widely used for shape representation and pattern recognition, they may also be well-adapted for hand posture recognition. The invariance properties of FD are discussed, and we provide a comparison of the performances with Hu moments. First, experiments are performed on the Triesch hand posture database. Then we define our own gesture vocabulary, with 11 gestures, and we perform the acquisition of a large number of images, with 18 persons. Hence tests are performed on a more realistic database, with various hand configurations realized by non-expert users. Results show that FD give very good recognition rates in comparison with Hu moments. This confirms the efficiency of FD and shows their great robustness in real-life conditions.

1. INTRODUCTION

With the need to find new means of interactions with computers, hand gestures recognition has become a challenging topic of research [1]. Vision-based hand gesture recognition techniques have many advantages, compared with devices such as mice, keyboards or electronic gloves. It allows to use directly the hand to interact or communicate with a computer, thus it provides more intuitive means of interaction. However recognizing the shape (posture) and the movement (gesture) of the hand in images is a complex task. The hand is indeed a deformable object with a large number of degrees of freedom.

In an appearance-based approach, it is necessary to define a vocabulary of gestures, by taking into account some constraints: gestures must be simple, intuitive, easily reproducible, different enough from each other. This vocabulary can be inspired from Sign Languages [2]. Applications can be found in augmented and virtual reality, in video games, or in user interfaces to replace the computer mice (tactile screen with video cameras) and to enhance means of interaction.

This paper presents a study of hand posture recognition using Fourier descriptors (FD). FD are indeed efficient features for pattern recognition, and we propose to study their performances for hand posture recognition. FD have already been experimented for hand gesture recognition [3], but only as a part of a complete recognition framework, thus the performances of FD were not analyzed in details.

We provide a comparison of the performances of FD with a widely used method, Hu invariants [4]. First experiments are performed on the Triesch gesture database [5]. Then to provide more significant experiments, we define our own vocabulary, with 11 gestures, and we make the acquisition of a database with 18 persons. A large amount of images has been collected, with various postures of the hand in order to test extensively the performances of FD in regard to invariances.

The paper is organized as follows. Section 2 summarizes related work. In section 3 an overview of the hand databases is given, followed by the description of Fourier descriptors and Hu moments in Section 4. Experiment setup, results and analysis are discussed in Section 5.

2. RELATED WORK

Hand gesture recognition is a challenging task in which two main approaches can be distinguished: hand model-based and appearance-based methods [1, 2]. Although appearance-based methods are view-dependent, they are more efficient in computation time. They generally aim at recognizing a gesture among a vocabulary, with template gestures learned from training data, whereas hand model-based methods are used to recover the exact 3D pose of the hand. The majority of appearance-based models consist in extracting features that represent the content of the images. To cope with the point of view dependency, these methods must have invariance properties to translation, rotation and scale changes. These features can be region-based such as Hu invariants [4] and Zernike moments [6], or contour-based such as Fourier descriptors (FD) or orientation histograms [7]. Caplier *et al.* [8] reports good classification results for eight gestures, with Hu moments and a multi-layer perceptron classifier. These models can also be based on deformable templates, for instance Ahmad *et al.* [9] use Point Distribution Models to track and classify five gestures. Kolsch *et al.* [10] use the Viola and Jones detector to detect the hand and recognize six hand postures.

FD are widely used for the description and classification of shapes with a closed contour, because they represent well the shapes and they have interesting invariance properties [11]. FD are computed with the Fourier coefficients given by the Fourier transform of the shape signature of the contour. Many different signatures can be associated with contours, for instance centroid distance, complex coordinates, curvature function, or cumulative angles. According to [12], the centroid distance gives the best results for synthetic shape retrieval, and the curvature function gives the worst results. Complex coordinates are also often used, particularly for FD-based hand gesture recognition [3, 13].

Classification is generally achieved by a distance metric and nearest-neighbor rules [3, 12, 14], but also with classifiers such as Radial-Basis Function (RBF). FD have also been used as input features for dynamic gesture recognition with Hidden Markov Models (HMM) [13, 15] and Neural
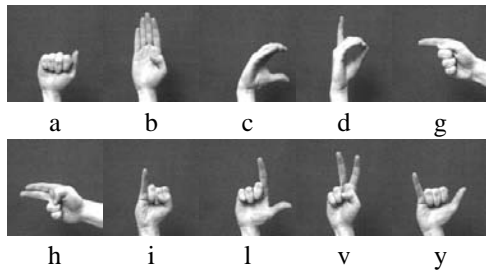
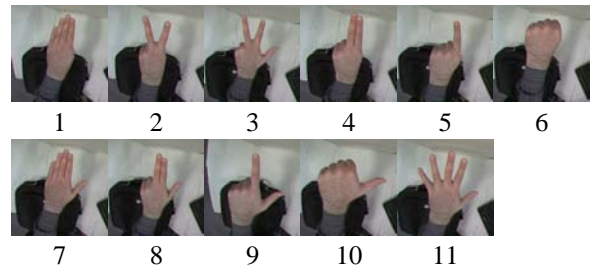Figure 1: Jochen Triesch hand posture database.

Figure 2: The 11 gestures of our database.

Networks [13]. The classification involves a learning procedure, for which the number of training images and the number of gestures are important factors. For instance in [15], 1200 sequences, 20 gestures and 20 persons are used and a recognition rate of 90.5% is obtained.

An interactive training method with both supervised and unsupervised steps has been proposed in [3]. The parameters of template gestures are refreshed with the well-classified gestures, and if a gesture is misclassified the user can send a feedback signal to the system and retrain the gesture. A recent study [14] provides a comparison on human posture recovery, between FD with complex coordinates, Hu moments and shape context histograms. The authors perform tests with deformed shapes to test the robustness of each descriptor, and they show that FD and shape context histograms outperform Hu moments.

3. SYSTEM OVERVIEW

In this paper we study the performances of Fourier descriptors (FD) for hand posture recognition, and we compare the results with Hu moments, in real-life conditions and with non-expert users. After experimenting FD with an existing benchmark database, we have defined our own gesture vocabulary - inspired by existing ones - and we have made the acquisition of a new database with 18 persons, most of which are not familiar with gesture recognition. Indeed it is very important to test the recognition with non-expert people, it permits to notice which gestures are easy to perform, and which are subject to confusions. This paper does not face all the different steps of a gesture recognition system, we focus on FD and their properties, which are invariances to translation, rotation and scale changes.

3.1 Triesch database

The first database used for experiments is the *Jochen Triesch Hand Posture Database* [5], available on the web[1], and which consists of 10 hand signs performed by 24 persons against three backgrounds. We use the sets of images with black and white backgrounds, which represents 479 images (one is missing for the 'v' gesture). Images are 128×128 pixels in gray levels (Fig. 1).

3.2 Our database

The Triesch database is too small to make representative experiments and gestures are always taken from the same point of view, with the same size and orientation. Thus we have made the acquisition of our own database with 11 gestures

[1]http://www.idiap.ch/resources/gestures/

(Fig. 2). We have chosen some very similar gestures to test the discrimination performances, such as gestures 4 and 5, and gestures 8 and 9.

Hand images acquisition has been realized by 18 persons in the following conditions: (i) indoor environment, (ii) gestures are realized over a desk, (iii) the camera is placed above the desk, images are 320×240, (iv) people wear long sleeve clothes, to avoid the issue of wrist detection and forearm separation, (v) no assumption is made about the entry point of the hand in the scene. If the arm was naked, it would be necessary to separate the hand from the forearm. This could be done by considering the width of the arm as in [3, 13]. But, as we focus on the recognition with FD, we have avoided this issue by making image acquisitions with people wearing sleeve clothes.

The purpose of our database is to test extensively the performances of recognition, and in particular to verify the invariances in translation, rotation and scale changes. Thus we asked users to move their hand in all the workspace, including depth for scale changes, and this for each gesture. This results in nearly 1000 images per gesture for each person.

4. FEATURES EXTRACTION

4.1 Hand segmentation

Images of the Triesch hand database are in gray levels, one set with black background and the other with white background. The hand is segmented with the Otsu threshold method [16], which computes the best threshold based on the histogram of gray levels.

For the color images of our database, the hand is segmented with a skin color detection method, which consists in thresholding the C_b and C_r components [17]:

$$C_b \in [77, 127] \quad \text{and} \quad C_r \in [133, 173] \quad (1)$$

This segmentation step is a common drawback of shape-based method. Despite active research on hand segmentation, there is still no perfect method for detecting a hand in all conditions, with complex background and varying illumination. Background subtraction methods are very sensitive to shadows, even in the case of background models with mixture of gaussians which are moreover quite computationally expensive. Skin color methods are the most suitable for hand detection, but they may fail on some types of skins (very dark or very bright), or if the background contains flesh-color objects. In the case of our database, the simple CbCr thresholds detection method works well on the different users skin, and the acquisition conditions are quite realistic as it is a common desktop in an open-space range. For a more robust application, we could use more efficient skin detection, using a skin
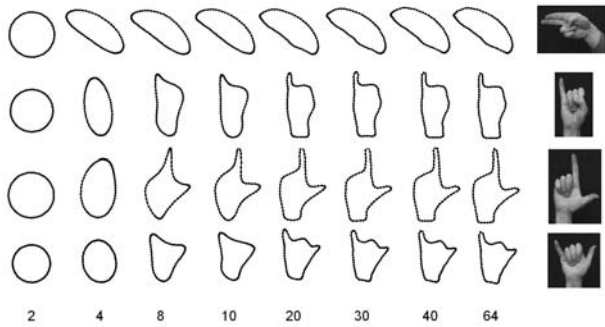
Figure 3: Example of reconstruction with FD, as a function of the cut-off frequency, with an initial contour sampled at 64 points.

model learning with histograms or with mixture of gaussians. Another solution would be to combine the informations from skin color, motion and edge such as in [15].

4.2 Fourier descriptors

FD are calculated on the contour of the hand region. Points of this contour can be represented with various signatures (complex coordinates, central distance, curvature, cumulative angular function) [12]. We use the complex coordinates, each point M_i of the shape contour is represented by a complex number z_i, with N the number of points of the contour:

$$\forall i \in [1, N], M_i(x_i, y_i) \Leftrightarrow z_i = x_i + jy_i \tag{2}$$

Before calculating the Fourier Transform — with the Fast Fourier Transform (FFT) — the contour is sampled to obtain a normalized contour length. The sampling is done by interpolating points which are at an equal arc length. The length must be chosen as a compromise between a good description of the shape, with enough details, and a smoothing on the shape, which eliminates the finest details which can be noise. Another factor is the computation time which increases with the number of points. For computational efficiency of the FFT, the number of points is chosen to be a power of two. The normalized length is generally chosen to $N = 64$. Hence the Fourier Transform leads to N Fourier coefficients C_k:

$$C_k = \sum_{i=0}^{N-1} z_i \exp\left(\frac{-2\pi jik}{N}\right), \quad k = 0, \ldots, N-1 \tag{3}$$

The first coefficient C_0 is discarded because it contains only the position of the hand shape (*i.e.* translation). Rotation of the shape affects only the phase information, thus rotation invariance of the Fourier descriptors is achieved by taking the magnitude of coefficients. Scale invariance is achieved by dividing coefficients by the magnitude of the second coefficient, C_1. Starting point invariance is also achieved by taking the magnitude, as a change of the starting point affects only the phase. Finally, we obtain $N - 2$ Fourier descriptors I_k:

$$I_k = \frac{|C_k|}{|C_1|}, \quad k = 2, \ldots, N-1 \tag{4}$$

Figure 3 shows that the low frequency coefficients contain information on the general form of the shape and the high frequency coefficients contain information on the finer details of the shape. We can notice that with more than 20 coefficients the hand shape is well reconstructed.
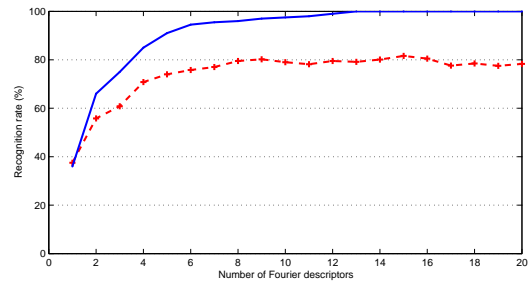

Figure 4: Classification rate as a function of the number of FD, for the Triesch database, with learning images in blue and test images in red dotted.

4.3 Hu moments

Hu invariants [4] are calculated with the geometrical moments of the hand region. The first six descriptors encode a shape with invariance to translation, scale and rotation. The seventh descriptor ensures skew invariance, which enables to distinguish between mirrored images. Hu moments computation is detailed in the Appendix. In contrast to FD, moments are region-based descriptors. Hu invariants are used in pattern recognition to provide a scale, orientation and position invariant representation of the shape of an object.

4.4 Classification

Hand shapes are classified using a bayesian distance. For each class i, corresponding to a gesture, a mean invariant vector μ_i and a covariance matrix Λ_i are learned using a training set. Then each image, represented by the invariant vector $\mathbf{x}$, is classified by minimizing the Mahalanobis distance:

$$\forall \text{class } i, \, g_i(\mathbf{x}) = (\mathbf{x} - \mu_i)^T \Lambda_i^{-1}(\mathbf{x} - \mu_i) \tag{5}$$

As the number of images in the Triesch database is small, we perform a cross-validation to estimate the recognition rates.

5. RESULTS AND DISCUSSION

5.1 Triesch database

With all the images for both learning and testing, we obtain a 79% recognition rate for Hu moments, with bad results for gestures 'b' and 'g'. For FD, the recognition rate depends on the number of descriptors (fig. 4). With 6 features the recognition rate is nearly 95%, and with more than 13 descriptors it reaches 100%.

Concerning the invariances, we perform tests with rotated and scaled images. With rotation angles 90°, 180° and 270° we still obtain a 100% rate of good classification with FD. However, for other angles that are not multiple of 90°, for instance with 30° and 60°, we notice that the recognition rate decreases to nearly 90%. This comes from the interpolation after the rotation to match the discrete grid of the images. Cross-validation gives also better results for FD: recognition rate are 65% for Hu moments and 78% for FD.

5.2 Our database

The learning is done with manually selected images of an expert user, with nearly 500 images per gesture. In the following tests, we take only 6 Fourier descriptors and the first

Figure 5: Example of images of our database.

	1	2	3	4	5	6
Hu	82.31	68.28	66.52	65.04	93.80	88.98
FD	86,58	90,75	96,38	60,77	97,81	94,31
	7	8	9	10	11	**Total**
Hu	48.29	65.42	82.78	74.35	40.57	**71.08**
FD	80,55	64,83	88,58	73,44	96,19	**84,58**

Table 2: Recognition results with our database, with 6 Hu invariants and 6 Fourier descriptors.

six Hu invariants. Indeed the seventh invariant enables to distinguish between mirrored images, but we have both left and right hands in our database. Thus without the seventh Hu invariant the recognition is independent of which hand has been used, and the recognition rate increases from 53% to 71%. In order to validate the learning stage, we run the classification on the learning images, and we obtain recognition rates of 98.11% for Hu moments and 99,96% for Fourier descriptors. Then images of the other users are classified using this learning data, with approximatively 1000 images per gesture for each user. Figure 5 shows various examples of hand images of test set.

Table 2 shows the recognition rates for each gesture. We obtain a total of 86.22% for FD versus 71.08% for Hu moments. This difference comes from a great improvement of recognition for the gestures 2, 3, 7 and 11. Hence FD outperform Hu moments in terms of discrimination between visually close gestures.

The analysis of confusion matrices (Tables 1 and 3) confirms the results and reveals which gestures are sources of errors. For instance, gestures 2, 3 and 9 are very similar and thus are confused with Hu moments method, whereas they are well recognized with FD. However for gestures 4 and 5, and gestures 8 and 9, the recognition rate is a little better but still does not exceed 76%. This is not really surprising, indeed gestures 4 and 8 have been chosen in order to test the limits of recognition. If we remove these two gestures and run the classification again, the global recognition rate increases to 90.52% for FD and 72.97% for Hu.

In these experiments we notice that recognition rates are very dependent on the users and on how they realize the gestures. Some users have sometimes difficulties to realize one

of the gestures, and thus its score is very bad. To improve the recognition an idea would be to add some images of the user's hand in the learning stage.

5.3 Application

The feasability of a posture recognition system based on FD has been experimented with a demo software and a basic webcam. The system runs in real time (20–25 fps) on a 2 GHz PC, with 320×240 images. The skin segmentation is robust enough to different users. With a little learning of the gestures, people manage easily to have their gestures been recognized.

6. CONCLUSION

In this paper, we have presented two shape descriptors for vision-based hand posture recognition. The performances of Fourier descriptors and Hu moments have been tested extensively. For this purpose, we first use a benchmark database, then we have defined our own gesture vocabulary and we have made the acquisition of a larger database, with 11 gestures, 18 persons, and approximatively 1000 images per gesture and per person. Most of these persons are not expert users of a hand gesture recognition system. They have realized the gestures in various configuration so that we could test the robustness of the Hu invariants and Fourier descriptors to changes in translation, rotation, scale, and point of view. We have shown that Fourier descriptors outperforms Hu moments for all deformations. Future work will aim at improving the hand detection with tracking methods, and taking advantage of the temporal stability of a given posture in a video stream.

	1	2	3	4	5	6	7	8	9	10	11
1	**82,3**	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
2	4,7	**68,3**	1,5	1,2	0,7	0,6	1,2	0,5	0,1	0,0	2,3
3	0,7	21,5	**66,5**	0,5	1,1	2,5	20,0	20,3	10,2	2,0	40,1
4	3,5	0,2	0,0	**65,0**	2,0	0,0	0,5	0,1	0,0	0,7	0,0
5	5,6	8,3	0,8	31,2	**93,8**	0,7	6,2	4,2	1,6	11,7	0,0
6	0,8	0,0	0,0	0,1	0,2	**89,0**	0,0	0,0	0,0	0,3	0,0
7	0,2	0,1	0,4	0,2	0,2	0,6	**48,3**	1,0	0,7	0,3	0,6
8	1,9	0,4	1,8	1,2	1,4	0,4	4,4	**65,4**	3,7	1,2	0,2
9	0,0	1,0	28,8	0,1	0,1	0,5	15,2	8,1	**82,8**	7,3	16,2
10	0,0	0,1	0,0	0,2	0,1	0,2	0,2	0,0	0,0	**74,4**	0,0
11	0,2	0,2	0,1	0,3	0,2	5,6	4,0	0,5	1,0	2,0	**40,6**

Table 1: Confusion matrix for Hu invariants recognition results.

	1	2	3	4	5	6	7	8	9	10	11
1	**86,6**	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
2	0,0	**90,8**	0,4	0,4	0,2	0,2	0,1	0,0	1,7	0,1	0,1
3	0,0	0,7	**96,4**	0,5	0,4	1,0	0,4	0,0	0,7	0,1	3,3
4	5,5	0,0	0,0	**60,8**	0,0	0,1	0,4	0,0	0,0	0,0	0,0
5	2,9	1,8	0,5	35,9	**97,8**	0,9	7,8	3,2	4,9	20,2	0,1
6	4,6	0,1	0,0	0,1	0,3	**94,3**	0,8	0,0	0,2	2,0	0,0
7	0,2	0,4	0,1	0,7	0,5	1,1	**80,6**	8,3	0,3	2,8	0,0
8	0,0	0,2	0,0	0,3	0,3	0,1	1,9	**64,8**	2,8	0,5	0,0
9	0,0	5,9	1,7	0,9	0,3	0,4	6,0	23,2	**88,6**	0,9	0,4
10	0,0	0,1	0,1	0,3	0,0	0,2	0,8	0,4	0,2	**73,4**	0,0
11	0,2	0,2	0,8	0,1	0,1	1,6	1,1	0,1	0,7	0,0	**96,2**

Table 3: Confusion matrix for Fourier descriptors recognition results.

7. ACKNOWLEDGEMENT

8. APPENDIX

Hu moments computation

Classical geometric moments m_{pq} of an image I_{xy} are calculated with the following equation:

$$m_{pq} = \sum_{x=1}^{M} \sum_{y=1}^{N} x^p y^q I_{xy} \qquad (6)$$

This allow to compute the center of mass of the image, and of a region in case of a binary mask. Centralized moment μ_{pq} are geometric moments of the image computed relatively to the center of mass $(\overline{x}, \overline{y})$:

$$(\overline{x}, \overline{y}) = (\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}) \qquad (7)$$

$$\mu_{pq} = \sum_{x=1}^{M} \sum_{y=1}^{N} (x - \overline{x})^p (y - \overline{y})^q I_{xy} \qquad (8)$$

Centralized moments are invariant under translation. To enable invariance to scale, normalized moments η_{pq} are used:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \text{, with } \gamma = \frac{p+q}{2} + 1, \forall p+q \geq 2 \qquad (9)$$

Hu [4] proposed a set of orthogonal moment invariants, which can be used for scale, position, and rotation invariant pattern identification. They are computed from normalized moments η_{pq} up to order three, with the following formulas:

$$I_1 = \eta_{20} + \eta_{02} \qquad (10)$$

$$I_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \qquad (11)$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \qquad (12)$$

$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \qquad (13)$$

$$I_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \qquad (14)$$
$$-3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})$$
$$[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$I_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \qquad (15)$$
$$+4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})]$$

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21}$$
$$+\eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} \qquad (16)$$
$$+\eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

REFERENCES

[1] V. Pavlovic, R. Sharma, and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *IEEE trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677–692, Jul. 1997.

[2] S. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873–891, Jun. 2005.

[3] A. Licsar and T. Sziranyi, "User-adaptive hand gesture recognition system with interactive training," *Image and Vision Computing*, vol. 23, no. 12, pp. 1102–1114, Nov. 2005.

[4] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE trans. on Information Theory*, vol. 8, pp. 179–187, 1962.

[5] J. Triesch and C. von der Malsburg, "Robust classification of hand postures against complex backgrounds," in *Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 1996, pp. 170–175.

[6] J. Schlenzig, E. Hunter, and R. Jain, "Vision-based hand gesture interpretation using recursive estimation," in *Proc. 28th Asilomar Conf. Signals, Systems, and Computer*, 1994.

[7] W. Freeman and M. Roth., "Orientation histogram for hand gesture recognition," in *Int'l Workshop on Automatic Face- and Gesture-Recognition*, 1995.

[8] A. Caplier, L. Bonnaud, S. Malassiotis, and M. Strintzis, "Comparison of 2d and 3d analysis for automated cued speech gesture recognition," in *SPECOM*, Sep. 2004.

[9] T. Ahmad, C. J. Taylor, A. Lanitis, and T. F. Cootes, "Tracking and recognising hand gestures, using statistical shape models," *Image and Vision Computing*, vol. 15, no. 5, pp. 345–352, May 1997.

[10] M. Kolsch and M. Turk, "Robust hand detection," in *Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2004.

[11] E. Persoon and K. Fu, "Shape discrimination using fourier descriptors," *IEEE trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 3, pp. 388–397, 1986.

[12] D. Zhang and G. Lu, "A comparative study on shape retrieval using fourier descriptors with different shape signatures," in *Int. Conf. on Intelligent Multimedia and Distance Education*, 2001.

[13] C. Wah Ng and S. Ranganath, "Real-time gesture recognition system and application," *Image and Vision Computing*, vol. 20, no. 13-14, pp. 993–1007, Dec. 2002.

[14] R. Poppe and M. Poel, "Comparison of silhouette shape descriptors for example-based human pose recovery," in *Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2006, pp. 541–546.

[15] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden markov models," *Image and Vision Computing*, vol. 21, no. 8, pp. 745–758, Aug. 2003.

[16] N. Otsu, "A threshold selection method from gray level histograms," *IEEE trans. on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, Mar. 1979.

[17] D. Chai and K. Ngan, "Face segmentation using skin-color map in videophone applications," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, 1999, pp. 551–564.