MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

Full-Passive Human Recognition from Image Sequences

Satoshi Abe†, Kozi Nakamura†, Mamoru Maekawa††, Takanobu Endo††, and Nobuyuki Sugiura††

† Department of Information Science, Faculty of Science, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113 Japan

†† Graduate School of Information Systems, University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182 Japan

Abstract

The prototype of the system that recognize the specified person from video image sequence was implemented. This system aims at the full-passive human recognition in home. It is therefore designed so that the influence of the experiment on the subjects' life should be minimum. This paper presents the overview of our approach including the system design and the general strategy of the recognition.

1. Introduction

Human recognition using image has attracted many researchers because of the simplicity and the difficulty of the problem. Large number of efforts has been made on this subject: some tried to identify the fingerprints or hand-written characters and others tried to recognize the face in the still image.

We also cope with this challenging theme but in a slightly different way; our goal is to develop the practical system that recognizes one or more persons in full-passive way. Though our primary application is the investigation of human action in the home, this system is also applicable to security systems and other experiments for psychological or social study.

As experiments, we took videos of ten families, and we digitized some parts of them for the analysis. Our conditions for the primary experiment are as follows:

a. The system should recognize less than ten people, that seem enough for the family.

b. Videos should be taken under no special lighting in order not to affect the subjects' life.

c. Viewing angle is wide enough to see the whole room.

d. The focus and zoom are fixed throughout the recording.

e. Videos are recorded with VHS format.

f. Video images are captured and digitized in 512pixels x 512pixels x 24bits.

g. The digitized images are analyzed on the SPARC-station 2 with the C language.

2. Experiments

2.1 General Strategy

Since observed people may change their clothes, the recognition should be performed with parts of them that is not affected by their clothes. There are two parts that satisfy this condition: the face and the hands. Hands are however smaller in area and have less characteristics than faces. Thus, we chose the face as the target of recognition.

Comparing with many previous works on the face recognition, the quality of face images are low [Harman77] [Shackleton91] [Turk90]. The orientation of face, the distance from the camera, the lighting condition change depending on the situation. The recognition rate from still image cannot be higher than the recognition from photograph based system.

On the other hand, our experiment has two advantages:

a. Image sequences can be used for the recognition.

b. The number of recognition targets is limited.

Our recognition strategy is based on these two points. Though the recognition rate is not so high in the single recognition, we can make it higher if we track the human figures and try to recognize them repeatedly. Fig. 1 shows the general structure of our recognition system. From the next section, we describe the method to utilize image sequences for the effective human recognition.

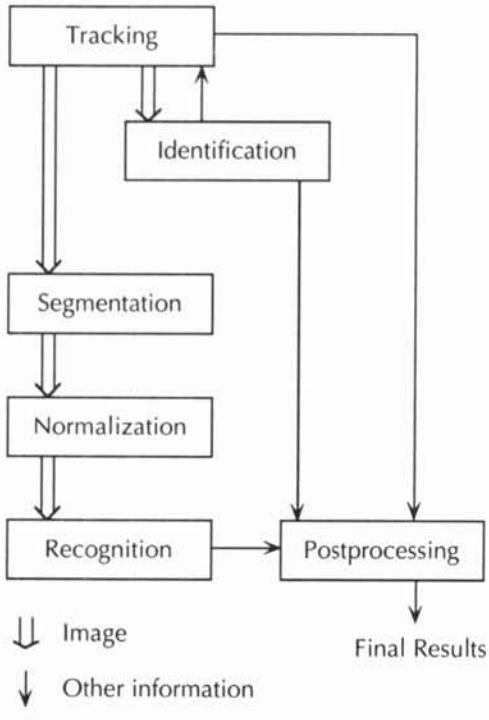

Fig.1 General Structure of Recognition

Fig.2 Subject Image

2.2 Human Figure Tracking

Fig. 2 shows an example of subject image. We start to track human figures by detecting moving objects with the image subtraction between different frames. First, difference between two frames of image are computed and the pixels where the difference exceeds the certain threshold are marked. Multiple objects are separated by testing continuity of the marked pixels.

Even if two objects are overlapped or one is occluded by another, we have to identify each object. We realize this using two properties. One is the distribution of the color. The colors of pixels contained in the traced objects are re-expressed in HVC model. When overlapping is over, the distributions of the hue and the value of two objects are compared with the those of the objects before overlapping using chi-square goodness-of-fit test. This method is also applicable to identify the person who goes out of the frame. The other property is the velocity of the objects [Limb75] [Fennema79] [Aggarwal88]. Since human acceleration is limited, we can track the two object when both are moving. This inference is particularly effective when the two objects move to the opposite direction.

2.3 Face Segmentation

Human faces can be segmented by detecting oval area in constant hue in upper part of the moving object. If the hue of the face and the hair of each recognition target is known in advance, they can help the segmentation.

If no face-like objects are found in the tracked object, there are the following two possibilities:

a. The object is not a person to be recognized.

b. The face cannot be seen from the camera.

In these case, we dare not segment a face candidate but just track that object.

2.4 Face Normalization

After segmenting faces, we look for the eyes and the mouth for the spatial normalization of face images including the position, the orientation, and the size using the hue and the value. Considering that the depth of the human figures are smaller than the distance between the camera and the recognition object, the transformation does not have to be perspective. Affine transform is therefore sufficient for the spatial normalization. Since three corresponding points are necessary to determine an affine transform uniquely,

we chose the inner corners of the eyes and the upper-center of the mouth. The left image of Fig. 3 shows the spatially normalized image of a face by finding these three points in the face image.

However, finding the mouth in low quality face images is sometimes far more difficult than finding the eyes because a face has only one mouth and the brightness of mouse is relatively similar to the brightness of other area on the face. We also try two-point transform to cope with the situation that we cannot find the mouth. In that case, the transform includes scaling and translate but not rotation. The right image of Fig. 3 shows this type of spatial normalization.

In our experiments of spatial normalization, the output image size is 64x64 and the bi-linear interpolation is used to resample images.

Fig.3 Normalized Face Image

Left: Normalized by Three Points Right: Normalized by Two Points

2.5 Face Recognition

Faces are recognized by K-L expansion. Table 1 shows a result of recognition of the scene shown in Fig. 1. The table shows three proposed results in its probability order with their distance value to the model when the specified number of eigenvectors are used. As this example shows, eight or more eigenvetors are required for the accurate recognition, generally.

2.6 Postprocessing

Postprocessing plays an important role in the improvement of the recognition rate. For each combination of the tracked objects and the models, we compute the value called *demerit* that reflects total invalidity of the combination. The demerit is determined by the following factors:

a. The distance between the model and the tracked object for several times of recognitions.

b. The probability of the assumption that the object is correctly tracked.

The final result of recognition is determined so that the summation of the demerits of all combination should be minimum.

a test data of A												
	1 EV			5 EVs			12 EVs			31 EVs		
1	***A***	(1)	56	***A***	(3)	1006	***A***	(3)	1328	***A***	(3)	1576
2	B	(1)	65	C	(2)	1316	C	(7)	1762	C	(2)	2136
3	B	(3)	108	C	(7)	1444	C	(2)	1803	C	(5)	2219
a test data of B												
1	A	(1)	52	***B***	(1)	1414	***B***	(5)	1985	C	(1)	2737
2	***B***	(1)	61	***B***	(5)	1515	***B***	(2)	2119	***B***	(2)	2742
3	***B***	(3)	113	***B***	(2)	1577	***B***	(1)	2299	***B***	(5)	2754
a test data of C												
1	D	(1)	72	***C***	(5)	252	***C***	(5)	510	***C***	(2)	959
2	***C***	(2)	72	***C***	(2)	338	***C***	(2)	628	***C***	(5)	1000
3	***C***	(5)	106	***C***	(1)	625	***C***	(1)	846	***C***	(1)	1136
a test data of D												
1	C	(8)	30	***D***	(1)	595	***D***	(1)	1107	***D***	(1)	1376
2	A	(3)	37	***D***	(8)	740	***D***	(3)	1115	***D***	(8)	1586
3	C	(6)	83	***D***	(3)	749	***D***	(8)	1380	***D***	(3)	1647

(a) Three-Point Normalization

a test data of A												
	1 EV			5 EVs			12 EVs			31 EVs		
1	C	(1)	28	C	(2)	559	***A***	(3)	956	***A***	(3)	1224
2	D	(4)	43	***A***	(3)	710	C	(2)	1164	C	(2)	1312
3	D	(3)	49	C	(3)	787	C	(5)	1198	C	(5)	1366
a test data of B												
1	A	(1)	15	C	(6)	1454	***B***	(2)	1684	***B***	(2)	1991
2	***B***	(4)	20	A	(1)	1561	***B***	(1)	1776	C	(6)	2052
3	***B***	(3)	53	***B***	(2)	1572	C	(6)	1785	C	(1)	2127
a test data of C												
1	D	(7)	36	***C***	(5)	286	***C***	(5)	377	***C***	(5)	604
2	A	(7)	120	***C***	(3)	325	***C***	(3)	602	***C***	(3)	992
3	D	(2)	168	***C***	(2)	496	***C***	(7)	706	***C***	(7)	1125
a test data of D												
1	***D***	(8)	16	***D***	(8)	528	***D***	(8)	1236	***D***	(8)	1431
2	B	(4)	80	***D***	(1)	705	***D***	(3)	1263	***D***	(3)	1530
3	A	(1)	116	***D***	(3)	898	***D***	(4)	1446	***D***	(1)	1730

(b) Two-Point Normalization

Table 1 Result of Recognition by K-L Expansion with Various Number of Eigenvectors
Eight models are used for each person.
Three potential answers are shown in order of probability.
Each cell contains the person name (image ID) and the distance to the model.
Person names are italicized if correct.

Since the experiment is still in the first stage, we have not yet made sufficient number of examinations. Thus we cannot precisely discuss the recognition rate in the current situation.

3. Remaining Problems and Future Works

There are still many remaining problems in our experiment including the followings:

a. Sometimes, we cannot see faces of some people because they do not face the camera or their face is hidden by things like a newspaper or a magazine.

b. The lighting condition varies much. For example, there is a large difference between the lighting of the daytime and that of the evening.

c. The effect of the hair style and the glasses, the change of face color such as the tanning or the make-ups should be carefully examined to realize the practical system.

d. Sometimes three or more people overlaps in an image. We have not yet deal with such cases.

4. Concluding Remarks

Experiment system that recognizes a person in the room from video image was implemented. It consists of 5 steps: figure tracking, segmentation, normalization, recognition, and postprocessing. Though there rest several problems for precise discussions, experiments show fairly good result.

References

[Aggarwal88]
Aggarwal, J. K. and N. Nandhakumar: "On Computation of Motion from Sequences of Images – A Review," Proc. of IEEE, Vol.76, No.8, pp.917-935, 1988

[Ballard82]
Ballard, D. H. and M. B. Christopher: *Computer Vision*, Prentice-Hall, New Jersey, 1982.

[Burt89]
Burt, P. J, J. R. Bergen, et al.: "Object Tracking with a Moving Camera: An Application of Dynamic Motion Analysis," Proc. of IEEE Workshop on Visual Motion, pp.2-12, 1989.

[Davis83]
Davis, L. S., Z. Wu and H. Sun: "Contour-Based Motion Estimation," CVGIP, Vol.23, pp.313-326, 1983.

[Dreschler82]
Dreschler, L. and H. Nagel: "Volumic Model and 3D-Trajectory of a Moving Car Derived from Monocular TV Frame Sequences of a Street Scene," CGIP, Vol.20, No.3, pp.199-228, 1982.

[Fennema79]
Fennema, C. L. and W. B. Thompson: "Velocity Determination in Scenes Containing Several Moving Objects," CGIP, Vol.9, pp.301-315, 1979.

[Harmon77]
Harmon, L. D. and W. F. Hunt: "Automatic Recognition of Human Face Profiles," CGIP Vol.6, No.2, pp.135-156, 1977.

[Harmon78]
Harmon, L. D.: "Identification of Human Face Profiles by Computer," Pattern Recognition Vol.10, No.5, pp.301-312, 1978.

[Healey89]
Healey, G.: "Using Color for Geometry-Insensitive Segmentation," Journal of the Optical Society of America, Vol.6, No.6, pp.920-937, 1989.

[Huang90]
Huang, T. S.: "Modeling Analysis and Visualization of Nonrigid Object Motion," Proc. of ICPR, Vol.1, pp.361-364, 1990.

[Limb75]
Limb, J. O. and J. A. Murphy: "Estimating the Velocity of Moving Images in Television Signals," CGIP, Vol.4, pp.311-327, 1975.

[Murray87]
Murray, D.. W. and B. F. Buxton: "Scene Segmentation from Visual Motion Using Global Optimization," IEEE Trans. on PAMI, Vol.PAMI-9, No.2, pp.220-228, 1987.

[Nagel86]
Nagel, H. and W. Enkelmann: "An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences," IEEE Trans. on PAMI, Vol.PAMI-8, No.5, pp.565-593, 1986.

[O'Rourke80]
O'Rourke, J. and N. L. Badler: "Model-Based Image Analysis of Human Motion Using Constraint Propagation," IEEE Trans. on PAMI, Vol.PAMI-2, No.6, pp.522-536, 1980.

[Shackleton91]
Shackleton, M. A. and W. J. Welsh: "Classification of Facial Features for Recognition," Proc. CVPR '91, pp.573-579, 1991.

[Shio91]
Shio, A. and J. Sklansky: "Segmentation of People in Motion," IEEE Workshop on Visual Motion, pp.325-332, 1991.

[Takagi79]
Takagi, M. and K. Sakaue: "The Analysis of Moving Granules in a Pancreatic Cell by Digital Moving Image Processing," Proc. 4th IJCPR, pp.735-739, 1979.

[Turk90]
Turk, M. A., and A. P. Pentland: "Recognition in Face Space," Intelligent Robots and Computer Vision IX: Algorithms and TECHNIQUES, PROC. SPIE VOL. 1381, PP.43-54, 1990.

[TURK91]
TURK, M. A., and A. P. Pentland: "Face Recognition Using Eigenfaces," Proc. CVPR '91, pp.586-591, 1991.