Face Recognition by Incremental Learning*

Weimin Huang, Beng Hai Lee, Liyuan Li and Karianto Leman
Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
{wmhuang, benghai, lyli and karianto}@i2r.a-star.edu.sg

Abstract - *One of the important features for human machine interaction is its ability to recognize human faces. This paper presents a novel architecture suitable for real time robotic face recognition by learning a person's face incrementally, where the Gabor features at respective feature locations of a face are used to derive a similarity measurement. A face tracking followed by a clustering technique is used to learn a person's face appearance variance when the system interacts with the person. The recognition by learning proposed in this paper is similar to the partial memory incremental learning method, where we proposed a novel approach to the learning and updating process. Experiment shows significant improvement in the face recognition performance after learning over the time and with more interaction between a person and the system.*

Keywords: face recognition, incremental learning, Gabor feature representation, face tracking, clustering

1 Introduction

Face recognition methods have been studied intensively in literature intensively. Models include template and feature based methods [1,2,7,15]. However, many techniques require the environment to be well controlled, such as constant lighting condition and the faces to be frontal. For a mobile system such as a robot, the changing of the surroundings poses considerable challenges in order to make it able to recognize people around it in real time and robustly. To overcome this problem, several methods have been proposed in literature [10,3,16,12].

One such method is to normalize the face data in order to reduce the variance of an object, such as the lighting normalization and elastic graph matching. However, due to the large variance and distortion of the objects, it is sometimes difficult to find the proper means to normalize the data.

An alternative approach is to introduce more samples to form a good boundary between different objects. In the literature, multiple view based face recognition has been proposed to cope with the face pose problem [6], where the pre-captured multiple view faces are stored and clustered for later recognition. Another attempt is to use a face sequence instead of just one image for human recognition [3,10,11]. Further, Sim et al [12,13] and Vetter [14] have proposed synthetic sample generation methods for face recognition, which synthesized more samples of faces for recognition and achieved quite good performance.

In this paper we will address a new face recognition approach for human-robot interaction using the incremental learning method, motivated by the work presented in [5,3,10,12].

With a robust face detection and tracking method, we select and accumulate a person's face samples automatically. A method for learning and clustering is proposed to refine the classifier in use. With the accumulation of different appearances of a person over time, the classifier gradually improves its recognition rate.

The paper is organized as follows. Section 2 presents the new framework for incremental face recognition. Section 3 discusses the similarity measurement using Gabor features. Section 4 gives the experimental results. Finally the discussion and conclusion come in Section 5.

2 Framework for incremental face recognition

2.1 Incremental learning

Most of the face recognition systems available today use one or several face images as the registered data. The problem is that the appearance variance of a person may be too big to be represented by only several samples. Therefore one way is to learn faces incrementally as the new samples come in.

It is known that knowledge learning for human and animal is incremental in nature, over time. One instinctive incremental learning method is to store all the samples in memory, which is however impractical for both the memory requirement and the efficiency in retrieval. To solve the problem Maloof and Michalski proposed a method for partial memory incremental learning (PMIL) [5] for user behavior analysis. Sim et al used reduced-resolution images and synthetic faces by using the direct difference L_p norm as the measurement [12]. However, Sim's method is not an incremental learning system although it can be extended to. The third one is reinforcement learning that requires no memory.

The learning scheme used in this paper is similar to the partial memory incremental learning, but different in the approaches. Assume that a person can be represented as a name or a concept, which is further represented by a set of face images. The partial memory incremental learning is to determine which samples establish the outer bounds of a concept and to deduce new concept and the boundary of it [5]. For a face space F, one person P can be represented by a set of complete and consistent samples. Then the representative samples for P are those lying at the boundaries of P. For face recognition system we can further represent P by several clusters, which may be corresponding to the appearance under different lighting conditions and poses.

A flowchart of the incremental learning face recognition is shown in Fig. 1. Initially we collect a training set to learn a classifier (representatives), which will give the system some ability to recognize faces. In this stage the recognition engine is quite sensitive to the appearance changes of samples. Once the initial system is built, we can use the PMIL to improve the face recognition performance. As shown in Fig. 1, we adopt a real time face detection and tracking module [4], together with the classification module (recognition engine), to play the teacher's role[1]. The rationale behind this is that given a long sequence X of a person P, the probability of recognizing at least one of the captured images as the person P becomes very high. So we have

$$X \leftrightarrow P, \text{ if } \exists X_i \in X,\ X_i \leftrightarrow P \tag{1}$$

where symbol '$\leftrightarrow$' means a sample is matched to a person who is represented by some representative face samples. To be more specific, let's say two faces X and Y are matched under a similarity measurement if $S(X,Y)>T$, where T is a threshold. In the incremental learning process we select a large T to ensure a near-zero false acceptance rate, which means that a match will usually give a positive sample for that person. This will eliminate the error of the teacher in the incremental learning process and ensure that only the faces with the correct identity are used for learning.

The details of the learning process will be described in the following sections. With the incremental learning the representative samples are updated, thereby improving the face classification performance. The process can be repeated indefinitely to learn the new changes being introduced.

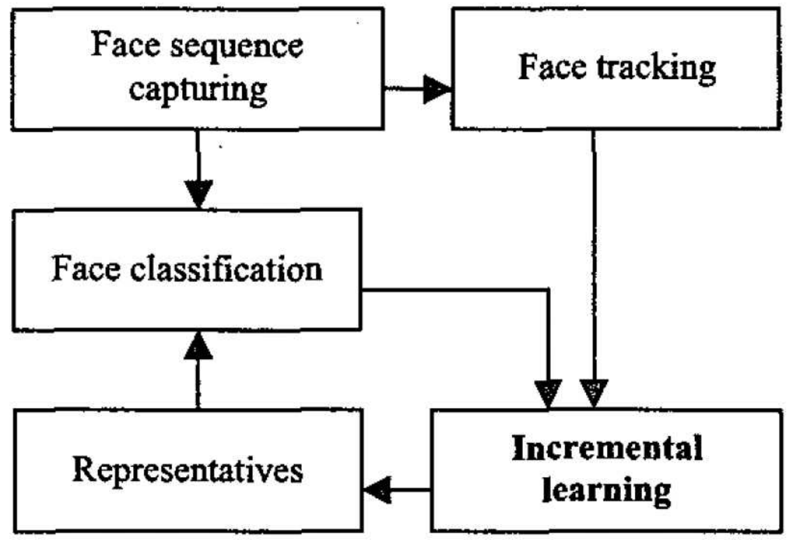

Figure 1 Incremental learning for face recognition

2.2 Searching the representative samples

Two issues should be addressed for face recognition by incremental learning. The first is the selection of the representative samples. The second issue is data updating which will be discussed later. Using the nearest neighbor algorithm, we cluster the face samples and then select the most representative data for each cluster. Different with that in [5], we use the centers of clusters as the representatives for incremental face recognition for the following reason. Assuming that a person's appearance will change greatly under different conditions, a mixture of clusters can therefore be used to represent it. And the cluster center can be used as a compact representative of the respective cluster. It also requires less memory.

Given a cluster of face samples for a person $C\{X_i, i=1,2,\ldots,n\}$, under a distance measurement $D(X,Y)$ for two faces X and Y, we can find the center X_c of the cluster using

$$X_c = arg\ min_k\ \Sigma D(X_j, X_k),\quad j \neq k \text{ and } X_j, X_k \in C \tag{2}$$

The center C will approximate the mean of the cluster if n is big enough. Fig. 2 shows a cluster with 6 samples, where the center is the sample marked with a dark '+'.

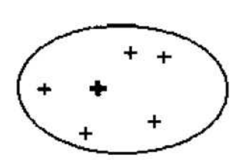

Figure 2 The representative of a cluster

[1] By the tracking technology, we know that the faces in the sequence are from the same person. Then in the whole sequence, any face matched with a registered person indicates that all faces in the sequence belong to that person.

To characterize a person's face, initially we use 5 samples as representative for 5 clusters. With incremental learning, we can update and increase the clusters if necessary. A nearest neighbor algorithm is used for the final face recognition.

2.3 Data updating by PMIL

Incremental learning means that the system can update itself with new incoming samples. To maintain a robust representative for a face and keep the system small and fast, we keep a limited number of clusters for each face and in each cluster we use n samples. The data updating includes the generation of new clusters and updating cluster centers and cluster members. All of these are based on the teacher's guidance. A new cluster will be generated when the faces labeled by the teacher can not be matched with the representatives.

2.3.1 Guidance from the teacher

Given a sequence of faces of a tracked person, $\boldsymbol{X} = \{X_i, i= 1,2,...M\}$ which are labeled by the teacher (tracker plus the classifier) as the person P. Assume that currently N clusters and one representative each cluster exist for P. The set $\boldsymbol{X}$ can be partitioned into two sets, $\boldsymbol{X_r}$ and $\boldsymbol{X_f}$. The set $\boldsymbol{X_r}$ is the collection of all matched faces while $\boldsymbol{X_f}$ is the set of non-matched faces, i.e. not the person P, by the classifier according to the teacher.

2.3.2 Cluster update

For $\boldsymbol{X_r}$ we randomly select some to update the current clusters. Assume a new sample Y in $\boldsymbol{X_r}$ is nearest to a cluster $C(X_c, X_1,X_2,...,X_{n-1})$, we have a new set $C'\{Y,X_c, X_1,X_2,...,X_{n-1}\}$. Using (2) we can find the new representative $X_{c'}$ for the cluster. Further, to keep each cluster having n samples, we delete a sample X_d from C' by using

$$X_d = \arg\min_{j \neq c} D(X_j, X_{c'}) \tag{3}$$

2.3.3 New cluster creation

It is easy to generate new clusters $C_{(N+1)}$, $C_{(N+2)}$, ...from $\boldsymbol{X_f}$ using k-means algorithm. To keep all samples in each cluster with high similarity, the previous threshold T for matching is used to control the cluster. The equation (2) and (3) are used thereafter to get the representative and delete extra samples from it. Currently we set a limit for the total number of the clusters and no new clusters will be generated beyond the preset limit. Also we are working on a deletion strategy that will eliminate the old clusters.

2.3.4 Cluster Prioritization and deletion

Over time a person's appearance may change greatly. To speed up the matching and reduce the memory for storage of the representative faces, the system will prioritize the clusters and delete the 'useless' faces based on the observation that the matching for new face is mainly with the recent images. A probability is assigned to each cluster,

$$P(C_i) = N_m/N_t \tag{4}$$

N_m is the number of the face images matched to the cluster, while N_t is the number of total faces in certain period, e.g. two months. The clusters will be sorted according to $P(C_i)$.While those cluster with the probability less than a threshold will be deleted.

3 Similarity measurement by gabor feature

In this paper we use the Gabor wavelet feature vector on some fixed feature locations of a face for similarity calculation due to their low sensitivity to local distortions, translations and rotations [8,15]. However other similarity function can also be used. We will compare the PMIL recognition result with the one without PMIL.

The Gabor kernels will be applied on the fixed facial feature points to extract the multi-scale and multi-orientation features. Those feature points are defined regarding to the eyes detected on each face using our face and eyes detection module. Then we form a feature vector on each point using all the Gabor responses f_{ij}. There are 18 Gabor kernels used to represent features at 3 scales and 6 orientations:

$$f_i = (f_{i1}, f_{i2}, ..., f_{i18})^T \tag{5}$$

The similarity between two feature points f_i and f_i' are the normalized inner product of the feature vectors:

$$S_{f_i f_i'} = f_i f_i' / |f_i||f_i'| \tag{6}$$

And the final similarity between two faces F and F' is calculated by

$$S(F,F') = S_{FF'} = \Sigma_{i=1}^{N} S_{f_i f_i'} / N \tag{7}$$

where $N=19$ is the number of feature points around eyes, nose mouth and cheek contours [9]. A large similarity indicates that the two face images are similar. In the PMIL, the distance used in (2) and (3) is defined as $D(F,F')=1-S(F,F')$, S is calculated by (7).

For the comparison, the result using a standard face database XM2VTSDB from University of Surrey is used.

It achieved a low equal error rate (*EER*) of *4.2%* for frontal face while *EER=8.42%* if the near frontal faces (within 5 degree of rotation) included in the testing. For non-incremental recognition, the testing of the Gabor feature method against some others achieved the highest recognition rate [8,9]. However performance was not satisfying when we tested it against real data collected from a running system. In the next section we will see that it can only achieve 30% equal error rate, without incremental learning.

4 Experiments

Several experiments are conducted to study the proposed approach. First, we test the recognition accuracy using multiple samples as the representatives for each person. The similarity of Gabor feature vectors is used to find the best match S_{best}, which is returned as the identification result if $S_{best}>T_r$, where T_r is a pre-defined threshold. Next, we test the face recognition using the PMIL. It can also be combined with other traditional and statistical method to obtain better results.

At the beginning, we collected 5 samples per person to build the face recognition engine. About 10 persons were registered in the database. To further compare the partial memory incremental learning results we illustrate the recognition rate for 5 persons. As shown in Fig 3, the initial samples are collected manually in order to cover some appearance variance. However the number of samples one can obtain during the registration phase is limited. Therefore the PMIL becomes significant since it can learn the variance of objects by itself, thereby overcoming the bottleneck of acquiring many images for each person.

The testing data were collected within 10 months, where the environment changed greatly and there are quite a lot of faces are not frontal. The total testing database contains 100 images per user for the first 5 users who is under the PMIL learning, i.e. we will let the system to learn the changing appearances for those 5 users. The testing database also includes 50 images per user for the second 5 users and 750 images from around 70 impostors, who are not registered into the system.

Fig. 3 The initial 5 samples for a person used for registration

Without the incremental learning, the overall system tested using *750* users' trials and *750* impostors' trials gives the false acceptance rate (FAR) and false rejection rate (FRR) as in Fig 4, which shows an equal error rate *EER=35%* where there are large changes of the environment and the person's appearance. With PMIL for 5 of the users the EER reduced to 11%, see Fig 5.

We also tested the method on another test sets containing 750 images from impostors and the 500 images from only PMIL users. The results are showed in Figure 6 and 7. The EER reduced from to 8.48%. Fig. 8 shows some representatives for a person after the incremental learning. Note that even with many impostors, the recognition rate can be improved greatly (35% reduced to 11%).

Taking the 500 trials from the 5 PMIL users and 250 trials from the other users, the system can achieve an EER=5.4% with PMIL from 33.5% without PMIL. See the result in Fig 9 and 10, where the EER is reduced from 30% to 1.9%.

Notice that the system performance is almost the same if without the PMIL training. The EER is around 30% no matter how many users registered into the system and whether there are any impostors. Experiments show that for PMIL user only, the system can reduce the EER to a very small figure.

For an interaction system such as a robot, we hope it can recognize a person correctly, especially a person with whom it interacts frequently. A large threshold is therefore used to reduce the FAR, which is nearly zero when the threshold is set to 95.

5 Discussion and Conclusion

In this paper we presented a novel incremental learning approach for face recognition. Although it is only applied to face recognition here, we believe that it can also be extended to other applications. For simplicity only the nearest-neighbor classification is used. However other methods such as K-NN or support vector machine can also be used to learn each face class with the new incoming samples.

The scheme presented here boot-strapped the new samples using a real time face detection and tracking system. The samples can be easily learned and updated into the classifier. The Gabor features extracted from feature location are used to measure the similarity between faces. An efficient learning and updating method is proposed that can characterize the new knowledge the robot received.

From the experiments, we find that, with the multiple samples, the recognition rate can be increased. But the enhancement is not satisfying since the samples selected manually are usually not representative enough. While with the incremental learning, the recognition rate is increased greatly, compared to that of non-learning scheme.

The advantage of the method proposed here is its ability to capture and learn the new samples. However, the classification using nearest neighbor is not an optimal method considering that the samples are limited and some of the past samples are actually discarded. The future work is to conduct study to find a way to learn the samples more efficiently while maintaining the real time processing capability.

Examined the result we found that many of the errors in the above testing are caused by the rotated faces. The test by keeping only the near frontal faces shows that the EER is reduced further to 1.25% for the users after PMIL learning. Thus how to learn the multi-view face appearances are critical to improve the performance.

From the experiments we can conclude that the recognition accuracy can be improved greatly by PMIL approach, while the more learning the better performance can be achieved. A system with all users' faces learned by PMIL shows a better result than the one with part of the users' faces learned. On the other hand, although an impostor's face can be matched with a user's face with a higher similarity given that there are more user's representatives in the database after PMIL learning, the overall recognition accuracy can still be improved by the proposed approach.

References

[1] R. Brunelli and T. Poggio, ``Face Recognition: Features versus Templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10), pp. 1042-1052, 1993

[2] R. Chellapa and S. Sirohey, "Human and Machine Recognition of Faces: A Suevey," Proc of IEEE, Vol 83, No. 5, pp. 705-740, May, 1995

[3] S. Gong, A Psarrou, I. Katsoulis and P. Palavouzis, "Tarcking and Recognition of Face Sequences," Proc of European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, Germany, pages 96-112 , Nov, 1994

[4] W. Huang et al, "Real Time Face detection and Eyes location," IEEE TENCON'02, pp.507-510, Oct, 2002

[5] M. A. Maloof and R. S. Michalski, "A method for Partial-Memory Incremental Learning and Its Application to Computer Intrusion Detection," Proc of IEEE Int Conf on Tools with AI, Washington DC, pp. 392-397, Nov, 1995

[6] A. Pentland, B. Moghaddam, T. Starner, View-based and modular eigenspaces for face recognition, Proceedings of IEEE. Conference On Computer Vision and Pattern Recognition, Seattle, pp. 84-91, June 21-23, 1994

[7] A. P. Pentland and M. A. Turk, "Face Recognition Using Eigenfaces," Proc of ICPR, pp. 586-591, 1994

[8] M. Rajapakse and Y. Guo, "Multiple Landmark Feature Point Mapping for Robust Face Recognition," Proc. of the Third Int Conf of Audio- and Video-based Biometric Person Authentication, pp.96-101, June, 2001

[9] M. Rajapakse and Y. Guo, "Efficient Gabor Feature Based Approach for Face Recognition," RWCP Tech Report-07, Kent Ridge Digital Labs, Sept, 2001

[10] B. Raytchev and H. Murase, "Unsupervised face recognition from image sequences," Proc of ICIP 2001, Greece, Vol 1, pp. 1042-1045, Oct, 2001

[11] G. Shaknarovich, John W. Fisher and T. Darrell, "Face Recognition from Long-term Observatipns," ECCV 2002

[12] T. Sim, R. Sukthankar, M. Mullin, S. Baluja, "Memory-based face recognition for visitor identification," Proceedings of IEEE Face and Gesture, France, pp. 214-220, March, 2000

[13] T. Sim and T. Kanade, " Combining Models and Exemplars for Face Recognition: An Illuminating Example," Proceedings of the CVPR 2001 Workshop on Models versus Exemplars in Computer Vision, December, 2001

[14] T. Vetter, "Synthesis of novel views from a single face image," International Journal of Computer Vision 28, No. 2, 102-116, 1998

[15] L. Wiskott, J. M. Fellous, N. Krer, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), pp.775-779, 1997

[16] L. Zheng and J.-K. Wu, "A New Model-based Lighting Normalisation Algorithm And Its Application in Face Recognition," RWCP Tech Rept-05, Kent Ridge Digital Labs, Sept, 2001

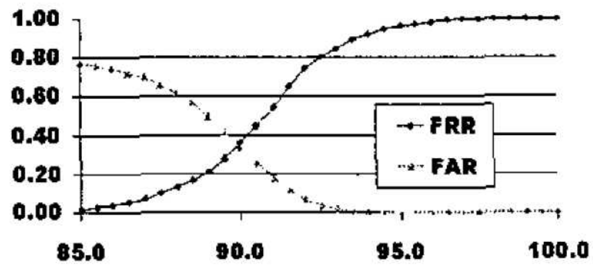

Figure 4 Face recognition result without PMIL, 750 tests for registered users and 750 tests from 70 persons as impostors. EER=35%.

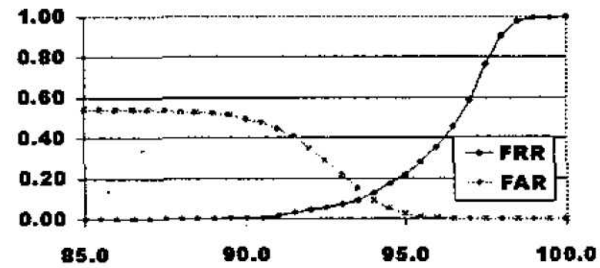

Figure 5 Face recognition result with PMIL for 5 users, same testing data are used as those in Figure 5. EER=11%.

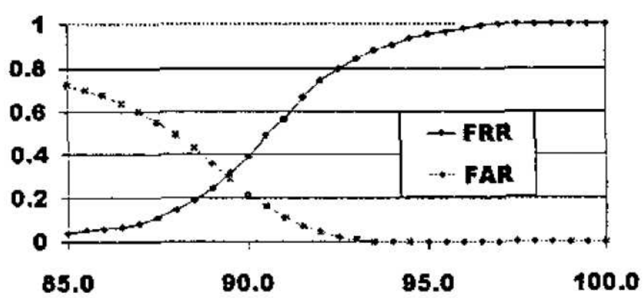

Figure 6 Face recognition result without PMIL, 500 tests from registered users and 750 tests from impostors. EER=31%

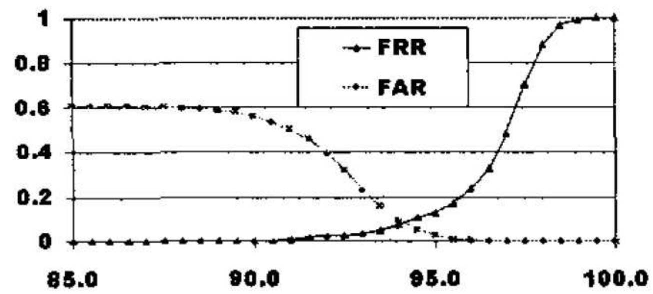

Figure 7 Face recognition result with PMIL, same testing sets are used as the one in Figure 7. EER=8.48%

Figure 8 Representatives after incremental learning

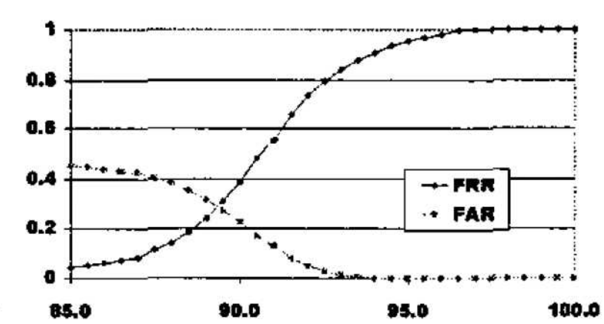

Figure 9 Face recognition result without PMIL, 500 tests from registered users only. EER=29.1%

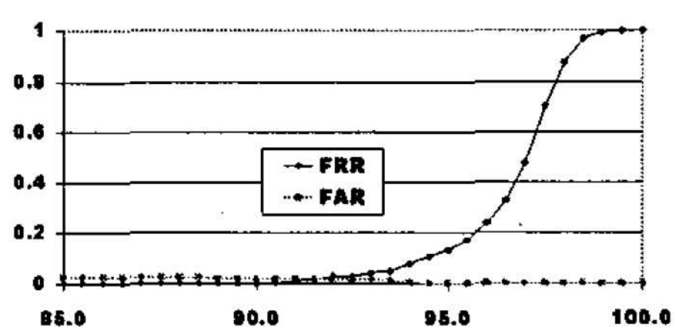

Figure 10 Face recognition result with PMIL, same testing set as the one used for Figure 10. EER=1.9%