CARA: Character Appearance Retrieval and Analysis for TV Programs

Byunghee Jung, Sungchoon Park and Kyeongsoo Kim

E-mail : bhjung@kbs.co.kr
Broadcast Technical Research Team, Korean Broadcasting System

Abstract

This paper describes a character retrieval system for TV programs and a set of novel algorithms for detecting and recognizing faces for the system. Our character retrieval system consists of two main components: Face Register and Face Recognizer. The Face Register detects faces in video frames and then guides users to register the detected faces of interest into the database. The Face Recognizer displays the appearance interval of each character on the timeline interface and the list of scenes with the names of characters that appear on each scene. These two components also provide a function to modify incorrect results, which is helpful to provide accurate character retrieval services. In the proposed face detection and recognition algorithms, we reduce the computation time without sacrificing the recognition accuracy by using the DCT/LDA method for face feature extraction. We also develop the character retrieval system in the form of plug-in. By plugging in our system to a cataloguing system, the metadata about the characters in a video can be automatically generated. Through this system, we can easily realize sophisticated on-demand video services which provide the search of scenes of a specific TV star.

1. Introduction

As the digital multimedia services such as distributing TV programs or movies through the Internet are widely used, searching a shot or scene that contains specific characters becomes useful to both users and program-producing staffs. For providing character retrieval services, automatic face detection and recognition algorithms are necessary, where the algorithms satisfy the requirements of fast execution and high accuracy. Thus, we propose a novel fast face recognition method and develop a character retrieval system using these algorithms. To elevate the applicability of the character retrieval system in various digital archive environments such as networked production system, and contents-providing applications on the Web and mobile devices, we develop the system in the form of both standalone and plug-in. Our system achieves real-time execution and shows high applicability.

The proposed system consists of *Face Register* and *Face Recognizer* as illustrated in Figure 1. Face Register detects faces in video frames and helps users to register faces into a database. Detected faces are then recognized by the proposed component-based DCT/LDA which reduces the dimension of a signal using DCT for LDA. The component-based DCT/LDA utilizes features of the holistic face and facial components such as eyes and nose. The component-based DCT/LDA maintains its high recognition rate under various types of face variation such as pose, facial expression and illumination. Once Face Recognizer identifies appearance intervals of each character using face recognition, the intervals are displayed on the timeline along with the names of characters. It also performs the scene change detection whenever appearing characters alter. Since the current system can return incorrect results, it is necessary to provide error correction by editing the names of recognized characters.

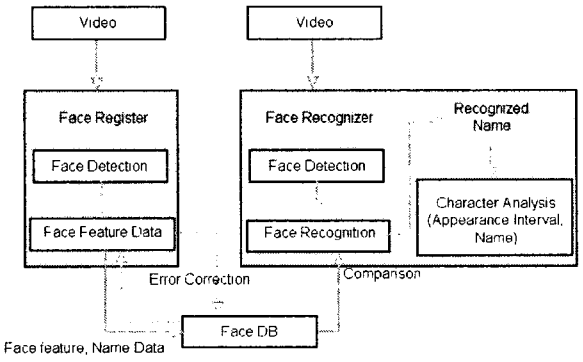

Figure 1. Framework of the System

The rest of this paper is organized as follows. In Section 2, we describe the character retrieval system in detail. Then, we propose an advanced DCT/LDA face recognition method in Section 3. Finally, we present the experimental result of our method and discuss the conclusion and the future work in Section 4 and 5.

2. Character Retrieval System

The proposed character retrieval system consists of two steps: 1) registering faces of interest and 2) recognizing the registered faces. In this section, the functionalities and user interfaces of both standalone and plug-in system are presented in detail.

2.1 Face Register

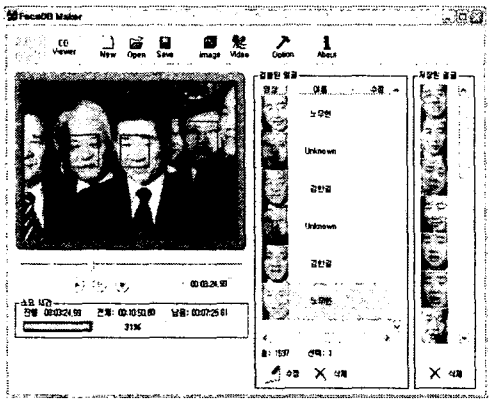

Figure 2. GUI of Face Register

Face Register helps users to register faces of characters into a database. When a character is recognized by the registered face images in the database, the corresponding name of the recognized face image is displayed.

To improve the recognition rate, it is better to register several faces with pose variation. When an unknown face is presented to the system, user has an option of registering the face one at a time whenever it occurs or registering several faces together at a time. For the latter, a module of clustering or grouping unregistered faces is developed to avoid repeated registration of the same character. The user interface of Face Register is shown in Figure 2.

2.2 Face Recognizer

The purpose of Face Recognizer is to analyze and display the appearance intervals of each character on the timeline using face recognition. Also, a scene change is detected when the recognized character disappears from the scene or a new character appears. Detected scenes are verified by the edge histogram because of recognition errors in the face recognition module. A module of face tracking utilizes temporal information between video frames in order to improve face detection speed by reducing search regions and face detection accuracy by updating wrong results. Figure 3 shows the user interface of Face Recognizer.

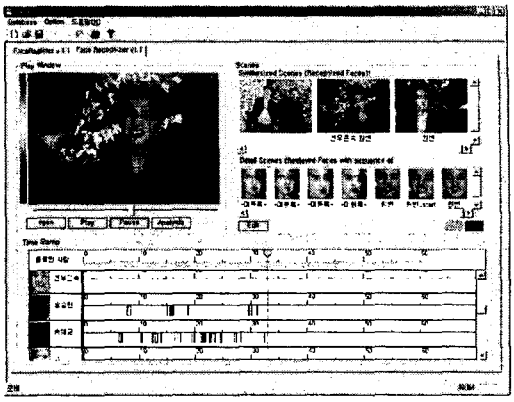

Figure 3. GUI of Face Recognizer

2.3 Plug-in Cataloguing System

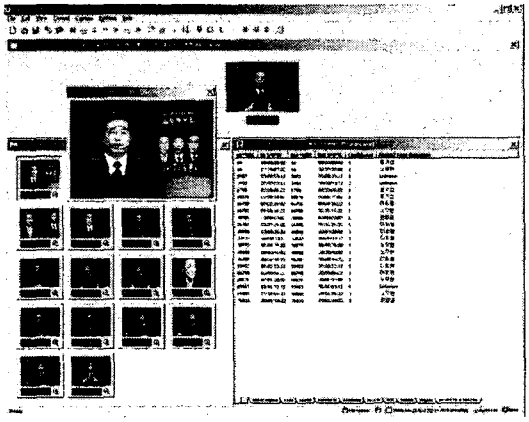

Figure 4. Result of plug-in

The standalone version of Face Register and Face Recognizer can be transformed into a plug-in for cataloguing system. The plug-in automatically generates and adds metadata (e.g. name, appearance interval, etc) of each character when cataloguing system indexes a video. We implement the plug-in for the commercial application '*VideoLogger*' which is a product of *Virage*, Inc. The result of plug-in is automatically displayed in the separate page of VideoLogger interface as shown in Figure 4. The metadata generated by the plug-in is usually stored in the database and utilized in the digital archive through various retrieval functions.

KBS already built up a digital archive system titled "news archive system" and applied it to produce news programs. The implemented plug-in is integrated into the cataloguing system of ours. Through the cataloguing process, the metadata of character are stored in the digital archive and are utilized in retrieval functions on the archive portal. Now, an evaluation about the applicability of the plug-in is in progress through feedback from users visiting this web site. In the near future, we will be able to report the result of the evaluation.

3. Fast Face Recognition Method

This section describes a set of novel algorithms for detecting and recognizing faces for the character retrieval system. In the proposed face detection and recognition algorithms, we reduce the computation time without sacrificing the recognition accuracy by using the DCT/LDA method for face feature extraction.

3.1 Face Detection

The face detection module is implemented based on SVM (1). To detect faces more quickly, the facial skin color, the simple feature, and the face edge model are combined sequentially to detect candidate regions of a face. In TV drama, since the variation of facial size is quite substantial, an image pyramid is constructed with a factor of 1.2 instead of 2 in sub-sampled pyramid. The 24×24 size of mask first scans a video frame and then the video frame is reduced to 1.2 times smaller. This process is equivalent to the scan of video frames with a larger mask.

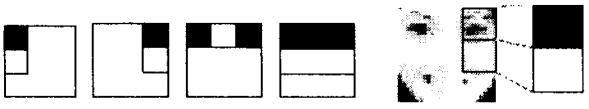

Figure 5. Four types of simple features

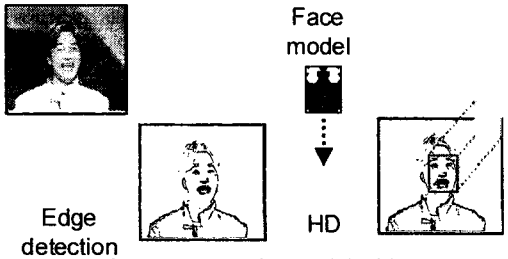

Figure 6. Face edge model with HD

At first, face candidate regions are found by facial skin color (2). The mask scans the face candidate regions and the regions are verified according to four types of simple features shown in Figure 5. These simple features measure the difference of intensity between the mask regions and this operation is performed quickly using integral image (3). If the difference of intensity between the regions is less than a threshold, the mask region is a non-face region.

The detected candidate regions are again verified by the face edge model with the Hausdorff distance (HD) as illustrated in Figure 6 (4). After the verification using the face edge model, the face region is finally determined by SVM. This SVM utilizes pixels to speed up the verification process. As a result, SVM is carried out with 108 coefficients (12 coefficients per each sub-block) instead of 576 pixels.

3.2 Face Recognition

The face recognition method is based on the component-based DCT/LDA. The DCT/LDA is one of the subspace methods of LDA, which significantly reduces the dimension of a signal using DCT. Compared with PCA based LDA (PCA/LDA), one of the advantages of DCT/LDA is to extract features more quickly and show a higher face recognition rate than the PCA/LDA method. The component-based DCT/LDA utilizes facial components such as eyes and nose to improve the performance of face recognition under various types of face variation such as illumination, facial expression, and pose. It extracts features from the holistic face, eyes, and nose. These facial components can be cropped from the predefined region, because a face image is 46×56 in size and is normalized based on position of eyes. Before extracting features, illumination of face image is normalized using Zero Mean Unit Variance (ZMST) (5). DCT coefficients are extracted from an illumination image.

In our approach of face recognition, it compares features of one holistic face and three facial components – left eye, right eye, and nose. An unknown face is recognized as the face of the smallest weighted Euclidean distance among the registered faces in the database.

4. Experimental Results

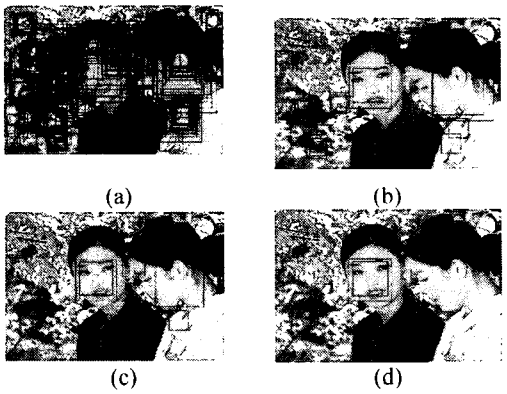

Figure 7. Result of Face Detection

In the face detection procedure, as shown in Figure 7, initial candidate regions are decided by skin color, then the candidate regions are verified by both the simple features and the face edge model with HD. Finally they are further verified by SVM trained with 296 faces and 3736 non-faces using bootstrapping. The computation time is 62 milliseconds in average. The video contains 1,340 face images with several types of face variation. Faces are detected by the method aforementioned and normalized automatically based on eye positions.

The result of the component-based DCT/LDA method is compared with that of MPEG-7 Advanced Face Recognition descriptor (AFRD), which is one of the best face recognition methods based on PCA/LDA (6). The component-based DCT/LDA is trained with the same conditions as AFRD was trained (7). The average rates of 20 face recognition are shown in Figure 8. Although AFRD shows the face recognition rate of 68.88%, component-based DCT/LDA shows face recognition rate of 77.14%. Moreover, it shows the much higher performance, 83.13%, than AFRD when it is trained with the KBS face dataset which comes from different sources. The reason why the component-based DCT/LDA shows the high performance is that it utilizes appropriate DCT coefficients extracted from each facial component for LDA discrimination.

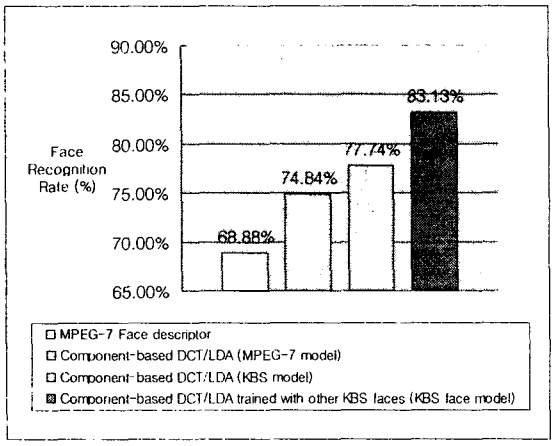

Figure 8. Result of Face Recognition

5. Conclusion

In this paper, we have presented a character retrieval system based on face recognition. The system consists of two modules: Face Register and Face Recognizer. The function of Face Register is to detect faces in video frames and allow users to register the faces into a database interactively. For the face detection, candidate face regions are first selected using the facial skin color, the simple feature, and the face edge model. To detect faces of variable size, a pyramid structure of the image is utilized. SVM then verifies whether the candidate region is a face or not. The detected faces are recognized using the component-based DCT/LDA method along with the face tracking method. Face Recognizer analyses the appearance intervals of each character using the proposed face recognition method and displays these intervals on the timeline. In case face detection and recognition are performed every third frame, the execution time spends 150 milliseconds per frame (in average) and shows 83.13% of accuracy in a dataset of KBS videos.

Face recognition in the broadcasting video is known as a difficult task due to substantial variations of illumination, facial expression, and poses. Registration errors, different hairstyles, and background noises are additional factors that confuse the face recognition. Therefore, it is important for a system not only to recognize faces accurately but also to modify the system-generated recognition results easily, which is the most essential function for feasibility of application. For this reason, both Face Register and Face Recognizer provide utilities to edit incorrect results of face recognition. For efficiency in editing the recognition results, a grouping method of unknown faces of the same person is added to avoid repeated editing if the unknown is the same person.

The character retrieval system is developed in two forms: standalone and plug-in. The plug-in is easier to apply to the cataloguing system of the digital archive system. We have already integrated the plug-in to our digital archive system. Moreover, we are evaluating the feasibility and applicability of our system through feedback from users of KBS archive portal.

Reference

[1] E. E. Osuna, R. Freund, and F. Girosi. "Support vector machines: Training and applications." AI Memo 1602, AI Lab, Massachusetts Institute of Technology, March 1997.

[2] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K. Jain, "Face detection in color images," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.696-706, 2002.

[3] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features,"Computer Vision and Pattern Recognition, Vol. 1, pp. 8-14, December 2001.

[4] O. Jesorsky, K. Kirchberg and R. Frischholz, "Robust Face Detection Using the Hausdorff Distance." In Proc. Third International Conference on Audio- and Video-based Biometric Person Authentication, Springer, Lecture Notes in Computer Science, LNCS-2091, Halmstad, Swden, pp. 90-95, June 2001.

[5] Yongping Li, "Linear Discriminant Analysis and its application to Face Identification," Ph. D. Thesis Paper, University of Surrey, September 2000.

[6] S.K.Kim and A. Yamada, "Description of MPEG-7 Visual Core Experiments," ISO/IEC JTC1/SC29/WG11 N5699, July 2003.

[7] M. Bober ed., "Description of MPEG-7 Visual Core Experiments," ISO/IEC JTC1/SC29/WG11 N4925, October 200.