Person Location Service on the Planetary Sensor Network

Ting Shan
Intelligent Real-Time Imaging and Sensing Group, EMI, School of ITEE, University of Queensland
Brisbane, QLD, Australia
shanting@itee.uq.edu.au

Brian C. Lovell
Intelligent Real-Time Imaging and Sensing Group, EMI, School of ITEE, University of Queensland
Brisbane, QLD, Australia
lovell@itee.uq.edu.au

Shaokang Chen
Intelligent Real-Time Imaging and Sensing Group, EMI, School of ITEE, University of Queensland
Brisbane, QLD, Australia
shaokang@itee.uq.edu.au

Abstract

This paper gives a prototype application which can provide person location service on the IrisNet. Two crucial technologies – face detection and face recognition underpinning such image and video data mining service are explained. For the face detection, authors use 4 types of simple rectangles as features, Adaboost as the learning algorithm to select the important features for classification, and finally generate a cascade of classifiers which is extremely fast on the face detection task. As for the face recognition, the authors develop Adaptive Principle Components Analysis (APCA) to improve the robustness of Principle Components Analysis (PCA) to nuisance factors such as lighting and expression. APCA also can recognize faces from single face which is suitable in a data mining situation

Keywords

Face Detection, Face Recognition, Adaboost, PCA, APCA.

1 INTRODUCTION

Multimedia data, such as speech, music, images and video are becoming increasingly prevalent on the internet and intranets as bandwidth rapidly increases due to continuing advances in computing hardware and consumer demand. An emerging major problem is the lack of accurate and efficient tools to query these multimedia data directly, so we are usually forced to rely on available metadata such as manual labeling. This is already uneconomic or, in an increasing number of application areas, quite impossible because these data are being collected much faster than any group of humans could meaningfully label it. Some driver applications are emerging from heightened security demands in the 21st century, postproduction of digital interactive television, and the recent deployment of a planetary sensor network overlaid on the internet backbone.

2 FAST FACE DETECTION

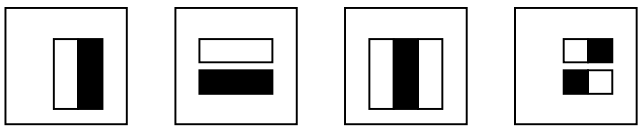
2.1 Face Detection

Face detection is a challenging and valuable work and has attracted much attention in recent years. Face detection is a necessary first-step in face recognition system, with the purpose locating the face from the cluttered background. It also can be used in wide areas such as human-computer interaction, content-based image retrieval, and intelligent surveillance. The survey paper [2] by E. Hjelmas and B. K. Low classify the previous work on face detection into two categories: feature-based approaches and image-based approaches.

Feature-based approaches such as using edges [3, 4], skin color [5], motion [6] etc, are applicable for real-time systems due to their fast feature extraction but suffer from their low detection rate. Image-based such as PCA [7], Neural Networks [8], support vector machine [9] generally achieve a good performance, but most of them are computationally expensive and not suitable for real-time applications. In recent years, Viola and Jones[10] proposed a real-time face detection system. The main idea of the method is to combine weak classifiers based on simple features which can be computed extremely fast. In their work, simple rectangle Haar-like features are extracted; face and non-face classification is done by using a cascade of successively more complex classifiers which are trained by AdaBoost learning algorithm. Our face detection system is based on their work.

2.2 Feature

Each weak classifier is constructed based on a simple rectangle feature. Four types of rectangle features are used, as shown in Fig. 1



(A) (B) (C) (D)

Fig. 1. The four types of rectangle features defined in a sub-window: the sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles.

Given the base resolution of the sub-window is 24*24, the exhaustive set of rectangle features is 116,300 (86,400 for 2 rectangle features, 27,600 for 3 rectangle features, and 2,300 for 4 rectangle features), which is overcomplete.

Rectangle features can be computed very fast using integral image. The integral image at location x, y contains the sum of pixels above and to the left of x, y, inclusive:

$$II(x,y) = \sum_{x \le x', y \le y'} I(x', y')$$

where II(x, y) is the integral image and I(x', y') is the original image.

Using the integral image any rectangular sum can be computed in four array references (Fig. 2). More clearly, two-rectangle features can be computed in six references, eight for the three-rectangle features and nine for four-rectangle features.

2.3 Learning Algorithm – Adaboost

Adaboost algorithm was mainly developed by Freund and Schapire [11]. They proved that the training error of the strong classifier approaches zero exponentially in the number of rounds.

The weak classifier is designed to select the single rectangle feature which can best separate the positive and negative examples. A weak classifier h_j contains a feature f_i a threshold θ_i and a direction ρ_i

$$h_j = \begin{cases} 1 & if \quad \rho_i f_i(x) < \rho_i \theta_i \\ 0 & otherwise \end{cases}$$

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \ldots, T$:
  1. Normalize the weights,
     $$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$
     so that w_t is a probability distribution.
  2. For each feature, j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t, $\epsilon_j = \sum_i w_i \left| h_j(x_i) - y_i \right|$.
  3. Choose the classifier, h_t, with the lowest error ϵ_t.
  4. Update the weights:
     $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$
     where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is:
  $$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \ge \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

Fig. 2 AdaBoost algorithm for classifier learning.

In our system, each classifier is trained with the 4916 training faces samples and 7872 non-face samples (both of them have the size 24*24 pixels) using the Adaboost learning algorithm.

2.4 Cascade Classifier

The goal of a cascade of classifiers is to enhance the classification rate which reduces the computing time. A positive result from the first classifier will trigger the second classifier which is more complex than the first one, a positive result from the second classifier will trigger a third classifier, and so on. A negative result at any stage will lead to the immediate rejection to the sub-window. In this way, the detection process is extremely fast.

3 NEED FOR FACE RECOGNITION FROM SINGLE FACE

3.1 Robust Face Recognition

Robust face recognition is a challenging goal because of the gross similarity of all human faces compared to large differences between face images of the same person due to variations in lighting conditions, view point, pose, age, health, and facial expression. Most systems work well only with images taken under constrained or laboratory conditions where lighting, pose, and camera parameters are strictly controlled.

Recent research has been focused on diminishing the impact of nuisance factors on face recognition. Many approaches have been proposed for illumination invariant recognition [12][13] and expression invariant recognition [14][15]. But these methods suffer from the need to have large numbers of example images for training, which is often impossible in many data mining situations when only few sample images are available such as in recognizing people from surveillance videos from a planetary sensor web or searching historic film archives.

Table 1. Data mining applications for face recognition

Person recognition and location services on a planetary wide sensor net
Recognizing faces in a crowd from video surveillance
Searching for video or images of selected persons in multimedia databases
Forensic examination of multiple video streams to detect movements of certain persons
Automatic annotation and labeling of video streams to provide added value for digital interactive television

3.2 Principle Component Analysis

Principal Components Analysis (PCA), also known as "eigenfaces," is originally popularized by Turk and Pentland [16]. PCA is a second-order method for finding a linear representation of faces using only the covariance of the data. It determines the set of orthogonal components (feature vectors) which minimizes the reconstruction error for a given number of feature vectors. Consider the face image set $I = [I_1, I_2, \cdots, I_n]$, where I_i is a $p \times q$ pixel image, $i \in [1 \cdots n]$, $p, q, n \in Z^+$, the average face of the image set is defined by the matrix:

$$\Psi = \frac{1}{n}\sum_{k=1}^{n} I_i \ . \quad (1)$$

Normalizing each image by subtracting the average face, we have the normalized difference image matrix:

$$\tilde{D}_i = I_i - \Psi \ . \quad (2)$$

Unpacking $\tilde{D}_i$ row-wise, we form the N $(N = p \times q)$ dimensional column vector d_i. We define the covariance matrix C of the normalized image set $D = [d_1, d_2, \cdots d_n]$ corresponding to the original face image set *I* by:

$$C = \sum_{i=1}^{n} d_i d_i^T = DD^T \ . \quad (3)$$

An eigen decomposition of C yields eigenvalues λ_i and eigenvectors u_i which satisfy:

$$Cu_i = \lambda_i u_i , \quad (4)$$

$$C = DD^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T , \quad (5)$$

where $i \in [1 \cdots N]$.

The eigenvectors of C are often called the eigenfaces and are shown as images in Figure 3. Generally, we select a small subset of $m < n$ eigenfaces to define a reduced dimensionality facespace that yields highest recognition performance on unseen examples of faces. For good recognition performance the required number of eigenfaces, m, is typically chosen to be of the order of 6 to 10.

Fig.3 Typical set of eigenfaces as used for face recognition. Leftmost image is average face.

3.3 Robust PCA Recognition

The authors have developed Adaptive Principal Component Analysis (APCA) to improve the robustness of PCA to nuisance factors such as lighting and expression [17][18]. In the APCA method, we first apply PCA. Then we rotate and warp the facespace by whitening and filtering the eigenfaces according to overall covariance, between-class, and within-class covariance to find an improved set of eigenfeatures. Figure 4 shows the large improvement in robustness to lighting angle. The proposed APCA method allows us to recognize faces with high confidence even if they are half in shadow. Figure 5 shows significant recognition performance gains over standard PCA when both changes in lighting and expression are present.

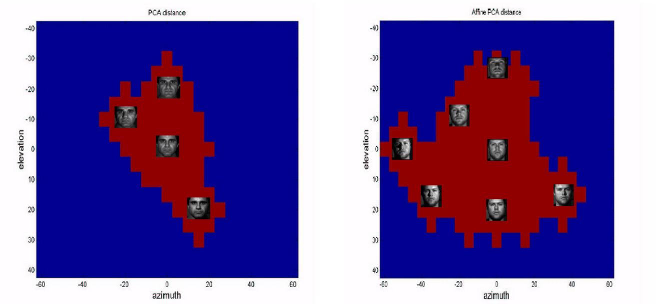


Fig.4 Contours of 95% recognition performance for the original PCA and the proposed APCA method against lighting elevation and azimuth.

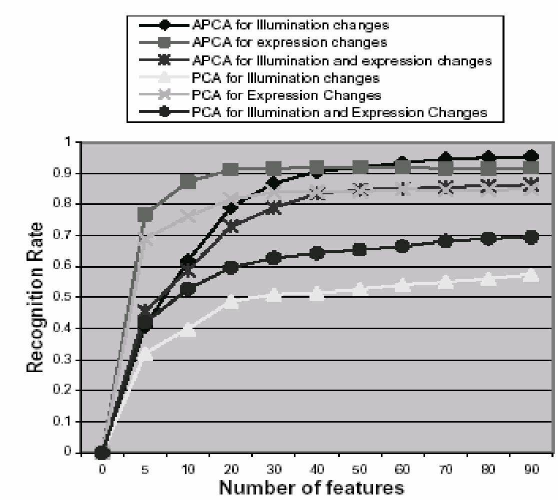
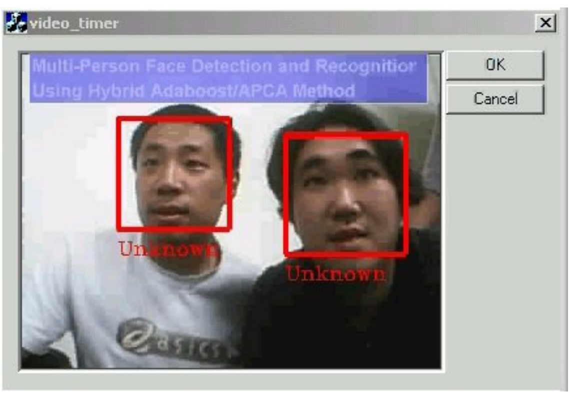


Fig.5 Recognition rates for APCA and PCA versus number of eigenfaces with variations in lighting and expression from Chen and Lovell (2003).

4 EXPERIMENTAL RESULTS

We present some experimental results here. There are 15 people (each person has one orientated face image) in our face database. The demo video shows the progress of detecting and recognizing of multiple persons from “unknown” to “confident”. Some selected frames are shown on Fig.6



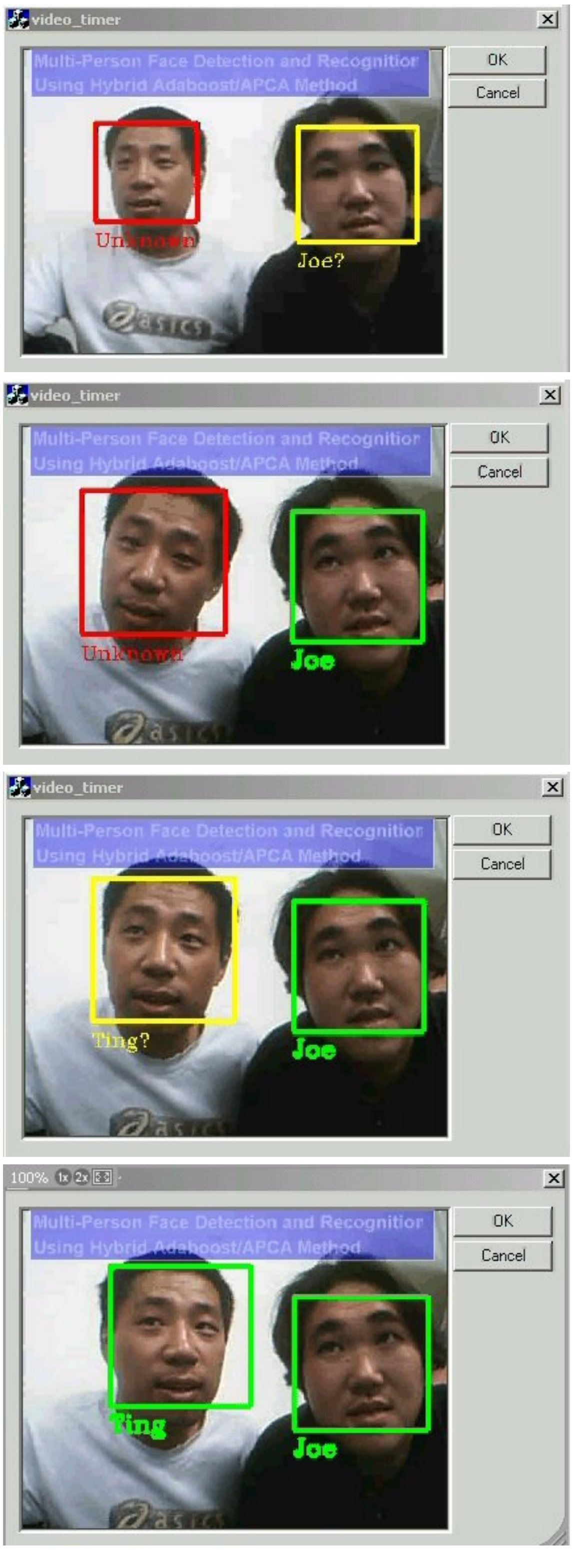


Fig.6 Selected frames from application demo video

Red rectangle: Unknown person

Yellow rectangle: Not confident enough to recognize, person’s name is under the rectangle with a “?”

Green rectangle: Very confident to recognize, person's name is under the rectangle

5 CONCLUSION AND FUTURE WORK

It has been argued that by the end of the 20th century computers were very capable of handling text and numbers and that in the 21st century computers will have to able to cope with raw data such as images and speech with much the same facility. The explosion of multimedia data on the internet and the conversion of all information to digital formats (music, speech, television) is driving the demand for advanced multimedia search capabilities, but the pattern recognition technology is mostly unreliable and slow. Yet, the emergence of handheld computers with built-in speech and handwriting recognition ability, however primitive, is a sign of the changing times. The challenge for researchers is to produce pattern recognition algorithms, such as face detection and recognition, reliable and fast enough for deployment on data spaces of a planetary scale.

In our application, currently face detection module can detect faces with rotated angles very well, but APCA can't recognize well on the rotated faces. Our future work will be focused on dealing with this problem. Some potential solutions include detect the positions of eyes or nose, and rotate the face back to orientation position depends on the face component geometry.

REFERENCES

[1] Gibbons, P.B., Karp, B., Ke, Y., Nath, S., and Sehan S, "IrisNet: An Architecture for a Worldwide Sensor Web," Pervasive Computing, 2(4), 22-23, Oct – Dec, 2003

[2] Erik Hjelmas and Boon Kee Low "Face Detection: A Survey" April 17, 2001

[3] V.Govindaraju "Locating human faces in photographs" Int. J. Comput. Vision 19,1996

[4] J.Huang, S.Gutta, and H.Wechsler "Detection of human faces using decision trees", in IEEE Proc. of 2nd Int.Conf. on Automatic Face and Gesture Recognition, Vermont, 1996

[5] C.H.Lee, J.S.Kim, and K.H.Park, "Automatic human face location in a complex background" Pattern Recog. 29,1996,1877-1889

[6] S.McKenna, S.Gong, and H.Liddell, "Real-time tracking for an integrated face recognition system", in 2nd Workshop on Parallel Modelling of Neural Operators, Faro, Portugal, Nov, 1995

[7] K.-K.Sung and T.Poggio, "Example-based learning for view-based human face detection", IEEE Trans. Pattern Anal.Mach.Intelligence 20, 1998, 39 -51.

[8] H.A.Rowley, S.Baluja, and T.Kanade, "Neural network-based face detection", IEEE Trans. Pattern Anal. Mach. Intell. 20, January 1998,23-38.

[9] *Osuna, E,Freund, R.; Girosit, F.;* "Training support vector machines: an application to face detection" 1997 IEEE Computer Society Conference on, 1997

[10] Paul.Viola, Michael.Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", In Pro on CVPR, pp. 511-518, 2001

[11] R.E.Schapire, "A Brief Introduction to Boosting", In Proc. of 16th Int.Joint Conf. on A.I., 1999

[12] Yilmaz, A. and Gokmen, M., "Eigenhill vs. eigenface and eigenedge", In Procs of International Conference Pattern Recognition, Barcelona, Spain, 827-830, 2000

[13] Gao, Yongsheng and Leung, Maylor K.H., "Face Recognition Using Line Edge Map", IEEE PAMI. 24(6), June, 764-779,2002

[14] Beymer, D., and Poggio, T. "Face Recognition from One Example View", Proc. Int'l Conf. of Comp. Vision, 500-507.1995

[15] Black, M. J., Fleet, D. J. and Yacoob, Y., "Robustly estimating Changes in Image Appearance", Computer Vision and Image Understanding, 78(1), 8-31.2000

[16] Turk M. A., and Pentland, A. P., "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 3(1), 71-86.1991

[17] Chen, Shaokang and Lovell, Brian C., "Illumination and Expression Invariant Face Recognition with One Sample Image," Proceedings of the International Conference on Pattern Recognition, Cambridge, August 2004, 23-26.

[18] Chen, Shaokang and Lovell, Brian C., "Face Recognition with One Sample Image per Class," Proceedings of ANZIIS2003, Sydney, December 10-12, 2003, 83-88.