ISSN: 2321-7782 (Online)

Special Issue, December 2013

International Journal of Advance Research in Computer Science and Management Studies

Research Paper

Available online at: www.ijarcsms.com

A Survey on Video Summarization using Face Recognition Methods

S. Sangeetha[1]
PG Student
Department of CSE
Velammal Engineering College
Chennai – India

S. Deepa[2]
Faculty
Department of CSE
Velammal Engineering College
Chennai – India

Abstract: Video Surveillance is the monitoring of the behavioural activities. Video Surveillance systems are not only needed to track the moving objects, but also to interpret by solving the given information and to combine the samples of behaviour. Human Face Detection and tracking of moving objects in the video are important tasks of the computer vision. This paper presents a long-take video shots, in which a particular face is recognized and a restriction to perform a video with a temporal sequence. In the projected system an unauthorized person can be identified by analyzing in different angles. The frames taken from various angles are grouped together in the form of a video that has been extracted from the existing collection of videos. Included with the videos, temporal detail is also being collected. A novel framework to compose descriptive long-take video with content-consistent shots retrieved from a video pool based on time series.

Keywords: Face recognition; one-shot video; temporal measurement.

I. Introduction

The video surveillance system is the most important issue in the homeland security ground. It is used as a security system because of its facility to track and to detect a particular person. With the popularity of private digital devices, the amount of home video data is growing explosively. These digital videos have several characteristics: (1) compared with former videos recorded by non-digital camcorder, nowadays videos are usually captured more casually due to the less constraint of storage, and thus the number of clips has been often quite large; (2) many videos may only contain a single shot and are very short; and (3) their contents are diverse yet related with few major subjects or events. The user frequently needs to preserve their own video clip collections captured at different locations and time. These unedited and unorganized videos bring difficulties to their management and handling. For example, when users want to share their narrative with others over video sharing websites and social networks, such as YouTube.com and Facebook.com, they will need to put more efforts in finding, organizing and uploading the smaller video clips. This could be an extremely difficult "problem" for users. Previous efforts towards efficient browsing such large amount of videos mainly focus on video summarization.

Image processing operations can be generally divided into three major categories, Image density, Image enrichment and renovation and Measurement Extraction.

II. Video Surveillance

Surveillance is the monitoring of the performance, actions, or other changing in sequence, generally people for the purpose of influencing, organization, directing, or protecting them. This can include inspection from a distance by means of electronic equipment (such as CCTV cameras), or interception of automatically transmitted information (such as Internet traffic or phone calls); and it can submit to uncomplicated, comparatively no- or low-technology methods such as human intellect agents and postal interception. The video surveillance system is the most important issue in the homeland security ground. It is used as a security organization because of its facility to track and to detect a particular person.

Fig.1 Surveillance cameras

To overcome the lack of the conservative video surveillance system that is based on human imminent, we introduce a narrative cognitive video surveillance system (CVS) that is based on mobile representatives. CVS (Customer Value and Service) proposes significant features such as expect object recognition and smart camera hold up for public tracking. According to numerous studies, an agent-based approach is proper for disseminating structures, since movable agents can relocate copies of themselves to other servers in the system.

III. FACE RECOGNITION SYSTEM

Face recognition presents a challenging problem in the field of image analysis and Computer visualization, and as such has received a great deal of consideration over the last few years because of its many applications in a variety of domains. Face recognition techniques can be broadly divided into three categories based on the face data acquirement methodology; methods that operate on intensity images; those that deal with video sequences; and those that require other sensory data such as 3D information or infra-red imagery. [10] Face recognition systems are comprised of three steps. Their basic flowchart is given in (Fig.2.) Among them, detection might include face edge detection, segmentation and localization, namely obtaining a pre-processed intensity face image of an input picture, moreover simple or cluttered, locating its position and segmenting the image out of the background. Feature extraction may denote the acquirement of the image features from the image such as visual features.

Applications

Face recognition is used for two main tasks:

1. Verification (one-to-one matching): When presented with a false image of an unidentified entity along with a state of identity, determining whether the person is who he/she claims to be.
2. Identification (one-to-many matching): Given an image of an unidentified person, determining that person's identity by comparing (possibly after encoding) that image with a record of (possibly encoded) images of known persons

There are some practical areas in which face recognition can be demoralizing for these two reasons, a few of which are outlined below.

- Security (access control to buildings, airports/seaports, ATM machines and border; computer/network security ; email authentication on multimedia workstations).
- Surveillance [11].

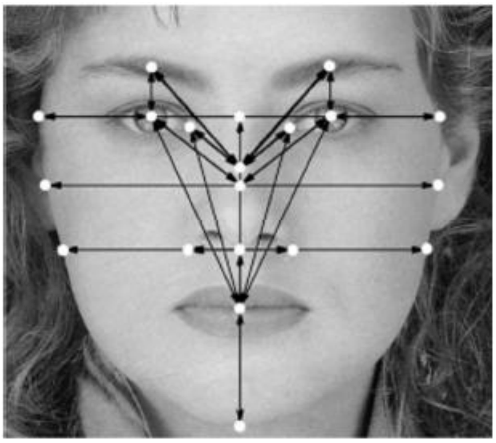

Fig.2.Some facial points and distances between them are used in face recognition.

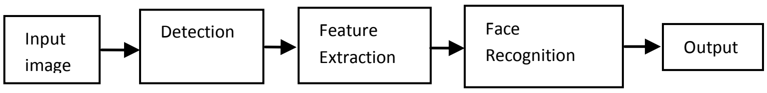

Fig.3. The basic flowchart of Face Recognition

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A. *Video Based Face Recognition*

Video based face recognition [12]in image sequences has increased curiosity based mainly on the idea expressed by psychophysical studies that motion helps humans be familiar with the faces, particularly when the spatial image excellence is low. Video-based face recognition systems consist of three modules: a discovery module, a tracking module and a recognition module. Given a frame of a video sequence, the detection component locates face applicants, while the tracking module finds the exact location of facial account in the current frame based on an estimate of face or feature locations in the preceding frame(s). The recognition module identifies or verifies the face, integrating information from previous frames[1].

B. *Face Detection And Localization*

Face detection and localization from images is an input problem and a crucial first step in face recognition systems, with the motive of localizing and extracting the face region from the surroundings. It also has some requests in areas such as content-based image recovery, video coding, video conferencing, mass surveillance, and intelligent human-computer interfaces. In terms of demonstration process used the approaches to face detection might fall into two main categories:[10]

(1) Local feature-based ones; (2) global methods

Viola and Jones propose an efficient machine learning approach for combining a small set of features from a large set to detect features in images. During the training stage, a weighted ensemble of weak classifiers is trained to distinguish faces from other objects, where each weak classier operates on a particular feature. A variant of the AdaBoost learning algorithm chooses the weighted combination of weak classifiers and, hence, the grouping of features that offers the best classification performance on the instruction set. The features, Haar-like wavelets, can be worked out with a minute number of processes by using a novel data structure called the essential image. The resultant detector functions on overlapping windows in the input images, determining the approximate positions of features [16].

IV. Video Summarization

Video summarizations are commonly presented as a set of static key frames or dynamic video skims. After extracting the key frames of video sequence there are different options for presenting them to the user. One of the most common video summarization presentation techniques is a storyboard, which is usually a static grid of extracting key frames. According to a recent study on evaluation of video summarization techniques, the storyboard has a capability to give an informative summary of the original video content. However, according to the user studies the storyboards lacked in their representativeness and ability to replace the original video content. These are all the methods used for video summarization.

A. VIDEO TAPESTRIES USING K-MEANS CLUSTERING

A method for summarizing video in the form of a multi scale image. A key frame is selected and then chronological ordering and continuity between the scales are maintained using subset constraints. Our video tapestries combine the best aspects of two common images, providing the visual precision of DVD chapter menus with the information density and multiple scales of a video editing timeline image. In addition, they give continuous changes between zoom levels. Advantages in this paper are key frame selection; BDS is optimized under the constraints that entire frames are conventional or discarded, while sequential ordering and continuity between the scales are maintained using subset constraints. In a static textile generation, our system again optimizes for bidirectional similarity, applying an additional of the time distance term to loosely enforce chronological ordering. Here scale-space continuity is not clearly enforced, but quite inherited absolutely from the subset constraints of key frame selection [6]. Finally, when constructing zoom animations, the stability constraint is forced by constructing and interpolating "islands," the coherence term from BDS is optimized to fill in the remaining space, and chronological ordering is inherited implicitly from the static tapestries. So the Constrained and Unconstrained regions can be taken and the result can be found. Accuracy of the algorithm is enhanced. No clear cut detection preprocessing. Face detector might fail to combustion on unusual faces [1].

i. K-means algorithm

Step 1: Place K points in the space represented by the objects that are being clustered. These points represent initial group centroids.

Step 2: Assign each object to the group that has the closest centroid.

Step 3: When all objects have been assigned, recalculate the positions of the K centroids.

Step 4: Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

ii. Key frame clustering

Clustering problems arise in various areas like pattern recognition and pattern classification, image processing, Bioinformatics etc. It is considered that the k-means algorithm is the best-known squared error based clustering algorithm. It is very simple and can be easily implemented in solving many practical problems.

The k-medoids clustering algorithm to minimize this objective. The input frames at each level are found by taking the known input frames at the previous level, and solving [1]

- K-medoids are also a partitioning technique of clustering that clusters the data set of *n* objects into *k* clusters with *k* known *a* priori.
- It could be more robust to noise and outliers as compared to k-means because it minimizes a sum of general pairwise dissimilarities instead of a sum of squared Euclidean distances. The possible choice of the dissimilarity function is very rich but in our applet we used the squared Euclidean distance.
- A method of a finite dataset is a data point from this set, whose average dissimilarity to all the data points is minimal i.e. it is the most centrally located point in the set.

For the best new key frames to add to the layout, while existing frames are not removed or changed. Thus the key frames at each level are a subset of the frames at the next stage. This programmed key frame selection process is usually effective at highlighting salient information, but it can optionally be replaced with simple sub-sampling in time (e.g. 2 fps), or the user can manually choose key frames to tell the narrative enhanced. Standardized sub-sampling offers the advantage that the final tapestry precedes approximately linearly with time, which may be desirable for video suppression applications where the duration of events is important.

B. Video Composition And Video Summarization

Video summarizations are commonly presented as a set of static keyframes or dynamic video skims. After extracting the keyframes of video sequence there are different options for presenting them to the user. One of the most common video summarization presentation techniques is a storyboard, which is usually a static grid of extracting keyframes. According to a recent study on evaluation of video summarization techniques,the storyboard has a capability to give an informative summary of the original video content. However, according to the user studies the storyboards lacked in their representativeness and ability to replace the original video content.

TECHNIQUES

- Static Summarization
- Dynamic Summarization

Static Summarization

The number of key-frame selected for presenting a shot is tailored to applications. We evaluate static abstraction from two aspects:single key-frame and multiple key-frames. The main concern of single key-frame representation is how informative a key-frame is as a representative icon of shot.

Dynamic Summarization

Dynamic video skimming is a technique that condenses the original video into a shorter version, while preserving important content with its time-evolving properties. Hence, video skims are practically short video clips cut from the original video sequence. Preservation of motion information is one of the great advantages of video skims, in addition to aural information, which can both enhance the expressiveness of the video summary.

Co-occurrence Matrices

Co-occurrence Matrics is useful for feature extraction.It represents an estimate of the probability of that pixel(i1,j1)has intensity z and a pixel(i2,j2) has intensity y.Suppose that the probability depends only on a certain spatial relation r between the pixel of brightness x and a pixel of brightness y then information about the relation r is recorded in the square co-occurrence matrix Cr,whose dimensions correspond to the number of Cr(z,y) for relation r is given by,

Step1:

Assign Cr(z,y)=0 for all z,y ε [0,L),where L is maximum brightness.

Step2:

For all pixels(i1,j1) in the image,determine(i2,j2) which has relation r with the pixel(i1,j1) and perform

Cr[f(i1,j1,f(i2,j2)]=Cr[f(i1,j1),f(i2,j2)]+1

The elements of the co-occurrence matrices represent the value of the probability density function Co(i,j),which measure the normalized frequency in which all pixel pairs within the image,with intersample spacing d along a direction θ,having gray-level values i and j, respectively.The co-occurrence matrix for various degrees is given by,

$Co_{0}{}^{o}$,d(a,b)=|{[k,1),(m,n)] ∈ D: k-m=0,|l-n|=d;

F(k,l)=a,f(m,n)=b|}

$Co_{45}{}^{o}{}_{,d}$(a,b)=|{[(k,l),(m,n)] ∈D: k-m=d,ln=d;

Or(k-m= -d,l-n=d),f(k,l)=a,f(m,n)=|b}

$Co_{90}{}^{o}{}_{,d}$(a,b)=|{[k,l),(m,n)] ∈ D: |k-m|=d,l-n=0;

F(k,l)=a,f(m,n)=|b}

$Co_{135}{}^{o}{}_{,d}$(a,b)=|{[(k,l),(m,n)] ∈ D: |k-m=d,ln=d);

Or(k-m=-d,l-n=-d),f(k,l)=a,f(m,n)|=b}

Where

|{……….}|Refers to set cardinality.

D=(M×N)×(M×N)

Video structure analysis

A video is composed of several video scenes {Sc_1...Sc_n}, each of which depicts an event like a paragraph does in the articles. A video scene is composed of several semantic-related video shots {sh_1...Shan}.A video shot's role is just like a sentence in articles. The visual content of a video shot can be represented by its key frames. A video shot group Sg_i is the intermediate entity between video scenes and video shots, which is composed of several visually similar and temporally nearby video shots[18].

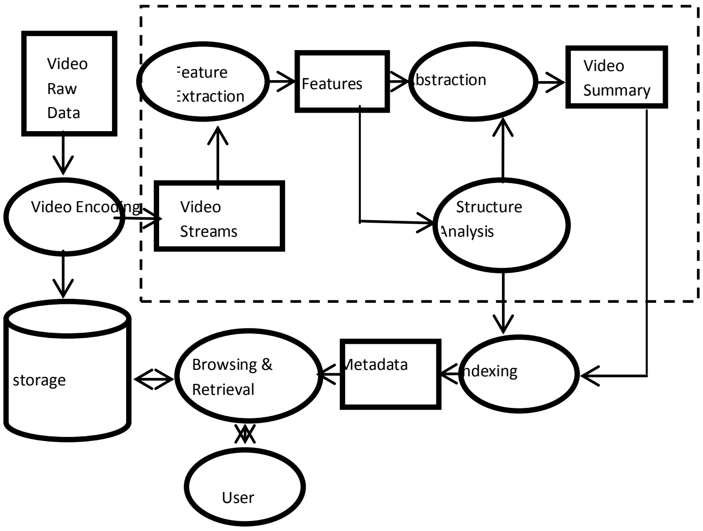

Fig.5 General application of Video Analysis

The detected video scenes can be classified into two types: loop scenes and progressive scenes. A loop scene is composed of more than one video shot groups, while a progressive scene is composed of a series of dissimilar video shots. Loop scenes are often used to depict an event happening in a place that needs detailed description, e.g., a conversation, while the progressive scenes are often used to depict changes between two events. Normally the loop scenes contain more important contents that need repeated illustrations.

For progressive scenes, simply use their length to measure their significance. For loop scenes, they are composed of video shot collections, define the content entropy of a scene Sc_i as:

$$\text{Entropy}(Sc_i)=\sum_{Sg_i \in Sc_i} - \frac{l_{Sg_j}}{,\ c_i} \log_2 \left(\frac{l_{Sg_j}}{l_{Sc_i}}\right)$$

With the above definition, the target video skimmimg length L_{vs} and the length of the video L_v,the skim ratio r_s is thus $\frac{L_{vs}}{L_v}$.

Skim length Sl of each scene and each group in the video as follows:

Step 1:

For each series scene Sc_x,$Sl_x=l_{Scx}\times r_s$. If $S_x<t_1$,it will discard this progressive scene.

Step 2:

Suppose that after the first round,the left skim length is$L^{\square}_{vs}$,for the loop

scenes{Sc_1,Sc_n},$Sl_i=L\square_{vs}\times \frac{\text{Entropy}(Sc_i)\times l_{sc_i}}{\sum nj=1\ \text{Entropy}(Sci)\times lscj}$

In a similar manner discard Sc_i if Sl_i is less than the preset threshold t2.

Step 3:

For the remaining loop scenes{$Sc\square_1 \ldots Sc^{\square}_m$},set $Sl_i=L\square_{vs}\times \frac{\text{Entropy}(Sc\square i)\times l_{sc_i}}{\sum^{m}_{j=1} \text{Entropy}(Sc\square_i)\times l_{sc\square_j}}$

The above skim length assignment algorithm ensures that most important scenes are assigned to more skim length.

V. Conclusion

Video summarization is an important technique for efficient video browsing and management. In this paper, we formulate the video skimming generation problem as a two-stage optimization problem. We obtain the video scene boundaries, determine each video scene's skim length and employ dynamic programming to find each scene's optimal skimming. The whole video skimming is concatenated by each scene's skimming. We implemented the proposed algorithm and obtained encouraging experimental results. In the future, we will further incorporate audio channel analysis to help our skimming generation. Moreover, intra shot compression will be studied to shorten the video shots' length in order to further magnify the content coverage.

References

1. C.Barnes, D. Goldman, E.Shechtman, and A.Finkelstein,"Video tapestries with continuous temporal zoom,"in proc. SIGGRAPH, 2010.
2. K.S.Bhat,.M.Seitz,J.K.Hodgins,P.K.Khosla,"Flow-based video synthesis and editing,"Acm Trans. Graph., vol. 23, no. 3, pp. 360-363, aug.2004.
3. J.Calic,D.Gibson, and N. Campbell,"Efficient layout of comic-like video summaries,"IEEE Trans. Circuits Syst. Video technol.,vol.17,no.7,pp.931-936,jul.2007.
4. Y. Caspi,A. Axelrod,Y. Matsushita, and A. Gamliel," Dynamic stills and clip trailers," Visual Compute., vol. 22,no.9,pp.642-652,sep.2006
5. S. Lu, I. King and M.R. Lyu,"Video summarization by video structure analysis and graph optimization," In Proc.ICME,2004.
6. P. Chiu, A. Girgensohn, and Q. Liu,"Stained-glass visualization for highly condensed video summaries,"In Proc.ICME,2004.
7. D. lee andQ.ke,"Partition min-hash for partial duplicate image discovery,"In Proc.ECCV,2010.
8. Hisateru kato, Goutam Chakraborty, and Basabi chakraborty," A real-time angle- and illumination-aware face recognition system based on artificial neural network" HPC, volume 2012, article id 274617,may 2012.
9. Gulrukh Ahanger and Thomas D.C. Little, Senior Member, IEEE," Automatic Composition Techniques"for Video Production IEEE transactions on knowledge and data engineering", vol. 10, no. 6, november/december 1998.
10. Yongzhong Lu, Jingli zhou, Shengsheng yu," A survey of face detection, extraction and recognition" computing and informatics, vol. 22, 2003.
11. Rabia Jafri,Hamid R. Arabnia,"Aa survey of face recognition techniques",journal of information processing systems, vol.5, no.2, june 2009.
12. Shailaja a patil and Pramod J Deore," Video-based face recognition: a survey" april 2012 .
13. Marryam Murtaza, Muhammad sharif, mudassar raza, and Jamal hussain shah department of computer sciences, comsats institute of information technology," Analysis of face recognition under varying facial expression:a survey", the international arab journal of information technology, vol. 10, no. 4, july 2013.

14. Cha zhang and Zhengyou zhang,"A survey of recent advances in face detection" june 2010.
15. Kandla arora," Real time application of face recognition concept" international journal of soft computing and engineering (IJSCE) ISSN: 2231-2307, volume-2, issue-5, November 2012.
16. Jeremiah R. Barr, Kevin W. Bowyer, Patrick j. Flynn, soma Biswas," Face recognition from video a review", international journal of pattern recognition and artificial intelligence, april 2012.
17. Parvinder S. Sandhu, Iqbaldeep Kaur, Amit Verma, Samriti Jindal, Inderpreet Kaur, Shilpi Kumari," Face Recognition Using Eigen face Coefficients and Principal Component Analysis", International Journal of Electrical and Electronics Engineering 2009.

AUTHOR(S) PROFILE

Sangeetha.S, is currently pursuing her ME in Computer Science Engineering at Velammal Engineering College which is affiliated to the Anna University of Chennai, Tamil Nadu, India.

S. Deepa, completed by Bachelor of computer science and engineering in Sri Venkateswara college of engineering and technology and Master of computer science and engineering in Jaya Engineering college. I have 7.5 years of experience in teaching and I am currently working in Velammal engineering college for the past 2.5 years.