Design and Implementation of Video Content Classification Algorithm Based on Face Recognition

Xiaokang Wu [a], Chenggang Xie [b], Shu Tang and Yuanbo Yang

National University of Defense Technology, Changsha 410073, China

[a]172896292@qq.com, [b]qingqingzijin_k@126.com

Keywords: face recognition, video content, video classification, and key frame.

Abstract. To further facilitate users' quick browsing the behaviors and expressions of the characters in videos, in which they are interested, we need to classify the according person's frame images in accordance with the contents of the group scene. In this paper, we use the technology of face recognition to classify the frame images of different person which are coupling in the video, and then we form the video abstraction of different person.

Introduction

In the substance level of video, there are four structures, they are frames, lenses, scenes and videos. Classifying the video content refers to classifying the frames, lenses and scenes of a video according to the different content. The frames, lenses and scene are the minimum logic units when we use them to classify a video. Due to the large number of frame layers and complicated scenes, both of them are inappropriate to the objects to classify. Generally, we use the contents of the lens's key frame as the classifying and grouping criterion. Classifying of video content is an important issue, however, there are no scholars exploring and doing research on it thoroughly. This poses a great challenge for the realization of classifying of video content.

In this paper, we use the video which is related to the person as processing objects and classify the persons who are appeared on the video, then we can obtain the key frame collection of every object, so that users can quickly browse the behavior and expression of the characters who they are interested in. After analyzing the specific needs of classifying frame images when we process the video data of person, we proposed a method that we use the face detection and recognition as the criterion of classifying the content in the video. On this basis, we designed and realized the video content classification algorithm based on face recognition, at last, analyzed and summarized the testing results of this algorithm.

Requirement of Analysis

People have a variety of behaviors in video, including body moving, expression transformations and so on. We can get the important frame of the video by detecting the features of the human shape, but there is almost no difference in those features when we used the data to describe each person's body, so we can't discriminate them clearly by using this method when there are many people appearing in a video. For those in a video, the most significant feature is the facial feature.

Obtaining the top face information of the object preliminary by using the face detection, and thereby we can remove the irrelevant and the blur background. After doing personal identification based on the human facial features and classifying the face in the video into the corresponding set of the object, we can obtain each person's information.

Face Recognition

Since we need to extract and classify the related frame in which the target exist, we have to do a detection and recognition of human face in the image.

The main technology is consists of three components:

Face detection, this is the foundation of the face recognition. Its main task is to locate the person's face, cut out the face and normalize the face's size from the frame image.

Facial feature extraction, this is the key to recognize a face. By extracting the face feature which has the strong capabilities in characterization, we can do well in distinguishing different objects to achieve better results in recognition.

Facial feature match is the core issue of the face recognition technology. And the main task of it is to design and make a sound selection of the classification criteria in face recognition.

A face detection method based on Attentional Cascade was proposed by P.Viola, which used a cascade structure to improve the speed of face detecting system, and the main idea of this design is to increase the accuracy of this detection progressively. Firstly, he used the strong classifier which is rather simple in structure(less number of) to preclude the window which the faces didn't exist in. Then the number of weak classifiers in subsequent strong classifier is increasing, and the detecting precision keeps getting higher with the number of sub-window which needs to detect becoming less and less, so as to achieve the purpose of improving the speed of detection.

The linearsub-space is one of the main methods of the facial feature extracting at present. The linearsub-space is to find a linear or non-linear spatial alternation based on fixed performance objective, compress original data signal into a low-dimensional subspace in order to make the data in subspace keep the maximum information in the original space. There is no need for the PCA to know and extract geometric knowledge of human face. It is a simple, fast and practical algorithm for the feature of transforming coefficient. The low-dimensional face got by the PCA is similar with facial shape, so it can be used to locate a human-face.

In facial feature matching, if the extractive facial feature can form a column vector x in some order, we can compute the distance between the facial feature vector x which is to be matched and the facial feature vector y in sample symbols to calculate similarity between x and y, the distance is defined as: $d(x,y)=\|x-y\|=\left[\sum_{i=1}^{n}(x_i-y_i)^2\right]^{1/2}$.

You can also use the cosine of included vector angle between them to measure the similarity. The cosine can be defined as: $\cos(x,y)=\frac{x^{\mathrm{T}}y}{\|x\|\|y\|}=\frac{x^{\mathrm{T}}y}{\left[(x^{\mathrm{T}}x)(y^{\mathrm{T}}y)\right]^{1/2}}$.

The smaller the distance between them is, the higher their level of similarity is. The closer the cos(x,y) are to 1, the bigger the value of the two feature vectors' similarity is.

Classification Algorithm

We sign a video data of human as V and suppose the number of processing object is n. We sign their facial feature vectors as $U=\{u_1,u_2,\cdots u_n\}$ and build $n+1$ sets for classifying. Sign n results of detecting observed object as $\{R_1,R_2\cdots R_n\}$ and one set of unrecognized staff as R_{n+1}. The implementation procedure is as following:

First step: Decompose V to get the image frame sequence. $A_1=\{f_1,f_2\cdots\}$

Second step: Preprocess the image frame and use the isometric sampling method that the step-size is k to obtain image sequence. $A_2=\{f_1,f_{k+1},f_{2k+1}\cdots\}$

Third step: Do background subtraction and frame difference filtering for image frame in A_2 to do away with the redundancy information and then get key-frames. $M=F_x(A_2)=\{f_{key_1},f_{key_2}\cdots\}\subseteq A_2$

Fourth step: Detect face in frame f_{key_i} and crop, correct and normalize the size after positioning the face $I(f_{key_i})=\{I_1,I_2,\cdots I_p\}$, p presents the number of detected human face in f_{key_i}

Fifth step: Extract the main features of the face images $\{I_1, I_2, \cdots I_p\}$ respectively, that is $v_1, v_2, \cdots v_n$

Sixth step: Compare the facial feature $v_1, v_2, \cdots v_n$ with the features in the face database $U = \{u_1, u_2, \cdots u_n\}$ one by one and get the Similarity $S(u_i) = \max(v, U)$ between them, then determine the identity of human faces

Seventh step: Mark this frame image and send it into the set R_i of u_i corresponding person.

Repeat the steps until it detects all key frame images in M.

Eighth step: Generate a video summary $\{\tilde{V}_1, \tilde{V}_2 \cdots \tilde{V}_n\}$ for the results of detecting faces $\{R_1, R_2 \cdots R_n\}$ and semantic description of each collection.

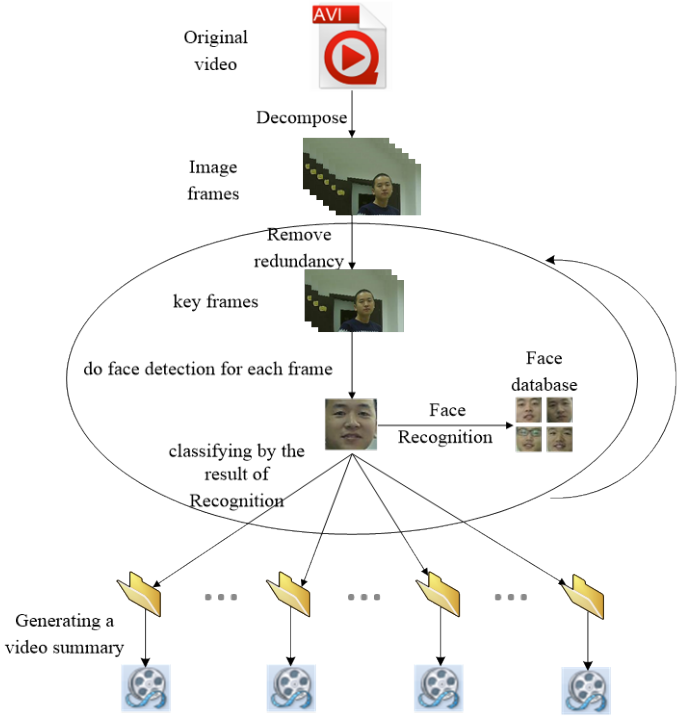

Fig. 1 process of video content classification

Result Analysis

The camera we used in the experiment is 1080P and has 2 million pixels. There is one step which is doing facial feature matching among face database in the process, so we need to build a face database.

In this paper, we gather different attitude information of ten people and obtain the average size of the face in each person's training set for matching.

We recorded three videos in laboratory. Every video has three observation objects. Sometimes the objects show up respectively, sometimes they both all appear in the video.

The three objects appeared in the video early or late, and we shoot the video indoors, outdoors and in dimly lit interiors respectively. Experimental results are as following:

Fig. 2 The matching effect of object A

Table 1 Experimental results

Video name	Scene1	Scene2	Scene3
Time	42.24s	47.32s	54.40s
All frames	1056	1183	1360
Key frames	42	47	54
Facial frames	24	25	23
Multiplayer frame image	6	4	0
Frames to A	6	9	10
Right classifying frames to A	4	7	4
Frames to B	4	6	7
Right classifying frames to B	4	4	3
Frames to C	8	6	5
Right classifying frames to C	6	4	2
Rate of accuracy	77.78%	71.43%	40.91%
Time-consuming (s)	1304.16s	1460.37s	1650.79s
Average time-consuming	30.88s	30.87s	30.35s

From the experimental results, we know that the effect is not bad. The rate of accuracy in in dimly lit interiors is being reduces, and the reason is that the differences among facial features have being decreased.

Summary

This article is for users to browse and query videos in everyone's behavior movements and facial expressions to form an initial impression of the different target audiences conveniently. We proposed video classification method based on face detection and recognition. In terms of the face detection based on Haar-like characteristics, PCA method is mainly used in face recognition. Then on this basis, we designed video classification algorithm based on face detection and recognition. Test results show that whether the facial feature extraction has good discriminative influence classification or not the results are great. Besides, to get the rid of the week point in the processing time, we need to do more in faster algorithm.

References

[1] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In CVPR 2011. 2, 4, 6.

[2] PAUL VIOLA, MICHAEL J. JONES, Robust Real-Time Face Detection. International Journal of Computer Vision 57(2), 137–154, 2004 Kluwer Academic Publishers. Manufactured in the Netherlands

[3] Paul Viola and Michael J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR, 2001.

[4] Xiaokang Wu, Chenggang Xie, Qin Lu. Algorithm of Video Decomposition and Video Abstraction Generation Based on Face Detection and Recognition[C]. (ICCDIM 2014), June 22-23, 2014, Guilin Guangxi, China.