Vision-based Hand Gesture Recognition

Hong Zhou

March 29, 2000

Abstract

We present a new gesture recognition method based on linear models. The idea comes from the linear models in tracking. Three kinds of experiments based on manual segmentation, color-based segmentation and without segmentation were made. By comparing with the eigenspace method, it showed that in most cases the linear method can be comparable with the eigenspace method. The model construction and image segmentation are the two main problems involved in recognition. Since hand gestures are a fusion of temporal and spatial properties, combining gestural motion could be a promising direction.

1 Introduction

As the computing, communication and display technologies progress even further, more natural and efficient human-computer interfaces (HCI) are tremendously addressed. With the motivation, there has been rapid growth of various studies related to vision-based gestural human-computer interfaces. Hand gestures are a natural form in interactive and communication. It has being explored to provide an attractive alternative to cumbersome interface device for HCIs. Recently, there has been a significant amount of research on vision-based hand gesture recognition (See surveys [5], [4]).

The first phrase of hand gesture recognition is choosing models of gestures. The models may consider both the spatial and temporal characteristics of the hand and hand gestures. There are two main kinds of models, 3D models and appearance-based models. 3D models can be classified in two large groups, volumetric models and skeletal models. The second group of models is based on appearance of hands in the visual images. A large variety of models belong to this group. Some are based on deformable 2D templates of the human hands. Some use 2D hand image sequences as gesture template. The majority of appearance-based models, however, use parameters derived from images in the templates, such as contours, edges, image moments, image eigenvectors and orientation histograms. Another groups of models uses fingertip positions as parameters.

Once the models are decided, a gesture analysis stage is used to compute the model parameters from the image features. These parameters consist some description of the hand pose or trajectory. Hand localization, hand tracking, and feature extraction are some important problems involved in the analysis.

The computation of model parameters is followed by gesture recognition. Here, the parameters are classified and interpreted in the light of the accepted model and perhaps the rules imposed by some grammar. Evaluation is based on if a particular gesture recognition approach encompasses both accuracy, robustness, and speed, as well as the variability in the number of different classes of hand movement it covers.

The problem of recognition can be solved by some parameter clustering techniques. Considering that gesture actions involve both the temporal and the spatial context, Hidden Markov Models (HMM) has been introduced in hand gesture recognition. The time-space context of any specific gesture can also be established by a grammatical element into recognition procedure.

In literature, there are many reports in hand gesture recognition, addressing different problems. For example, in [1], they introduced a learning based prediction-and-verification segmentation scheme for hand sign image sequence. They used 19 fixations with different position and size for each hand shape to complement the motion-based segmentation. The motion-based visual attention and masks were used to get the initial hand contour, then a dynamic deformation model were introduced to get a better contour.

In [6], they introduced RIEVL:recursive induction learning in hand gesture recognition. They use the color segmentation approach to segment the hand from the background. Twenty-eight features were extracted from the segmented hand poses, such as, the area of the bounding box, the compactness of the hand, the centroid of the hand, the central moments, the principal axes, the normalized moments and the rotation invariants. A number of conjunctive rules for each class were produced by RIEVL. In total, there were 20 classes, and 42 conjunction rules were generated by RIEVL.

In [3], they used several detection processes for tracking hands, image differencing from successive images, image differencing from background and skin detection. The confident factor of visual processes were introduced for control the three processes.

Although the current progress in gesture recognition is encouraging, further theoretical as well as computational advances are needed.

Here we introduce a different gesture recognition method based on linear models. The idea comes from the linear model in tracking [2]. It is a natural way to link the tracking and recognition in gesture recognition since the gesture recognition is a fusion of temporal and spatial properties. We tested linear recognition models on the manually segmented images and the color-based segmented images and images without segmentation. By comparing with the methods based on image eigenspaces, it showed that in most cases the linear method can be comparable with the eigenspace method. The eigenspace has an advantage of feasible to search a large database of images. However an advantage of linear method is potentially capable of considering the continuity of motion. Since hand gestures have both temporal and spatial properties, linear methods could be a promising direction.

2 Linear Models for Gesture Recognition

Let us model the image motion of the gesture by a parametric motion model $f(p;\mu)$ parameterized by $\mu = (\mu_1, \mu_2, ..., \mu_n)^T$, the location $p = (x, y)^T$ in an image has the brightness value $I(x, y)$. Let the set $R = p_1, p_2, ..., p_N$ be a set of N image locations which define a gesture.

A database of gesture models $IM = I_1, I_2, ..., I_M$ are normalized in size: that is the image frame is minimum rectangle enclosing the largest appearance of the gesture and the energy of each model is normalized to 1.

The correlation between the image and the models can be computed as the distance δI.

$$\begin{aligned}
\delta I &= ||I - I_i|| \\
\delta I &= M \delta \mu \\
M &= [I_x | I_y | I_r | I_s] \\
I_x &= I(x, y) - I(x-1, y) \\
I_y &= I(x, y) - I(x, y-1) \\
I_r &= -y I_x + x I_y \\
I_s &= \frac{\mathbf{1}}{\sqrt{\mathbf{x^2+y^2}}}(x I_x + y I_y)
\end{aligned} \tag{1}$$

Where $I_i \in IM$, $p(x,y) \in R$, I_x, I_y, I_r, I_s are N-dimensional vectors, and M is a 4xN matrix.

After simple transformation, the distance can be expressed as:

$$\delta I = M(M^T M)^{-1} M^T \delta I \tag{2}$$

The recognized gesture is the model with the minimizing distance:

$$||\delta \mathbf{I} - \mathbf{M}(\mathbf{M^T M})^{-\mathbf{1}} \mathbf{M^T} \delta \mathbf{I}||_{min} \tag{3}$$

3 Experiments

The main problems related to the linear recognition are model construction and image segmentation. We tested three kinds of experiments based on manual segmentation, color-based segmentation and without segmentation.

3.1 Experiments on manually segmented gestures

Procedure:

- Construct the template gestures with predefined size (i.e. 30x15) and normalized intensity
- Capture color images.
- Get the bounding box of gestures in the images manually.
- Extract the gestures in the bounding box.
- Resize and normalize the extracted gestures.
- Compute the distance as function (3).

The template models are shown as figure1, including five gestural models. The manually segmented gestures are shown as figure2-4. The recognition results are shown as table1-3. From the experimental results, it showed that the linear models are efficient when the extracted

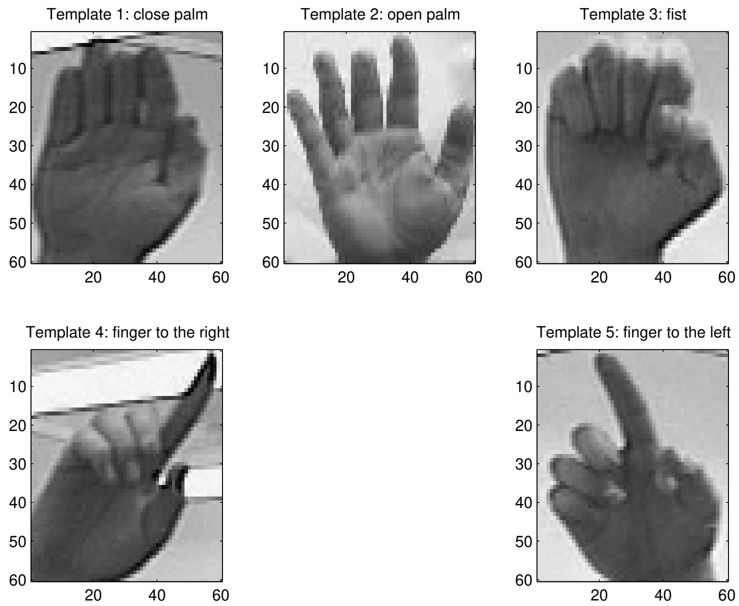

Figure 1: Gestural models in experiment1

gestures have trivial shape variation compared to the corresponding template.

3.2 Experiments on the color-based segmented gestures

Procedure:

- Construct the template gestures with predefined size (i.e. $30x15$) and normalized intensity
- Capture color images.
- Segment the images using color-based segmentation.
- Extract the gestures in bounding boxes.

Image No	Close palm	Open palm	fist	finger to the right	finger to the left
1	4.2065	6.1133	6.3760	7.7478	7.2756
2	5.2884	7.1142	6.3720	7.0914	6.6033
3	4.0645	6.9486	6.3571	7.9679	7.4749
4	5.0386	6.5357	6.3532	7.9947	7.4611
5	3.9248	6.4498	6.1939	8.0044	7.4680
6	5.9840	6.7410	6.3625	7.4547	6.3831
7	6.0940	7.1595	6.2459	7.4421	6.4872
8	6.0940	7.1595	6.2459	7.4421	6.4872
9	5.4029	6.7241	6.2226	7.7139	6.3133
10	5.7001	7.2800	6.0814	7.1342	6.6412
11	3.7515	6.4772	6.2931	7.9203	7.4960
12*	6.2681	7.4385	6.1639	6.7574	7.0629
13	4.7323	6.9448	6.4073	7.9826	7.5185
14	5.6863	6.8212	5.8379	7.2379	6.3858
15	0	6.8959	6.2846	7.9946	7.1183
16*	7.4661	8.3027	5.6867	6.2014	7.0643

Table 1: Experimental results of palm gestures in figure2 based on the templates in figure1. * denotes the uncorrect recognition.

Image No	Close palm	Open palm	fist	finger to the right	finger to the left
1	5.8454	6.8625	5.1486	6.7827	6.4689
2	6.6113	6.7245	5.2218	5.5360	5.4771
3	5.1530	6.2991	1.6804	7.0475	6.6768
4	5.0924	6.5379	2.5941	7.0395	6.7588
5	5.4769	6.7254	5.1694	6.8275	6.5253
6	5.5489	6.4147	4.8834	6.6913	6.4345
7	5.7919	6.9142	5.3362	6.9585	6.6889
8	5.2580	6.9022	3.8241	6.9831	6.8627
9	6.8555	5.5308	4.8026	6.5621	6.8432
10*	6.3558	6.9632	5.9068	5.7176	6.237
11	5.3040	7.1415	4.8744	6.5440	6.5410
12*	6.6269	7.6929	6.4682	5.8090	6.3228
13	5.3898	6.5805	4.9953	6.8575	6.3482
14	6.0588	7.0362	5.7262	6.9391	6.8021
15	6.0982	7.5269	5.7638	6.1906	6.5010
16	5.7879	6.6769	5.0294	6.8410	6.5226

Table 2: Experimental results of fist gestures in figure3 based on the templates in figure1. * denotes the uncorrect recognition.

Image No	Close palm	Open palm	fist	finger to the right	finger to the left
1	8.4236	8.2654	7.1560	6.1824	6.3275
2	7.8942	7.0056	6.6098	7.4429	4.3751
3	7.4830	8.2414	6.8001	4.5790	6.9396
4	8.5907	8.2052	7.7714	7.9930	6.8295
5	7.4556	6.8384	6.0138	7.2327	4.6024
6	7.3678	6.5562	5.9742	6.9424	5.0436
7	8.3646	8.9423	8.1579	5.6897	7.3975
8	8.1260	6.5136	6.8628	7.0893	5.3176
9	8.3380	8.8139	7.8792	5.6905	7.7855
10	8.0947	7.1819	6.9609	7.6335	5.2885
11	8.1650	8.7704	8.0567	4.9419	7.8740
12	8.1097	6.6442	6.6932	7.3713	4.6565
13	8.2196	6.6454	6.4585	7.4319	1.5397
14	7.5299	6.4488	6.2736	7.0136	5.2833
15	7.5584	8.6700	7.6090	5.5911	7.9593
16	7.6094	8.1174	6.9625	2.0389	7.4042

Table 3: Experimental results of point gestures in figure4 based on the templates in figure1.

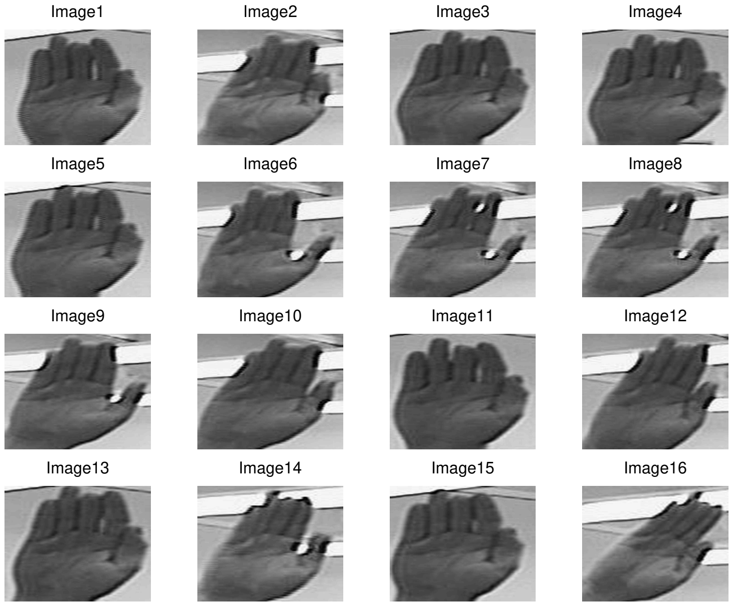

Figure 2: Manually segmented palm gestures in experiment1

- Resize and normalize the extracted gestures.
- Compute the distance as function (3).

The template models are shown as figure5, including six gestural models. The color-based segmented gestures are shown as figure6-8. The recognition results are shown as table4-6. From the experimental results, it showed that the linear models are efficient to recognize the palm and fist. Because some figures were cut in the color-based segmentation, some points were recognized as palm or fist.

3.3 Experiments on comparing PCA and Linear Models

Procedure:

- Choosing training set to construct the templates for PCA.

Image No	palm	palm to the right	fist	fist to the right	finger to the right	finger to the left
1	5.8188	* 5.4671	6.9139	6.6149	7.3336	7.3023
2	* 5.5317	7.9174	6.0137	7.6799	8.4380	8.0460
3	4.4894	7.2694	5.6357	7.2249	8.0796	7.4934
4	7.0480	* 1.8934	7.2275	6.4988	7.2042	8.0454
5	* 4.8605	7.3476	6.3785	7.0700	7.9912	7.6321
6	* 5.3409	7.1937	6.4888	6.9034	7.7024	7.5596
7	* 4.3346	7.1120	5.9164	6.9796	8.2972	8.0565
8	* 5.3119	7.8420	6.4714	7.9167	8.4029	7.7765
9	7.2474	* 3.2891	7.5745	6.9070	7.6588	8.1973
10	7.1439	* 4.4931	7.3423	6.3072	7.5781	8.2385
11	7.1439	* 4.4931	7.3423	6.3072	7.5781	8.2385
12	* 5.0531	8.0026	6.8899	8.1649	8.3105	7.7036
13	7.2522	* 3.3259	7.7977	6.9989	7.3202	8.1964
14	* 2.4460	* 7.5232	5.9311	7.4357	8.1562	7.6522
15	7.5082	* 3.4523	7.4916	5.8264	7.0951	8.4223
16	* 5.8128	5.8214	6.8250	6.7173	7.3746	7.4662

Table 4: Experimental results of palm gestures in figure6 based on the templates in figure5. * denotes the recognized gesture.

Image No	palm	palm to the right	fist	fist to the right	finger to the right	finger to the left
1	7.3617	6.1058	6.8497	* 4.3826	6.8951	9.0518
2	7.4798	6.1448	6.5881	* 4.6410	7.0581	9.1620
3	5.9354	7.2373	* 4.1449	7.0188	8.2475	8.1360
4	5.7729	7.5288	* 2.1273	6.8756	8.5187	8.0764
5	7.5895	6.1072	6.7051	* 4.8051	7.0560	8.8572
6	6.6871	6.2910	6.2190	* 4.3901	7.1800	8.9472
7	7.6516	5.6312	7.0765	* 5.2514	6.6744	8.7690
8	6.1341	7.5802	* 4.5533	6.9916	8.2772	8.1012
9	5.7784	6.9152	* 5.6810	7.0964	7.8384	7.4474
10	7.4301	6.2646	6.7530	* 1.7469	7.0910	9.0454
11	6.8459	6.2460	5.6671	* 4.8972	7.5133	9.0400
12	7.3887	6.2468	7.7971	* 6.1229	6.3045	7.5239
13	7.1027	6.1398	6.8143	* 4.6776	7.5520	8.8110
14	7.7602	6.4310	7.5407	* 6.3498	7.2248	8.1880
15	7.0919	6.2371	6.8434	* 4.7726	7.4450	8.9120
16	7.4193	5.8751	6.5156	* 5.0538	7.0961	8.7016

Table 5: Experimental results of fist gestures in figure7 based on the templates in figure5. * denotes the recognized gesture.

Image No	palm	palm to the right	fist	fist to the right	finger to the right	finger to the left
1	7.1326	7.7620	7.3116	8.3661	7.2165	* 5.0864
2	7.5441	6.7219	8.0009	* 5.9754	6.2070	7.8574
3	8.7861	* 7.5894	8.6575	7.9297	8.4859	8.1442
4	6.8488	7.9968	* 6.4536	8.3386	7.8362	6.8729
5	7.1359	8.2500	* 6.3865	8.3481	8.2032	7.0004
6	8.1392	6.3253	7.3957	* 4.2132	7.1271	9.3030
7	7.0561	7.6590	7.2679	7.9352	7.7624	* 5.8190
8	7.7599	6.1471	7.1868	* 4.6410	6.8727	8.7609
9	6.7154	8.2250	* 6.2394	8.3880	8.3511	7.1665
10	7.1200	7.5477	7.0613	8.0176	7.1506	* 6.2685
11	7.1200	7.5477	7.0613	8.0176	7.1506	* 6.2685
12	6.7608	7.2261	* 6.3250	7.7427	7.2648	6.6728
13	* 6.0603	7.0228	6.2030	7.2574	7.7483	7.5497
14	6.5311	6.5571	* 6.4432	6.7911	7.9104	8.0092
15	7.2524	7.4805	7.2737	8.1159	7.0423	* 5.7561
16	7.4880	* 5.3037	6.9221	5.5857	7.0552	8.3291

Table 6: Experimental results of point gestures in figure8 based on the templates in figure5. Because the figures were cut in most segmented gestures, some points were confused with fist or palm. * denotes the recognized gesture.

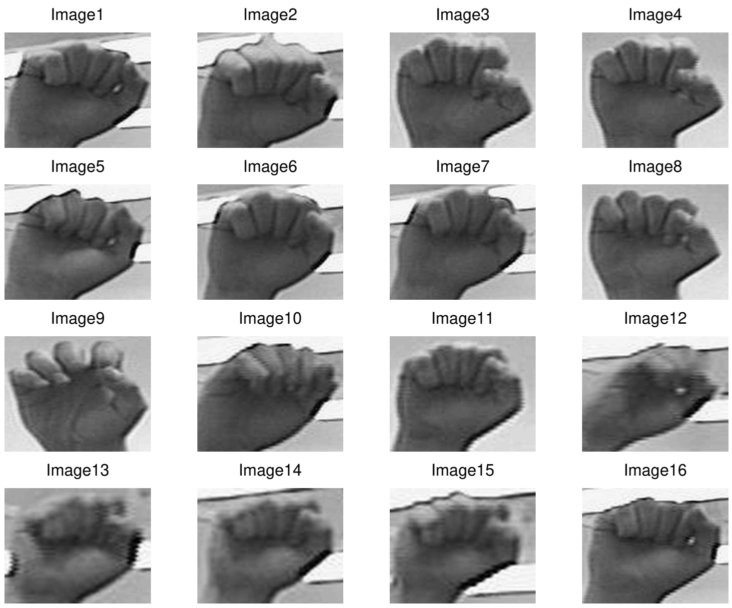

Figure 3: Manually segmented fist gestures in experiment1

- Select the template gestures for Linear Model.
- Resize and normalize the testing set.
- Recognize based on PCA.
- Recognize based on Linear Models.

The templates for Linear Models are shown as figure9, including three gestural models, palm, fist and open palm. The sample image sets in experiment 3 are shown as figure 10-15. The images only include gestures and simple background. No segmentation procedure is required. The experimental results are shown in table 7. From the experimental results, it shows that the linear models can be comparable with the PCA method. With training, PCA is more robust than linear models with respect to the templates. But linear models contain shape variation and the templates are easy to change. How to reduce

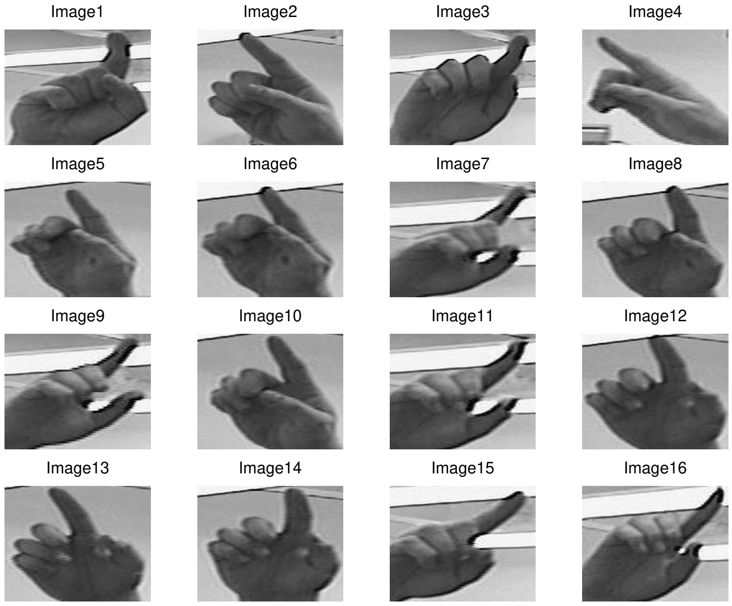

Figure 4: Manually segmented point gestures in experiment1

the template set in linear model will be an important problem.

4 Conclusion

A new hand gesture recognition method based on linear models is introduced in this paper. Three kinds of experiments based on manual segmentation, color-based segmentation and without segmentation were made. By comparing with the PCA method, it showed that the linear method is more sensitive to the templates than PCA method but in most cases, it can be comparable with PCA. Two difficulties in linear methods are template selection and segmentation. To improve the system, fusing the gestural temporal and spatial properties could be a promising direction.

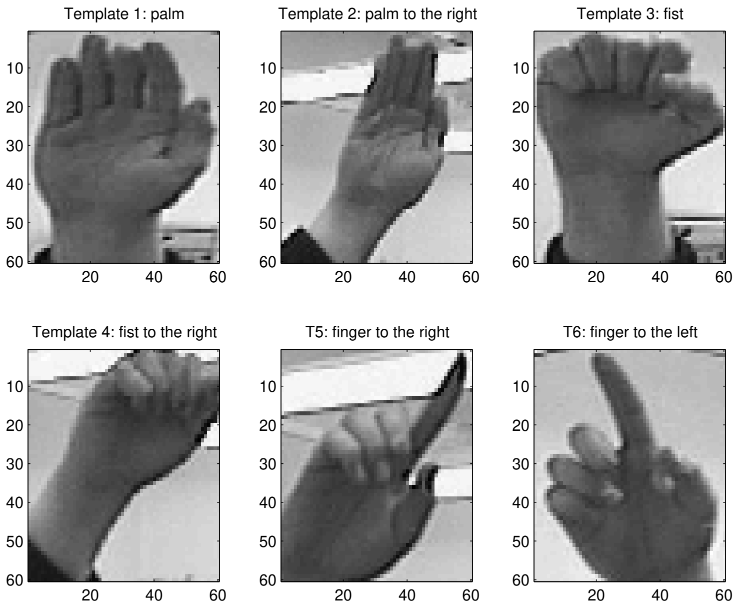

Figure 5: Gestural models in experiment2

The number of test images	PCA Recognized images	Linear Models Recognized images
45	45	41
102	102	94
156	156	134
171	165	155
264	264	235

Table 7: Comparison between PCA and Linear Method. The training set are randomly selected from the whole set (with 300 images). The left images are the testing set. Sample images are shown in figure 10-15.

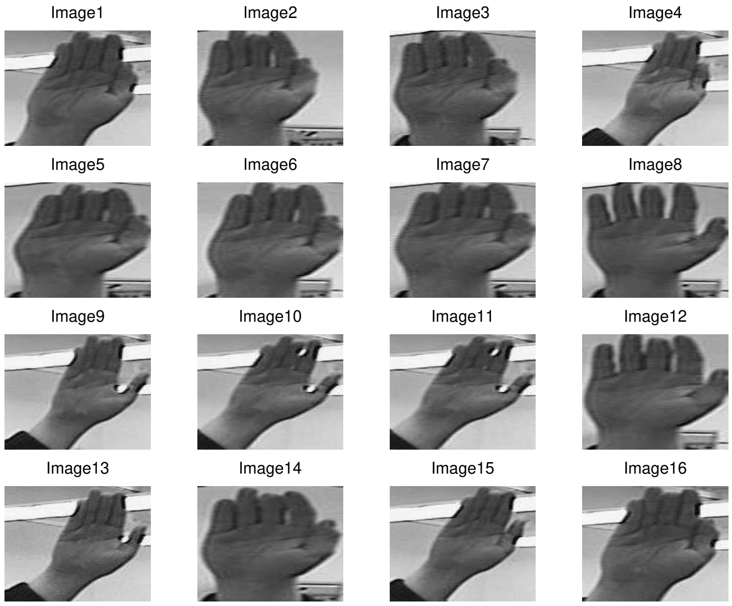

Figure 6: Color-based segmented palm gestures in experiment2

References

[1] Yuntao Cui and Junyang Weng. A learning-based prediction-and-verification segmentation scheme for hand sign image sequence. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 21(8):798–804, 1999.

[2] Gregory D. Hager and Peter N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 20(10):1125–1139, 1998.

[3] Jérôme Martin, Vincent Devin, and James L. Crowley. Active hand tracking. In *International Workshop Aotomatic Face and Gesture Recognition*, pages 573–578, 1998.

[4] Vladimir I. Palvlovic, Rajeev sharma, and Thomas S. Huang. Vi-

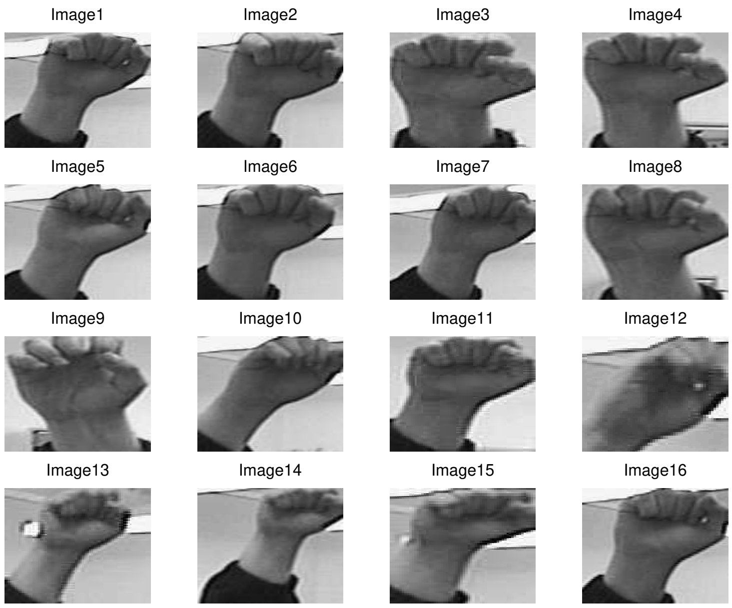

Figure 7: Color-based segmented fist gestures in experiment2

sual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 19(7):677–695, 1997.

[5] Richard Watson. A survey of gesture recognition techniques. In *Department of Computer Science, Trinity College, Dublin, Technical Report TCD-CS-93-11*, 1993.

[6] Maide Zhao, Francis K.H. Quek, and Xindong Wu. Rievl: Recursive induction learning in hand gesture reognition. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 20(11), 1998.

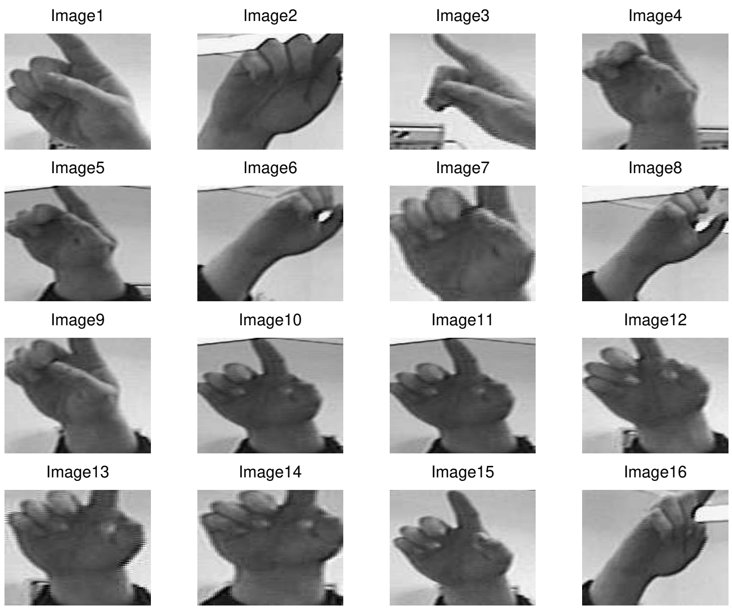

Figure 8: Color-based segmented point gestures in experiment2

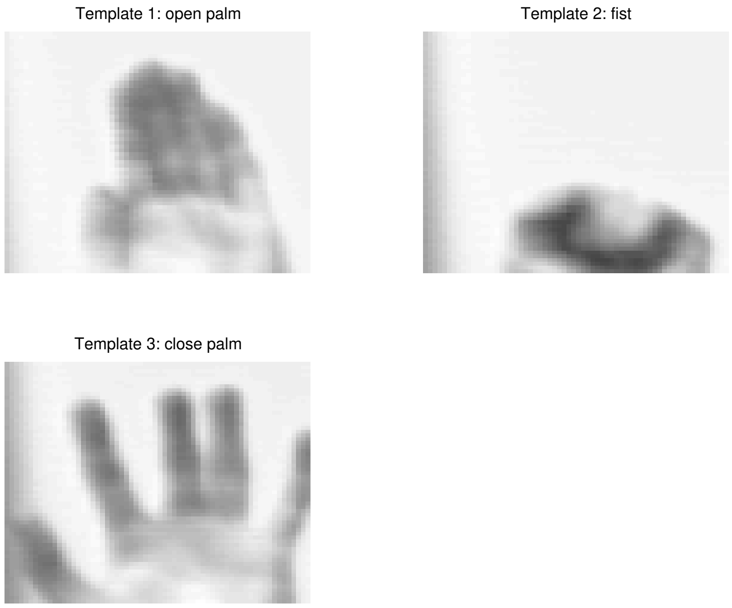

Figure 9: Gestural models in experiment3

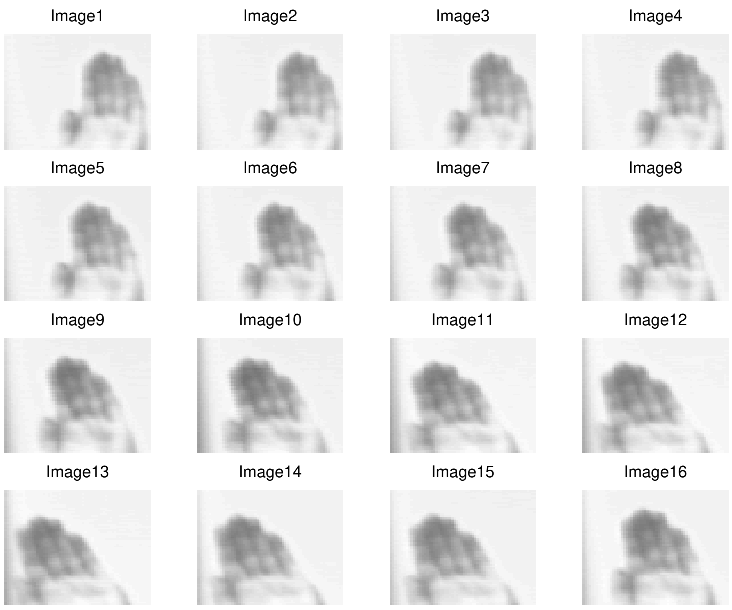

Figure 10: Sample palm gestures in experiment3

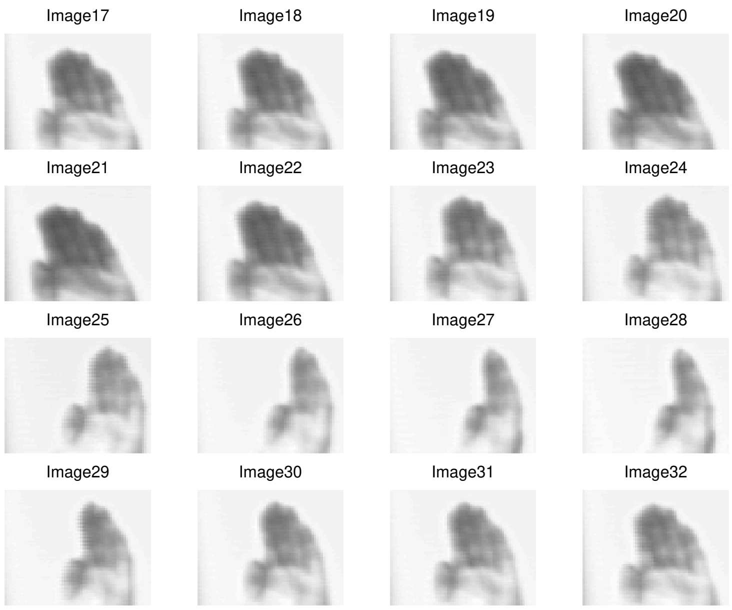

Figure 11: Sample palm gestures in experiment3

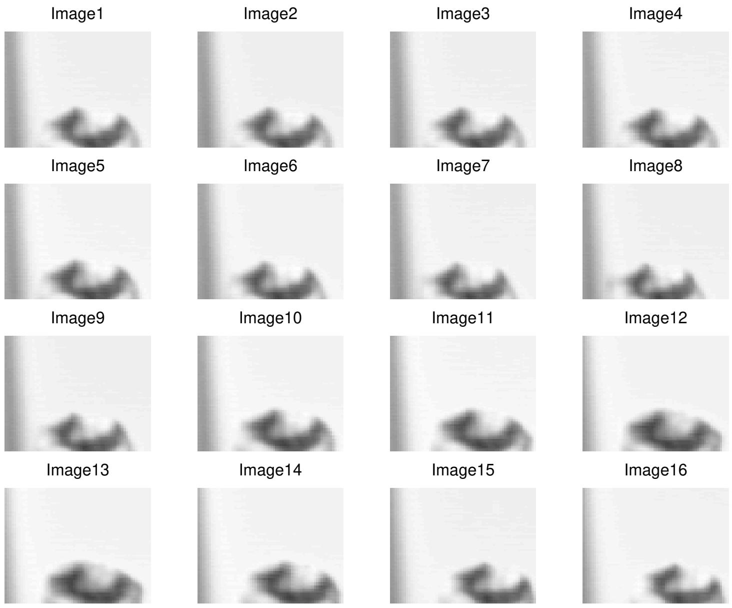

Figure 12: Sample fist gestures in experiment3

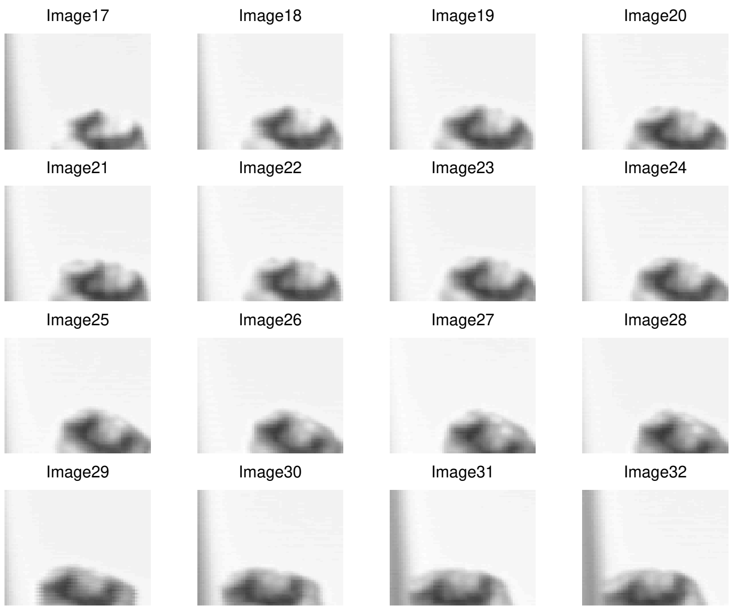

Figure 13: Sample fist gestures in experiment3

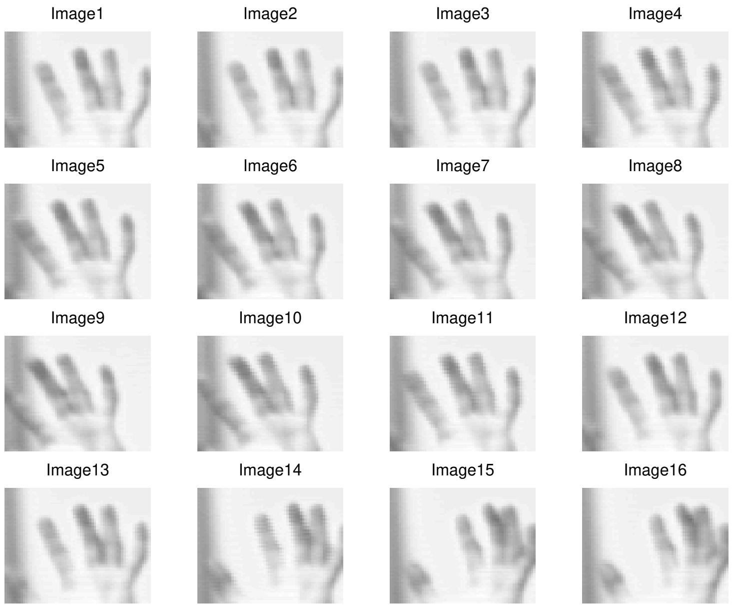

Figure 14: Sample open palm gestures in experiment3

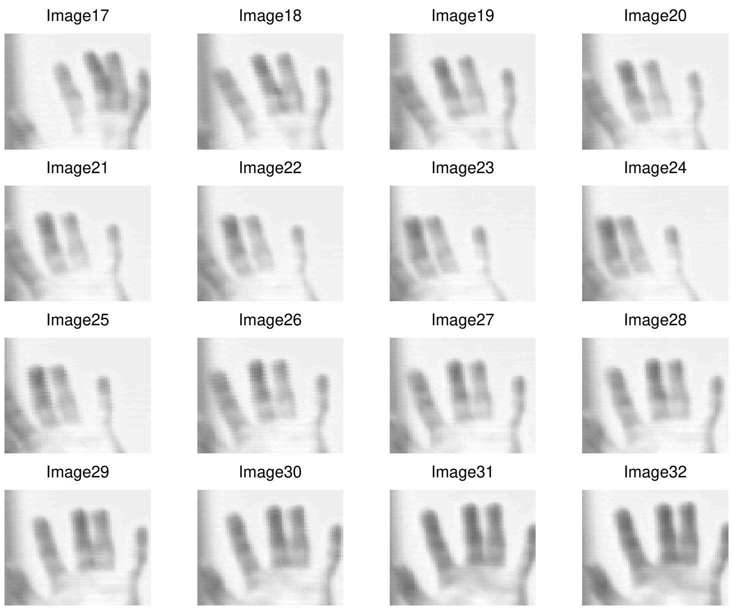

Figure 15: Sample open palm gestures in experiment3