Real-time Face Detection from One Camera

1. Research Team

Project Leader: Prof. Ulrich Neumann, *IMSC and Computer Science*

Other Faculty: Prof. Isaac Cohen, *Computer Science*
Prof. Suya You, *Computer Science*

Graduate Students: Douglas Fidaleo, Kyoungchin Seo

2. Statement of Project Goals

Nowadays, the importance of the automatic face detection and tracking system has increased as it is needed for video surveillance and new user interfaces [1]. The goal of this research effort is to construct an automatic face detection system using a standard PC camera in real-time.

The proposed system is composed of several modules: background subtraction using Gaussian modeling and face detection using machine learning techniques [2].

3. Project Role in Support of IMSC Strategic Plan

IMSC has developed several techniques on facial modeling and rendering, for example, facial expression cloning [3] and facial expression analysis [4]. The proposed real-time face detection system easily interacts with these developed methods. The interaction with head pose estimation means that we automatically obtain orientation of head pose from detected regions of the face. The detected face regions are also utilized as input to a face expression analysis system to realize online face expression recognition. A good example for describing the scenario of the system is the one of an ATM machine that can interact with users by recognizing the users' facial expression in real-time.

4. Discussion of Methodology Used

In this report, we describe a method of face detection using the statistical learning method. Whenever people are within the field of view of the camera, the system recognizes the existence of the face and obtains the location among segmented regions. In order to recognize and localize the face simultaneously, we adopt a face detection framework to decide whether the region of image is a face or a non-face. For improving the performances of the system, background regions (i.e. stationary part of the scene) are discarded by focusing on the segmented moving objects. We also use a pyramid approach allowing us to automatically detect faces of various sizes in the image.

4a. System Overview

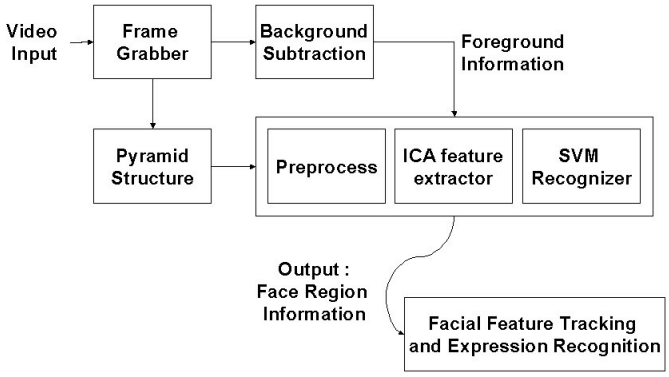

Figure 1: Proposed Face Detection System

Figure 1 shows the diagram of our proposed face detection system. Video stream from a camera is directly used as input to the face detection system. The video stream is divided into each frame by a software grabber. Foreground regions are segmented for each frame by background subtraction method. Within these segmented regions, we apply the face detection algorithm for detecting face regions. The detection module is composed of preprocessing, ICA feature extraction [5], and SVM classification [6].

Detecting faces of various sizes requires scanning for regions of various sizes in each frame. Instead, we use a 3-level pyramid structure for detecting faces of various shapes using the same scanning window. If the recognizer succeeds in finding a face, the proposed system gives output with a format of boundary box for the detected face region. In the future, the output result will be used for the facial feature tracking and facial expression analysis system.

4b. Background Subtraction

For robust segmentation of foreground and background, the proper background model is necessary. For a color camera, we use the difference model. Before the detection, we store the one video frame in the condition of no foreground objects. We can simply calculate the difference between the stored one and the current input image for each red, blue, and green color channel. We calculate a weighted sum of each band difference values and decide face or non-face by checking this summation value. After obtaining mask images including segmented moving regions, we make a segmented image using this mask image as depicted in Figure 2.

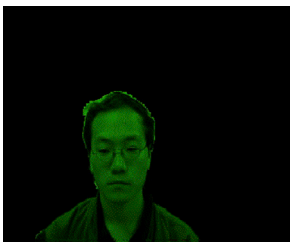

Figure 2: Segmented Result

4c. Hybrid Statistical Learning with ICA and SVM

For detecting faces in an image, Support Vector Machine (SVM) is suitable and it shows robust performance among statistical learning methods. In order to improve classification results, it is more efficient to use statistical features as an input to SVM instead of raw pixel data. ICA (independent component analysis) can be used as a good feature extractor suitable for SVM [7][8]. We use the ICA representation of database images as an input feature to SVM. Before we apply ICA to raw images, we use PCA (Principal Component Analysis) in order to reduce the dimension of input for efficient computation. Reduced number of Eigenvectors is used for calculating ICA. As shown in Figure 3, we extract ICA feature basis images from the eigenvectors. All the database images are projected into these ICA basis images. So we get a feature vector per image in the database and construct face recognition module by training SVM using these feature vector as an input.

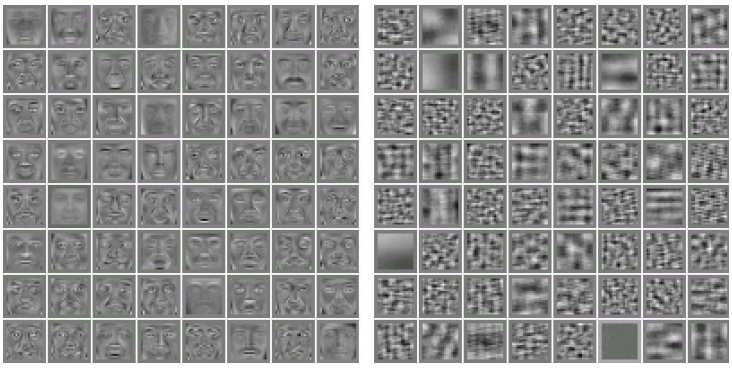

Figure 3: 64 Basis Images for face images (left) and non-face images (right) as a result of ICA

4d. Pyramid-based Search

The developed SVM classifier is analyzing each image region and classifies it as a face or non-face region. The confidence of the classification is given by the distance to the support vector separating the two classes. This classifier is trained on fixed face sizes that do not always correspond to the one in the image. Indeed, the size of a face in an image can vary according to the distance of the person to the camera and faces of various sizes can also be present. Therefore, in order to use the same training data for detecting faces of various sizes we have implemented a pyramid approach that scans for faces at various image resolutions. We use currently a 3-level pyramid. The system scans for faces in the foreground regions of the upper pyramid's level as shown in Figure 4. If the face region is detected in upper level image, the search for the selected region is ignored in the other level. Using this strategy we can reduce the searching time and achieve real-time performances.

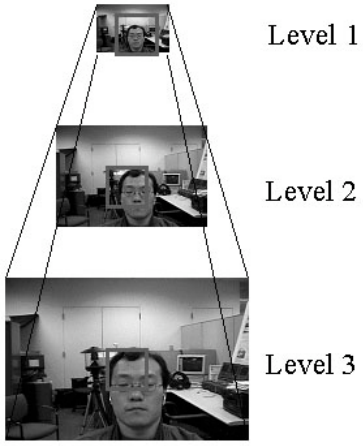

Figure 4: 3-level pyramid structure

5. Short Description of Achievements in Previous Years

We designed a face detection system, which works using one camera and implemented a prototype application as shown in Figure 5. From captured images, the system can find faces in the image.

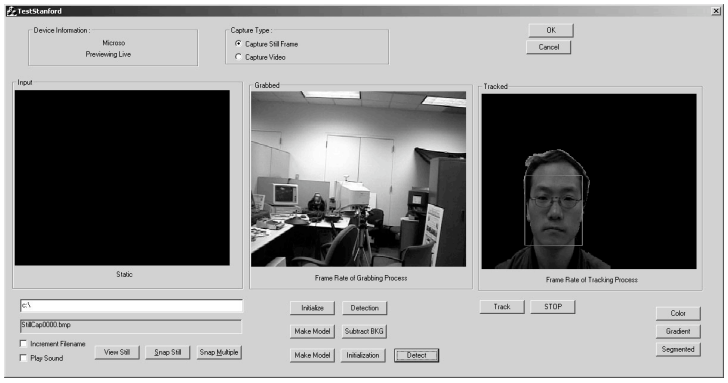

Figure 5: Prototype System

5a. Detail of Accomplishments During the Past Year

The implementation of our prototype for face detection system is composed of several modules: interface between PC and PC-camera, background subtraction process, pyramid structure-based search, face detection. A 3-level pyramid structure per each frame is constructed for adaptively detecting faces of different sizes as depicted in Figure 6. The SVM classifier was trained on approximately 2000 faces of the FERET database [9] and 1000 non-face images. The face images were aligned using an affine transformation and performed a histogram equalization. We have applied ICA to the constructed database and trained SVM for the extracted ICA feature.

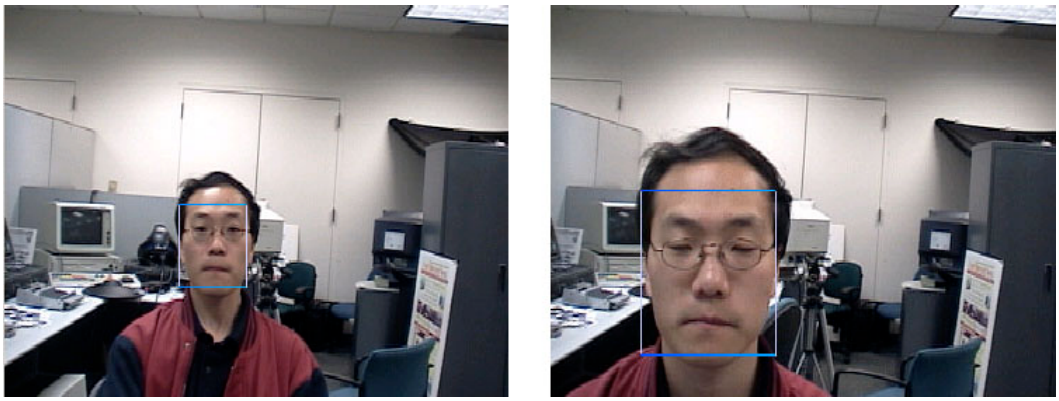

Figure 6: Detection of faces of various sizes using the 3-level pyramid.

6. Other Relevant Work Being Conducted and How this Project is Different

Our proposed face detection system is utilized in various fields. Head orientation estimation techniques are strongly related to our system. Extracted head pose is used to control the mouse point for perceptual user interface. The adjusted results using pose parameters for Localized face region become an input to facial expression system. Using this integration, we obtain more accurate expression result. Exact face information compensated by pose parameter and facial expression analysis will be used to transfer person's expression and motion onto a graphically generated face model based on expression cloning technique.

7. Plan for the Next Year

The proposed face detection system will be integrated into a head pose estimation system. The integrated system will be able to get the pose parameters of the detected face from the video. This pose information will be used to adjust the detected image region for recognizing accurately facial expression. In the end, face expression algorithm will be integrated into the whole system and will allow the recognition of facial expression. This will be extended to the simultaneous detection of multiple faces and their pose estimation for video conferencing applications.

8. Expected Milestones and Deliverables

- Implementation of real-time face detection system for multiple people
- Extension to head pose estimation technique
- Extension to face expression analysis technique

9. Member Company Benefits

The developed methodology is applicable in video surveillance application as well as multimodal interaction systems. We have initiated contacts with NCR on the application of such technology for next generation ATM machines.

10. References

[1] E. Hjelmas and B. K. Low, "Face Detection: A Survey," *Computer Vision and Image Understanding*, vol. 83, pp. 236-274, 2001.
[2] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," *IEEE Proceeding of Computer Vision and Pattern Recognition 97*, vol. 6, pp. 130-136, 1997.
[3] J.Y. Noh and U. Neumann, "Expression Cloning," *Computer Graphics, Proceedings of ACM SIGGRAPH 2001*, Los Angeles CA, August 2001, pages 277-288.
[4] D. Fidaleo and U. Neumann, "CoArt: Co-articulation Region Analysis for Control of 3D Characters," *Proceedings of IEEE Computer Animation*, pp. 12-17, 2002.Animation, pp. 12-17, 2002.
[5] A. Hyvarinen, P. O. Hoyer, and Mika Inki, "Topographic independent component analysis," *Neural Computation*, vol. 13, pp 1527-1558, July 2001.
[6] B. Heisele, A.Verri, and T. Poggio, "Learning and Vision Machines," *Proceedings of IEEE*, vol. 90, no. 7, July 2002.
[7] T.-K. Kim, S.-U. Lee, J.-H. Lee, S.-C. Kee, and S.-R. Kim, "Integrated approach of multiple face detection for video surveillance," *International Conference of Pattern Recognition 2002*,vol. 2, pp. 394-397, August 2002.
[8] Y. Qi, Doermann D. and Dementhon D., "Hybrid independent component analysis and support vector machine learning scheme for face detection," *IEEE International Conference on Acoustics, Speech, and Signal Processing 2001 Proceedings*., vol. 3, pp 1481-1484, May 2002.
[9] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Journal of Image and Vision Computing*, vol. 15, no. 5, pp 295-306, 1998.