International Journal of Pure and Applied Mathematics
Volume 116 No. 8 2017, 7-11
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

SEAMLESS VIDEO COMPOSITION BASED ON TRANSITION HINTS USING VIDEO SYNOPSIS

[1]Dr.A.R.Arunachalam, [2]Mr.S.Srigowthem
[1]Assistant Professor, Department of CSE, BIST, BIHER,
Bharath University,Chennai.
[2]Assistant Professor, Department of CSE, BIST, BIHER,
Bharath University,Chennai.
Email : [1]arunachalam.cse@bharathuniv.ac.in

Abstract: Video synopsis plays a vital role for extracting the necessary frames from a long video. Video synopsis or video composition is a method of removing the unnecessary frames in a video using algorithm. The videos are processed based on certain threshold values that are found using SIFT and SURF algorithm. The processed video is converted into frames and compared with the reference frames. This way face recognition is implemented for image retrieval and is useful in surveillance, crime detection, finding missing objects, etc,.

Keywords: Video composition, SIFT, SURF, threshold value, image retrieval.

1. Introduction

With the growth in technology, all personal devices are manufactured with digital cameras. In traditional camcorders, there were many sampling issues that were spotted. Digital cameras have many more added features that take clear videos thereby addressing the issues found in camcorders. Digital videos have many characteristics that give better clarity for the videos and images taken. In personal devices like mobile phones, we can take only videos of short duration. If a user wants to view all the relevant videos in a single shot it is not possible. This paper deals with the process of converting multiple videos into a single long-shot video which is content-consistent. Video browsing is done by mainly focussing on video synopsis or summarization. It retrieves only those frames that contain more relevant information thus reducing size of video to be browsed.

A long-shot video is nothing but a video with long duration that contains all important frames of the multiple input videos and is appealing to the user. Such long-shot videos are useful in various fields like surveillance, films, media etc,. The proposed technique called the "Seamless compositon" is used to generate an automatic long shot video. After the generation of a single-shot video, the next process is object level and human matching. This system focuses on automatically discovering content-consistent clips, merging the multiple videos into a single-shot video with spatial and temporal consistency. It focuses on simplifying the video browsing process by providing a tree structure collection with temporal smoothing. [1-5] In SIFT algorithm, we first find the key points in the objects from the given reference frame in the database. The new image is now compared with the reference image to check whether it matches the features. To find this matching we use Euclidian distance of the feature vectors. From this process we retrieve those images that match with the reference frame in terms of scale, orientation and location. The same procedure is followed for human matching. For finding the candidate key point, we use the difference-of-Gaussian scale-space function $D(x, y, \sigma)$. First we take the key point as the origin. We use Taylor series for finding the value of the key point. Taylor series is given by:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

Now D values are calculated at the keypoint and $\mathbf{x} = (x, y, \sigma)$ is the offset value. If this offset value is greater than 0.5, it considers the next point to be closer to the extremum and changes the candidate keypoint to the next point. Else, it is included as the candidate keypoint for feature matching.[6-9]

The matching process can be done more efficiently using SURF. SURF uses high performance descriptors. It first converts the image into pixel values using a method called multi-resolution. This technique will reduce the bandwidth of the original image thereby reducing the blurring effect. Thus scale-space is achieved. SURF uses a detector based on Hessian matrix which is of high accuracy to find the keypoints. By taking the determinant value of the matrix, we can get an expression that indicates the local change around the area. Given a point x=(x,y) in an image, the Hessian matrix H(x,σ) is calculated as follows:

$$H(x,\sigma) = \begin{pmatrix} Lxx(x,\sigma) & Lxy(x,\sigma) \\ Lxy(x,\sigma) & Lyy(x,\sigma) \end{pmatrix}$$

where σ is the scale value, $Lxx(x,\sigma)$ is the convolution value of second order derivative $\partial x/\partial x^2 g(\sigma)$. Similarly $Lxy(x,\sigma)$ and $Lyy(x,\sigma)$ values are calculated with their respective coordinates. The orientation of the selected keypoint defines the horizontality and verticality. The SURF descriptor is given as :

$$v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$$

This descriptor, v, is unique as well as robust to noise, geometric and photometric variations.

There are two ways of doing comparison between the descriptors. (1) find out the keypoints and descriptor values of both images. Now we can compare these values to find similarity. (2) find the keypoint and descriptor value of the first image and compare it pixel-By-Pixel With The Second Image.[10-16]

2. Related Work

In the VPDOS, since SIFT detects all the unnecessary points in a frame,[1] the computational cost is high and the order of complexity is O(N^3). It works only when the size of videos are large enough. Here in the proposed system we use SURF instead of SIFT which uses a descriptor to identify the important and necessary points in the frame thus reducing the complexity and cost.

In this proposed system, we use the idea in ACTVP [2], that is to align multiple videos into a single video. Here we aim to generate a smooth and consistent video whereas [2] aims to present all the video segments with music.

Here, we try to make the matching process faster by using SURF algorithm and thus overcoming the drawback in CTLV[4] which increases the exposure time of a frame thereby spoiling the picture and consumes lot of time.

The aim of creating content-consistent long-shot video came from the idea in [8]. In OAHVE [8], the video is processed and important parts are taken and merged together along with the songs that suit the videos. But this will not be continuous and will tend to be unrealistic. In the proposed system we focus on generating seamless video that contains important frames.

In the proposed system we perform face recognition based on the features obtained from the reference frame. If the features in reference frame match with the input frame then face is recognised. But in EP [10], when an input frame is given, the features are extracted and all possible expressions are detected and then the output is got. This increases the computational time.

In this paper we eliminate the unnecessary processing of unimportant frames which increases cost and time. In RIMFD [9], the frames with weak features are found out, Real Adaboost algorithm is applied to strengthen the weaker frames, thereby increasing the number of unnecessary computations.

3. Proposed Method

System Overview

The main objective is to convert multiple videos into a single-shot video. We further use this long-shot video for object-level and human matching. The process is divided into 4 stages: (1) Pre-processing (2) Categorisation based on transition hints (3). Video composition based on object matching (4). Video composition based on human matching. The overall system architecture is given below

4. Output

A. Pre-processing

Multiple videos are given as input. All the videos are converted into frames. They are then resized with corresponding resolution. They are then summarised and merged into a single long-shot video that contains frames which has the necessary and useful information. To generate the single-shot video we use hashing technique. Here video pair selection is done based on the similarity between the frames. The frames with highest similarity are considered as candidate video pair. Sequence matching method is used to generate continuous transition of frames. After the frames are generated, graph is constructed and the repeated frames are removed by edge pruning method.

B. Categorisation based on transition hints

Once the single-shot video is obtained after pre-processing, we use Viola-Jones algorithm to detect the face appearance of humans. Thus all the frames that contain humans are segregated as a separate video leaving behind the objects as another video. Thus, the long-shot video is categorised and we get human and object hints with which we will do the matching process in the later stages.

C. Video Composition based on object matching

After categorisation on single-shot video, we get all the object hints that appear in the frame as a video. Given a reference frame, matching is done between the test frame and the reference frame. This matching occurs irrespective of the background. We use SURF algorithm which first detects all the features of the reference frame and then compares it with the test frame. If the frame matches the reference frame, the image is retrieved. Thus, the object matching is done.

D. Video Composition based on human hints

Human video that is got after the categorisation of long-shot video. Now SURF algorithm is used to detect the features of the given reference frame. The test frame that contains human is taken and its features are found out using Eigen value method. Both the frames are compared. If the features of both frames are similar, then the image is retrieved.

5. Experiment Results

Face recognition is the main focus of this paper and it is done using SURF. The results of performing matching among frames are done using both SIFT and SURF and the difference in the parameters are shown in the below table.

Object Matching

Figure (a). Reference image

Figure (b). Object matching using SURF

Figure (c). Object matching using SIFT

HUMAN MATCHING

Figure (d). Reference image

Figure (e). Human matching using SURF

Figure (f). Human matching using SIFT

In the above results, fig.(a) is the input reference frame that is given for matching. Fig(b). Shows how the SURF descriptor takes only the required points in the frame and matches it with the input frame. Fig.(c) describes how SIFT takes all the points in the frame and processes unnecessary points in the frame.

The next figures depicts human matching. Fig(d) is the input human frame and fig(e) shows how SURF matches the required points in the reference frame with that of the input frame. Fig(f) shows how SIFT takes all points for matching thus increasing the cost for processing.

On comparing various parameters of SIFT and SURF, the following table is drawn:

IMAGES	SIFT	SURF
Detected points	1292	482
Average time	1646.53 ms	485.77 ms

This table clearly shows that with the use of SURF descriptor, we take only selected points for matching. The accessing time also shows that SURF is more efficient than SIFT and produces a better result in terms of time and cost.

	SIFT	SURF
Runtime	2887.13 ms	959.87 ms
Power consumption	0.04 %	0.02 %
Effectiveness	0.33	0.40

This table shows the comparison of SIFT and SURF in terms of runtime, power consumption and effectiveness. The runtime of SURF is less than half the time taken for SIFT. Also, the power consumption is less for SURF as it processes only the required points and finishes the matching quickly. As only selected points are taken and matched, the process is quicker and hence SURF will obviously be much efficient than SIFT.

6. Conclusion

Thus, small multiple videos are converted into frames which are resized according to specified resolution. These frames are converted into a single long-shot video. Now, we use Viola-Jones algorithm to distinguish human and object frames and convert them into separate videos. After the completion of this process we use SURF algorithm to detect the features of the frame and use these features for face recognition. Previously this was done by SIFT algorithm which detects all points in a frame including unwanted points which led to processing of unnecessary points which led to higher computational cost. In the proposed system, we use SURF algorithm which has a descriptor that identifies only the necessary points in a frame and the process takes place only for those points and thus complexity and computational cost is reduced thereby reducing the runtime.

References

[1] Udayakumar R., Kaliyamurthie K.P., Khanaa, Thooyamani K.P., Data mining a boon: Predictive system for university topper women in academia, World Applied Sciences Journal, v-29, i-14, pp-86-90, 2014.

[2] Kaliyamurthie K.P., Parameswari D., Udayakumar R., QOS aware privacy preserving location monitoring in wireless sensor network, Indian Journal of Science and Technology, v-6, i-SUPPL5, pp-4648-4652, 2013.

[3] Brintha Rajakumari S., Nalini C., An efficient cost model for data storage with horizontal layout in the cloud, Indian Journal of Science and Technology, v-7, i-, pp-45-46, 2014.

[4] Brintha Rajakumari S., Nalini C., An efficient data mining dataset preparation using aggregation in relational database, Indian Journal of Science and Technology, v-7, i-, pp-44-46, 2014.

[5] Khanna V., Mohanta K., Saravanan T., Recovery of link quality degradation in wireless mesh networks, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4837-4843, 2013.

[6] Khanaa V., Thooyamani K.P., Udayakumar R., A secure and efficient authentication system for distributed wireless sensor network, World Applied Sciences Journal, v-29, i-14, pp-304-308, 2014.

[7] Udayakumar R., Khanaa V., Saravanan T., Saritha G., Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction, Middle - East Journal of Scientific Research, v-16, i-12, pp-1781-1785, 2013.

[8] Khanaa V., Mohanta K., Saravanan. T., Performance analysis of FTTH using GEPON in direct and external modulation, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4848-4852, 2013.

[9] Kaliyamurthie K.P., Udayakumar R., Parameswari D., Mugunthan S.N., Highly secured online voting system over network, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4831-4836, 2013.

[10] Thooyamani K.P., Khanaa V., Udayakumar R., Efficiently measuring denial of service attacks using appropriate metrics, Middle - East Journal of Scientific Research, v-20, i-12, pp-2464-2470, 2014.

[11] R.Kalaiprasath, R.Elankavi, Dr.R.Udayakumar, Cloud Information Accountability (Cia) Framework Ensuring Accountability Of Data In Cloud And Security In End To End Process In Cloud Terminology, International Journal Of Civil Engineering And Technology (Ijciet) Volume 8, Issue 4, Pp. 376–385, April 2017.

[12] R.Elankavi, R.Kalaiprasath, Dr.R.Udayakumar, A fast clustering algorithm for high-dimensional data, International Journal Of Civil Engineering And Technology (Ijciet), Volume 8, Issue 5, Pp. 1220–1227, May 2017.

[13] R. Kalaiprasath, R. Elankavi and Dr. R. Udayakumar. Cloud. Security and Compliance - A Semantic Approach in End to End Security, International Journal Of Mechanical Engineering And Technology (Ijmet), Volume 8, Issue 5, pp-987-994, May 2017.

[14] Thooyamani K.P., Khanaa V., Udayakumar R., Virtual instrumentation based process of agriculture by automation, Middle - East Journal of Scientific Research, v-20, i-12, pp-2604-2612, 2014.

[15] Udayakumar R., Thooyamani K.P., Khanaa, Random projection based data perturbation using geometric transformation, World Applied Sciences Journal, v-29, i-14, pp-19-24, 2014.

[16] Udayakumar R., Thooyamani K.P., Khanaa, Deploying site-to-site VPN connectivity: MPLS Vs IPSec, World Applied Sciences Journal, v-29, i-14, pp-6-10, 2014.