*2011 International Conference on Computer Science and Information Technology (ICCSIT 2011)*
***IPCSIT vol. 51 (2012) © (2012) IACSIT Press, Singapore***
*DOI: 10.7763/IPCSIT.2012.V51. 112*

# Face Video Key-Frame Extraction Algorithm Based on Color Histogram

Xintao Li* and Tiongrong Xu

School of Computer Science & Technology, Soochow University, Suzhou, Jiangsu 215006, China

**Abstract.** Video-based face recognition has been widely studied in recently years. The common method of video face recognition is to recognize of every face frames that collected. The method not only cause large computing complexity, but also when face gesture and illumination change greatly, the face images collected contain very little information that can be used, these samples may have a negative effect to the results of recognition and even cause the recognition failure. Therefore, this paper proposes a method for facial video key frame extraction based on color histogram, the purpose is to remove the redundant frames, reduce the computational complexity and improve recognition efficiency.

**Keywords:** video; face recognition; key-frame extraction; colour histogram.

## 1. Introduction

In the system of video-based face recognition, it is commonly to recognize of every face frames that collected. Consequently, the system not only has a large amount of calculation, but also when face gesture and illumination change greatly, the face images collected contains very little information that can be used, these samples may have a negative effect to the results of recognition and even cause the recognition failure. Therefore, key frame extraction algorithm is used in this paper to solve the issue.

There are many key frame extraction methods. In earlier works on video summarization, key frames are selected by sampling video frames randomly or uniformly at certain time intervals [1]. This approach is simple and fast but neglects the video content. Another simple and intuitive method is to extract the first frame and final frame of each shot as key frames in [3], but the key-frames extracted are inaccurate and the method only can get fixed number of key frames. Zhang [4] proposed an unsupervised clustering algorithm to extract key frames. This method can not effectively maintain the dynamic information in the original lens. Cernekova [5] clustering on the inter-frame mutual information and obtained good results. Reference [6] gives a visual content-based key-frame extraction algorithm which can extract the key-frame dynamically. Wolf [7] computed the optical flow for each frame and then used a simple motion metric to evaluate the changes in the optical flow along the sequence.

In this paper, we propose an algorithm based on color histogram aimed at extracting key frame from face video, which can be used as a separate module in video face recognition system. The organization of this paper is as follows. In section 2, we focus on the color histogram feature extraction. In section 3, the key frame extraction algorithm is given. Experiment results are shown in section 4. Concluding remarks are provided in section 5.

## 2. Feature Extraction

Color feature is widely used in the object recognition, it is a numerical measure of the image in the color space. Color features include color histogram, color moment, color consistency and color relevant graph.

+ Corresponding author.
*E-mail address:* 20084227065051@suda.edu.cn, trxu@suda.edu.cn

Color histogram is an effective method which can describe the color information, it is widely used in image processing because of its simple computation, high computing efficiency and normalized against the impact of image size.

## 2.1. Selection of color model

Color difference distinguished by the traditional RGB model is nonlinear and is not intuitive. However, HSV color model is close to human visual characteristics, which consists of hue (H), saturation (S) and value (V). Because this model has a linear scalability and the perceived color is proportional to Euclid distance of the corresponding values of color components, HSV color model is more intuitive and more easily accepted than RGB color model. The transformation from r, g and b values in RGB color space to h, s and v values in HSV color space can be described as follows[8]:

$$v = \mathbf{max}(r, g, b) \tag{1}$$

$$s = \frac{\mathbf{max}(r, g, b) - \mathbf{min}(r, g, b)}{\mathbf{max}(r, g, b)} \tag{2}$$

Let:

$$r^{'} = \frac{v - r}{v - \mathbf{min}(r, g, b)} \tag{3}$$

$$g^{'} = \frac{v - g}{v - \mathbf{min}(r, g, b)} \tag{4}$$

$$b^{'} = \frac{v - b}{v - \mathbf{min}(r, g, b)} \tag{5}$$

$$h^{'} = \begin{cases} 5 + b^{'},\ r = \mathbf{max}(r, g, b)\ \ and\ \ g = \mathbf{min}(r, g, b) \\ 1 - g^{'},\ r = \mathbf{max}(r, g, b)\ \ and\ \ g \neq \mathbf{min}(r, g, b) \\ 1 + r^{'},\ g = \mathbf{max}(r, g, b)\ \ and\ \ b = \mathbf{min}(r, g, b) \\ 3 - b^{'},\ g = \mathbf{max}(r, g, b)\ \ and\ \ b \neq \mathbf{min}(r, g, b) \\ 3 + g^{'},\ b = \mathbf{max}(r, g, b)\ \ and\ \ r = \mathbf{min}(r, g, b) \\ 5 - r^{'}, \qquad others \end{cases} \tag{6}$$

Then:

$$h = h^{'} \times 60 \tag{7}$$

Here: $h \in [0, 360]$, $s, v \in [0, 1]$.

## 2.2. Color space quantization

The original face images in video sequence are generally true color images, which contain 24-bit colors and 8 bits per color component, if we calculate the color histogram of the three-dimensional images directly, we get 16 million dimensional color vector, it's a large computation when calculate the distance of histogram. So take into consideration of the instantaneity, we have to quantify the level of color histogram.

Color quantization means to designate a group of colors to represent the image color space, and then map the color space to the colors selected. There are many ways to quantify the color, such as vector quantization, clustering methods or neural network methods, the most common method is to partition color space use the same interval or different interval. In this paper, we use the later method. After the transformation of RGB color space to HSV color space, we use the following formula to quantify [9]:

$$h = \begin{cases} 0, & h \in (315,360] \cup [0,20] \\ 1, & h \in (20,40] \\ 2, & h \in (40,75] \\ 3, & h \in (75,155] \\ 4, & h \in (155,190] \\ 5, & h \in (190,270] \\ 6, & h \in (270,295] \\ 7, & h \in (295,315] \end{cases} \tag{8}$$

$$s = \begin{cases} 0, & s \in [0,0.2] \\ 1, & s \in (0.2,0.7] \\ 2, & s \in (0.7,1] \end{cases} \tag{9}$$

$$v = \begin{cases} 0, & v \in [0,0.2] \\ 1, & v \in (0.2,0.7] \\ 2, & v \in (0.7,1] \end{cases} \tag{10}$$

According to the quantitative level above, we combine the three color component into a one-dimensional vector:

$$G = HQ_S Q_V + SQ_V + V \tag{11}$$

Here, $Q_S$ and $Q_V$ are the quantitative level of $S$ and $V$ component. Let $Q_S$ =3, $Q_V$ =3, so the above equation turns into:

$$G = 9H + 3S + V \tag{12}$$

By this way, three components of H, S and V are distributed on one-dimensional vector. According to the above equation, $G$ is range of [0, 1, ⋯ , 71]. This means the whole color space is mapped into 72 kinds of colors, which can effectively compress the color features.

## 2.3. Color Histogram

Color histogram [10] reflects the probability of one kind of color pixels in an image. It is the estimation of pixels probability. Given a digital image $I$, its color histogram vector can be described as:

$$H = \left( h[c_1], h[c_2], \cdots, h[c_k], \cdots, h[c_N] \right), \quad \sum_{k=1}^{N} h[c_k] = 1, \ 0 \le h[c_k] \le 1 \tag{13}$$

Here, $h[c_k]$ is the probability of the color $C_k$:

$$h[c_k] = \frac{\sum_{i=0}^{N_1} \sum_{j=0}^{N_2} \begin{cases} 1, & I(i,j) = c_k \\ 0, & otherwise \end{cases}}{N_1 * N_2} \tag{14}$$

$N_1$, $N_2$ indicate the input image has $N_1$ rows and $N_2$ columns.

$I\,[i\,,j]$ is the pixel value of point $(i, j)$.

Assume that $G$ and $H$ is the color histogram vector to be compared, $N$ indicates the color level in the image, $g_k$, $h_k$ indicates the probability of color level $k$ in the color histogram of image $G$ and $H$. Then the similarity of $G$ and $H$ can be measured by the Euclidean distance between them:

$$d(G,H) = \frac{1}{N} \sum_{k=1}^{N} \left( g_k - h_k \right)^{\frac{1}{2}} \tag{15}$$

# 3. Key-Frame Extraction

Color histogram-based key frame extraction algorithm is described as follows:

*a)* Initialize, select one face image of good quality from face video sequence as the standard face graph, and calculate the color histogram $H$ as a standard color histogram.

*b)* Get the video sequence frame by frame, and calculate its color histogram $G_i$, then compared with the standard color histogram, calculate the Euclidean distance.

*c)* Set the threshold $T$ ($0 \leq T \leq 1$), $T$ can control the number of key frames selected. Take all the tested frames whose Euclidean distance is less than $T$ as the key frame face images.

The algorithm in details can be described as figure 1.

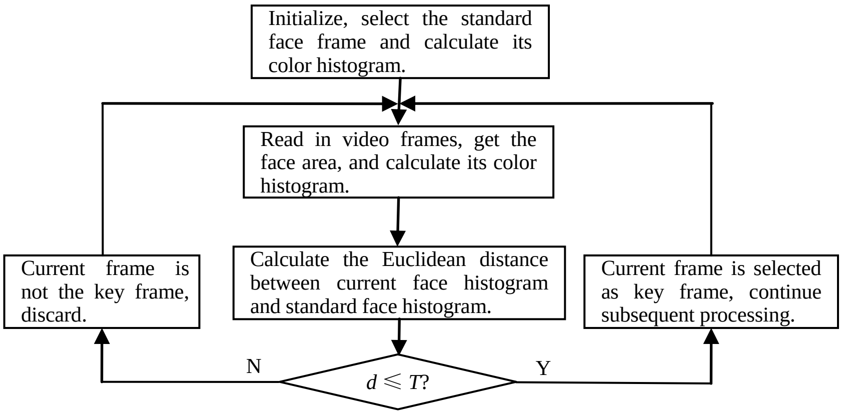



Figure 1. the process of face video key frame extraction

## 4. Experiment Results and Analysis

Experiment is carried out on the NRC-IIT face video database [11]. This database contains 11 individuals, each individual has two video clips in the “avi” format. The video clips are recorded in the office of the research institution by digital camera. The face images in the video have a great range of facial express and orientation changes, as well as scale changes and facial occlusion. Select the second clip of the first person for experiment, and select the first face frame as the standard face graph. Fig.2 gives the color histogram of the standard face. Fig.3 and Fig.4 are the face color histogram of the second frame and twentieth frame. Fig.5 gives the key frame results when the threshold is 0.15. Table 1 gives the extraction results under different thresholds. Table 2 gives the key frame extraction results on different video clips.

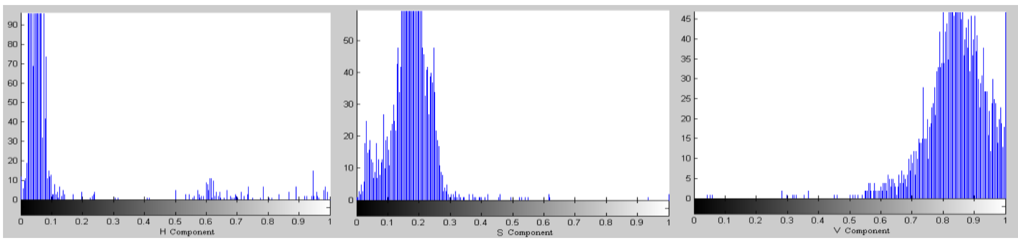



Figure 2. color histogram of the standard face image

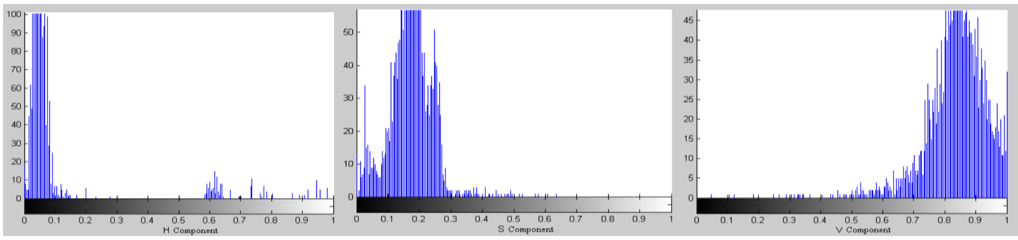



Figure 3. color histogram of the second frame

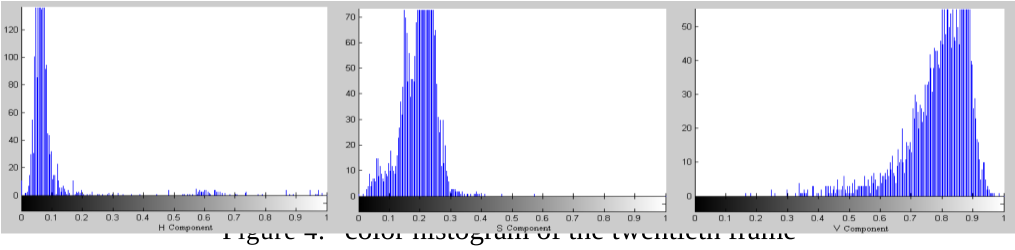
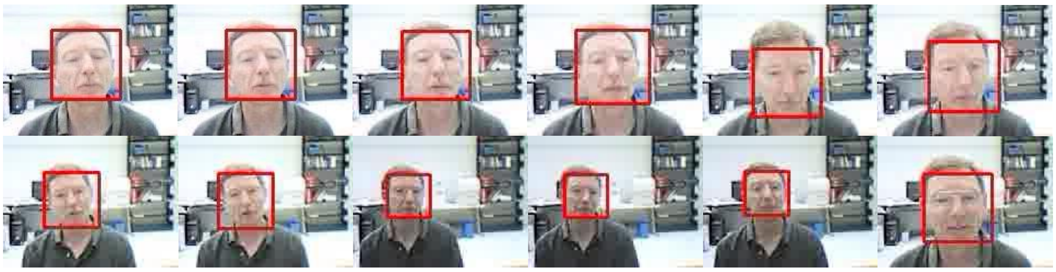
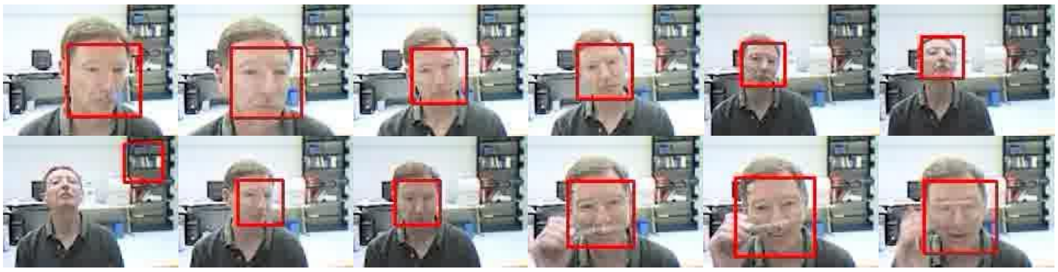



Figure 4. color histogram of the twentieth frame



(a)



(b)

Figure 5. (a) key frame samples,$T$=0.15. (b) non-key samples,$T$=0.15.

Results can be seen from the Fig.5. The key frame extraction algorithm can extract the face frames which have little change in the rotation and scale. From the results in the table 1 and table 2, we know that changes of threshold value can control the number of key frames selected, meanwhile, different video clips need select different threshold. In actually, we set the threshold manually according to the characteristics of video and the results of key frame extraction.

TABLE 1. FACE KEY FRAME EXTRACTION RESULTS UNDER DIFFERENT THRESHOLD

| **threshold(*T*)** | **key frame extracted** | **total** |
|---|---|---|
| 0.1 | 1-14, 50, 51, 108-117, 132-138 | 33 |
| 0.15 | 1-32, 50, 51, 69-80, 102-117, 131-138, 150, 233-239 | 78 |
| 0.2 | 1-36, 50-54, 68-117, 131-138, 149-155, 223-241 | 125 |
| 0.3 | all face frames | 165 |

TABLE 2. KEY FRAME EXTRACTION RESULTS ON DIFFERENT VIDEO CLIPS

| **video 00-1 （frame amount /threshold）** | **video 01-1 （frame amount /threshold）** | **video 02-1 （frame amount /threshold）** | **video 03-1 （frame amount /threshold）** | **video 04-1 （frame amount /threshold）** |
|---|---|---|---|---|
| 120/0.2 | 29/0.01 | 78/0.1 | 31/0.005 | 23/0.01 |
| 143/0.25 | 37/0.02 | 154/0.15 | 204/0.01 | 102/0.05 |
| 188/0.3 | 80/0.03 | 195/0.2 | 287/0.02 | 156/0.1 |

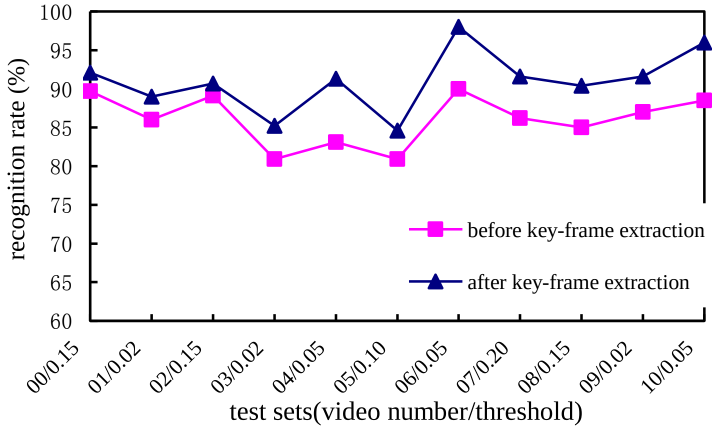



Figure 6. comparation of correct recognition rate before and after key frame extraction

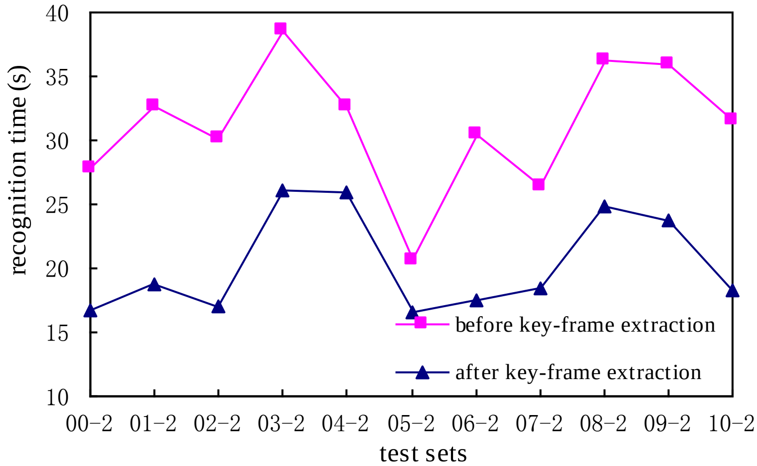



Figure 7. comparation of recognition time before and after key frame extraction

Key frame extraction greatly reduces the redundant information of the face video, so subsequent processing can only focus on the key face frames. Take the second video clip of each person in NRC-IIT database for experiment, use elastic graph match method[12](select 30 feature points), and recognize every face graph, finally count the proportion of correct identification face frames in the total face images as recognition rate. Statistical result is shown in Fig.6. Recognition time is shown in Fig.7.

From the results in Fig.6, we see that correct recognition rate is obviously improved after key frame extraction, this indicates that key frame extraction removed facial image frames easily to get wrong results and retained positive facial images, the overall effectiveness of video information is improved. Generally, key frame number is far less than original video frame number (decided by threshold), after key frame extraction, feature extraction and recognition is only focus on key frames, so greatly reduce recognition time.

## 5. Conclusion

In this paper, we propose an approach for face video key frame extraction. The algorithm is based on color histogram, which is simple for computation and normalized against the impact of image scale. By means of key frame extraction, most redundant information was reduced, follow-up processing can only implement on the useful information. Therefore, the method reduces the computational complexity and improves recognition efficiency. It can be used as a separate module in video face recognition systems. In this paper, the threshold is set by manual and based on experience. Our feature work is to set the threshold automatically and adaptively.

## 6. References


[1] Y. Taniguchi, A. Akutsu, Y. Tonomura, and H. Hamada. An intuitive

[2] and efficient access interface to real-time incoming video based on automatic indexing [J]. In Proceedings of the

third ACM international conference on Multimedia, ACM, New York, USA, 1995, pp. 25-33.

[3] Y. Tonomura, A. Akutsu, K. Otsuji, and T. Sadakata. VideoMAP and VideoSpaceIcon: Tools for anatomizing video content [J]. INTERCHI' 93 Conference Proceedings, 1993, ACM Press, pp. 131-141.

[4] Y. Zhang, Y. Rui, T. S. Huang and S. Mehrotra. Adaptive key frame extraction using unsupervised clustering [J]. In Proc. Of 1nt. Conf. On Image Processing, Chicago, 1998.884-887.

[5] Z. Cernekova, 1. Pitas, and C. Nikou. Information theory-based shot cut/fade detection and video summarization [J]. IEEE Transaction on Circuits and Systems for Video Technology, 2006, Vol. 16, No. I: pp.82-91.

[6] H. Kui, X. O. Chen. A key frame extraction method without considering video format [J]. Computer Application, 2003, 23(06):189-190.

[7] W. Wolf, Key frame selection by motion analysis [J]. In Proc. IEEE Int. Conf. Acoust., Speech Signal Proc., vol. 2, May 1996, pp. 1228-1231.

[8] Sun Jixiang, Image processing [M].Science press.2005:50-51.

[9] Zeng Zhiyong, Zhou Lihua. A novel region-based image retrieval technique using special color distribution feature [J].IEEE, ASPX, 2008.

[10] Zhang Bianli, Chang Shengjiang, Li Jiangwei, Wang Kai. Intelligent video surveillance system based on color histogram [J]. Physics Journal. 2006, 55(12):6399-6403.

[11] Dmitry O. Gorodnichy .Video-based framework for face recognition in video [J]. Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), pp. 330-338, Victoria, BC, Canada, 9-11 May, 2005.

[12] Wiskott L, Fellous J M, Kruger N, etal. Face recognition by elastic bunch graph matching [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7):775-779.