SURVING SELF LOADING VIDEO COMPOSITION

T.Dhanabalan[1], S. Suresh[2]

[1]PG Scholar, [2]Assistant Professor Dept of IT Anna University Coimbatore(India)

ABSTRACT

In the present or recent times people want to collect their memorable moments with the help of digital devices like camera. Digital videos becoming grown and found anywhere. So camera plays a vital role in our day to day life. However editing and organizing videos remains difficult for people by different reasons. Also searching takes more time. So people need a better solution for video edition and video organization in an efficient way. This paper presents various techniques used for video edition and composition for grouping the required portion of the video which has taken from different places at different time. Video storage helps to secure videos keep on by users. So, proper administration control will be there to maintain a recognized users record and its personal information to keep is as privacy one.

I. INTRODUCTION

In this modern world, people would like to recollect their memorable moments with the help of digital devices like camera. So camera plays a vital role in our day to day life. The image resolution varies according to the capacity of the camera. When people want to collect their videos which are taken in different places at different time, they find difficult to search and edit it. Searching the required image from the collection of video would need more time. So people need a better solution for video editing and organizing the videos. Image processing is a technique used to process the image. Normally the image is in the form of the pixels or it is said to be in matrix format i.e. group of pixels from an image. This image processing technique analyses the image as one by one pixel and performs the matching operation or else extract the useful information from the image. Image processing accepts image or video as input and produces an image or video as an output with better quality and efficiency. It is helpful for the engineers and scientist who works on feature detection, noise reduction, image segmentation, frame splitting etc., Image processing normally refers to digital image processing but sometimes it includes

analog and optical images too. Image processing includes five groups: Visualization, Image Sharpening and Restoration, Image Retrieval, Measurement of pattern and Image Recognition. Visualization is used to observe the objects that are not visibe.

II. RELATED WORK

A literature survey, or literature review, is a proof essay of sorts. It is a study and review of relevant literature materials in relation to a topic a person have chosen. A literature review is a text written by someone to consider the critical points of current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. Literature reviews are secondary sources, and as such, do not report any new or original experimental work. Also, a literature review can be interpreted as a review of an abstract accomplishment.

Video editing is a quite complicated one for users and professional's .When users want to edit their videos they need software and it takes some time for editing. By using software there is some problem in frames splitting. So in video editing with intelligent interaction technique Ahanger et al (l998).It has proposed an intelligent interaction technique called silver interface. This helps users to solve their problems and to provide an efficient edited video.

The user collects the videos and arranged them in tree shaped udder and then edits the video by rearranging the branches of the tree. Silver interface provides different formats for editing and it calculates and manipulates the audio and video separately. This silver interface helps us to edit the video. Even though it is useful for editing it has some disadvantages. Sometimes it fails to show the former position to the user after some change is performed in size of the video. Because of this problem people need a better solution for editing.

It have proposed a special technique called drag and drop interface for arranging the still images from videos Barnes et al(2010). This is mainly useful to media field. So that, users can relate the moving objects with graphical objects on the screen and organize the video to create a still image. It consists of several preprocessing techniques like particle tracking, particle grouping etc., to convert the video to still images and still images to video Bhat et al (2004). There are several limitations in s project. One is it takes some time for preprocessing. If the video length is large it takes more time to preprocess it and it runs slowly according to the length of the project. Another drawback is moving a video back and forth along a single path is difficult. So we need a best solution to resolve this problem.

Matching the scene is based on surf features. Surf algorithm has several advantages when compared to sift algorithm. Scene matching algorithm plays an important role in realizing the operational purpose of cruise missiles. Cutting (2002) have proposed a coarse to fine partial matching to realize the surf feature points. First the surf feature points are extracted from the base image and the real time image and the matching of surf key point is performed to find a match location of the images. After finding the surf key points coarse to fine partial matching is done to match the images.

Coarse match is based on bidirectional nearest neighbor method and fine matching is based on RANSAC method and dominant line direction method. RANSAC method stands for random sample consensus and it is used to estimate the parameters and to eliminate the outliers. The main disadvantages of this project are sometimes it may produce wrong matching.

SURF algorithm plays an important role in image matching. When we want to add a quality to an image which is taken at different places at different time we need an algorithm called SURF algorithm. Surf is the accelerated version of SIFT algorithm and is mainly used for object recognition and object tracking Chiu et al (2004). SURF is three times faster than SIFT algorithm. Yuan et al has proposed the SIFT algorithm for image matching with the use of the KD-tree. KD-tree is a useful data structure for organizing points in a k dimensional space. First the feature points are extracted from the image by using SURF algorithm and the KD-tree algorithm is used to improve the efficiency and compared the SIFT algorithm with surf algorithm. But there is a problem with this project. i.e. we have to convert the input image into a gray image to avoid some problems in matching. Also making color images with matching efficiency is another major problem. These problems are referred from the Table 2.1.

It have proposed the process of capturing, editing and composing the video segments. These segments are composed using algebraic operations like union, intersection and concatenation. Apart from this technique many evaluation techniques are used to measure the performanceAhanger et al (1998). A literature review is a text written by someone to consider the critical points of current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic.

Literature reviews are secondary sources, and as such, do not report any new or original experimental work. Evaluation and analysis of the automatic analysis technique consists of temporal ordering thematic composition, thematic nearness composition and time limited composition. These are used to demonstrate the viability of automatically composing new video. The main aim of the project is to edit the video and to improve the quality of the video. The figure shows a composition and customization of video segments. These metrics are used to compare the quality of the newly composed video with the original video. The edition of the video is based on the customer needs Taylor et al (1995).

The main disadvantage is sometimes it may analyze a wrong value to the quality of the video. Because of these wrong values people find difficult to edit and compose the video. And these composition techniques fail to produce the seamless transition to video. Even though these techniques present an automated way for video edition sometimes it needs user help to organize the edition part.

III.SYSTEM ARCHITECTURE

The objective of this concept is to automatically compose descriptive long take video with content consistent shots retrieved from a video pool that convert into a single shot video which must be efficient.

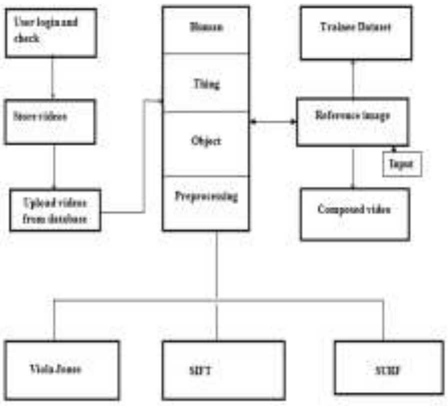

Figure 3.1 System Architecture

Initially a login page is created for the users to store the videos. After the storage, uploading the videos as an input then preprocessing technique will get started. In preprocessing the human and nonhuman objects are separated using the algorithm viola Jones, SIFT and SURF algorithm. Viola Jones is used for face detection and SIFT is used to describe and detect local features in images and SRUF algorithm is used for scene matching and to speed up the process. After preprocessing, a reference image is given as an input and using SURF algorithm it match the faces with the faces in the trainee dataset and then it automatically it compose the video which contain single person videos alone.

3.1 Algorithms

In this project three algorithms are used. They are

- SIFT
- SURF
- VIOLA JONES

3.1.1 Sift Algorithm

Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. The algorithm was published by David Lowe in 1999. Applications include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, individual identification of wildlife and match moving [9].

For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie in high-contrast regions of the image, such as object edges.

Another important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another. For example, if only the four corners of a door were used as features, they would work regardless of the door's position; but if points in the frame were also used, the recognition would fail if the door is opened or closed. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed. However, in practice SIFT detects and uses a much larger number of features from the images, which reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors.

SIFT descriptors robust to local affine distortion are then obtained by considering pixels around a radius of the key location, blurring and resampling of local image orientation planes [10].

Indexing consists of storing SIFT keys and identifying matching keys from the new image. Lowe used a modification of the k-d tree algorithm called the Best-bin-first search method that can identify the nearest neighbors with high probability using only a limited amount of computation. The BBF algorithm uses a modified search ordering for the k-d tree algorithm so that the bins in feature space are searched in the order of their closest distance from the query location. This search order requires the use of a heap-based priority queue for efficient determination of the search order. The best candidate match for each key point is found by identifying its nearest neighbor in the database of the key points from training images. The nearest neighbors are defined as the key points with minimum Euclidean distance from the given descriptor vector. The probability that a match is correct can be determined by taking the ratio of distance from the closest neighbor to the distance of the second closest.

Lowe rejected all matches in which the distance ratio is greater than 0.8, which eliminates 90% of the false matches while discarding less than 5% of the correct matches. To further improve the efficiency of the best-bin-first algorithm search was cut off after checking the first 200 nearest neighbor candidates. For a database of 100,000 key points, this provides a speedup over exact nearest neighbor search by about 2 orders of magnitude, yet results in less than a 5% loss in the number of correct matches.

Outlier can now be removed by checking for agreement between each image feature and the model, given the parameter solution. Given the linear least squares solution, each match is required to agree within half the error range that was used for the parameters in the Hough transform bins. As outliers are discarded, the linear least squares solution is re-solved with the remaining points, and the process iterated. If fewer than 3 points remain after discarding outliers, then the match is rejected. In addition, a top-down matching phase is used to add any further matches that agree with the projected model position, which may have been missed from the Hough transform bin due to the similarity transform approximation or other errors [11].

The final decision to accept or reject a model hypothesis is based on a detailed probabilistic model. This method first computes the expected number of false matches to the model pose, given the projected size of the model, the number of features within the region, and the accuracy of the fit. A Bayesian probability analysis then gives the probability that the object is present based on the actual number of matching features found. A model is accepted if the final probability for a correct interpretation is greater than 0.98. Lowe's SIFT based object recognition gives excellent results except under wide illumination variations and under non-rigid transformations.

3.1.2 Surf Algorithm

SURF (Speeded Up Robust Features) is a robust local feature detector, first presented by Herbert Bay et al. In 2006, that can be used in computer vision tasks like object recognition or 3D reconstruction. It is partly inspired by the SIFT descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images.

It uses an integer approximation to the determinant of Hessian blob detector, which can be computed extremely quickly with an integral image (3 integer operations). For features, it uses the sum of the Haar wavelet response around the point of interest. Again, these can be computed with the aid of the integral image. When we want to detect SURF features, we can use the syntax

POINTS = detectSURFFeatures(I)

POINTS = detectSURFFeatures (I, Name, value)

POINTS=detectSURFFeatures(I) returns a SURF Points object, POINTS containing information about SURF features detected in the 2-D grayscale input image I. The detectSURFFeatures function implements the Speeded-Up Robust Features (SURF) algorithm to find blob features.

POINTS = detectSURFFeatures (I, Name, Value) Additional control for the algorithm requires specification of parameters and corresponding values. An additional option is specified by one or more Name,Value pair arguments.

The task of finding point correspondences between two images of the same scene or object is an integral part of many machine vision or computer vision systems. The algorithm aims to find salient regions in images which can be found under a variety of image transformations. This allows it to form the basis of many vision based tasks; object recognition, video surveillance, medical imaging, augmented reality and image retrieval..

Feature detection is the process where we automatically examine an image to extract features that are unique to the objects in the image, in such a manner that we are able to detect an object based on its features in different images. This detection should ideally be possible when the image shows the object with different transformations, mainly scale and rotation, or when parts of the object are occluded. The processes can be divided into 3 overall steps.

- **Detection**Automatically identifies interesting features, interest points this must be done robustly. The same feature should always be detected regardless of viewpoint.
- **Description** Each interest point should have a unique description that does not depend on the features scale and rotation.

- **Matching** Given and input image, determine which objects it contains, and possibly a transformation of the object, based on predetermined interest points.

In order to detect feature points in a scale invariant manner SIFT uses a cascading filtering approach, Where the Difference of Gaussians, DoG, is calculated on progressively downscaled images. In general the technique to achieve scale invariance is to examine the image at different scales, scale space, using Gaussian kernels. Both SIFT and SURF divides the scale space into levels and octaves. An octave corresponds to a doubling of, and the octave is divided into uniformly spaced levels to detect the features of the object.

3.1.3 Viola Jones Algorithm

The Viola–Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection.

The basic principle of the Viola-Jones algorithm is to scan a sub-window capable of detecting faces across a given input image. The standard image processing approach would be to rescale the input image to different sizes and then run the fixed size detector through these images. This approach turns out to be rather time consuming due to the calculation of the different size images [12]. Contrary to the standard approach Viola-Jones rescale the detector instead of the input image and run the detector many times through the image – each time with a different size. At first one might suspect both approaches to be equally time consuming, but Viola-Jones has devised a scale invariant detector that requires the same number of calculations whatever the size. This detector is constructed using a so-called integral image and some simple rectangular features reminiscent of Haar wavelets.

The first step of the Viola-Jones face detection algorithm is to turn the input image into an integral image. This is done by making each pixel equal to the entire sum of all pixels above and to the left of the concerned pixel.

This allows for the calculation of the sum of all pixels inside any given rectangle using only four values. These values are the pixels in the integral image that coincide with the corners of the rectangle in the input image.

AdaBoost is a machine learning boosting algorithm capable of constructing a strong classifier through a weighted combination of weak classifiers. (A weak classifier classifies correctly in only a little bit more than half the cases.) To match this terminology to the presented theory each feature is considered to be a potential weak classifier. Since only a small amount of the possible 160.000 feature values is expected to be potential weak classifiers the AdaBoost algorithm is modified to select only the best features.

An important part of the modified AdaBoost algorithm is the determination of the best feature, polarity and threshold. There seems to be no smart solution to this problem and Viola-Jones suggests a simple brute force method. This means that the determination of each new weak classifier involves evaluating each feature on all the training examples in order to find the best performing feature. This is expected to be the most time consuming part of the training procedure. The best performing feature is chosen based on the weighted error it produces. This weighted error is a function of the weights belonging to the training examples.

The basic principle of the Viola-Jones face detection algorithm is to scan the detector many times through the same image – each time with a new size. Even if an image should contain one or more faces it is obvious that an excessive large amount of the evaluated sub-windows would still be negative (non-faces). This realization leads to a different formulation of the problem: Instead of finding faces, the algorithm should discard non-faces. The thought behind this statement is that it is faster to discard a non-face than to find a face. With this in mind a detector consisting of only one (strong) classifier suddenly seems inefficient since the evaluation time is constant no matter the input. Hence the need for a cascaded classifier arises. The cascaded classifier is composed of stages each containing a strong classifier. The job of each stage is to determine whether a given sub-window is definitely not a face or maybe a face. When a sub-window is classified to be a non-face by a given stage it is immediately discarded. Conversely a sub-window classified as a maybe-face is passed on to the next stage in the cascade. It follows that the more stages a given sub-window passes, the higher the chance the sub-window actually contains a face.

In the first algorithm two detections are merged if they have equal size and they overlap with 25 % or more. In the second algorithm two detections are merged if their centers coincide. As long as the detector is not yet done this merging affects the performance figures in a negative direction since the amount of visible true positives is more heavily reduced than the amount of visible false positives.

3.2 Methodology

The steps in this project are

- User Authentication & Video Storage
- Pre- Processing
- Categorization Based on Transition Clues
- Video Composition based on Reference image

3.2.1 User Authentication & Video Storage

Video Storage helps to secure videos keep on by users. So, proper administration control will be there to maintain a recognized users record and its information to keep is as privacy one.

3.2.2 Pre- Processing

First Our Input short videos are converted into frames. . Then we eliminate some frames like information less frames (Mean of Input frame<15). After we resize the each frame. Then all frames are merged into a single video for video categorization.

3.2.3 Categorization Based on Transition Clues

Videos are categorized by using transition clues like human, object. Then we are taking human clue for first categorization by using Viola-Jones algorithm, if faces are not detected in frames that frames are separated into another process for object matching. Viola- Jones algorithm are specially used for face detection and before using this algorithm some training had to made for easy face detection. So separation of human and non human is comes under the preprocessing technique.

3.2.4 Video Composition Based on Reference Image

Object & sequence matching process are done by using SIFT algorithm (Scale-invariant feature transform). Related Object frames and related sequence frames are categorized into a separate folder respectively. Also surf algorithm is used for speed and good quality. SURF stands for speed up robust features. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images. . Finally categorized frames are converted into Separate videos.

IV. CONCLUSION

Automatic content based video composition describes the composition of single shot video.People may want to collect their memorable moments for their happiness and to show their wealth. So they would collect their videos and photos which are taken at different places and in different time. This process requires an efficient algorithm for video collection and composition.

By considering this problem, an innovative solution has been proposed to help people to collect their videos and to produce a single-shot video which contain the requested person videos alone. Our main aim is to automatically retrieve videos from the video pool and pre-processing is done to separate the human and the non-human objects and recognize the face of a person to produce a single shot-video. A pre-processing technique has been used to separate the required portion from video. It comprises three algorithms namely SIFT, SURF and Viola – Jones

Login page was created to provide security for the users. Then the videos would be uploaded for pre-processing. In pre-processing the human and non-human objects would be separated. Thus separation of human video and object video would have been takes place. During this stage the sift algorithm would be used for the identification of objects and viola jones for detection of human faces. Back propagation algorithm is used to learn different features in each face so that identification of the face can be done easily. Then the reference image would have been as an input image. It goes to the trainee dataset to refer whether the images are in the trainee dataset. If it is available their then detection of face will be easy else the reference image is not related to the video. Trainee dataset contain the all the photos which contained in the vide

V. FUTURE ENHANGEMENT

When users want to collect their videos, which are taken in different places at different time, find difficult to edit and organize the videos. In some projects coarse to fine partial matching is used to match the human faces. But sometimes it may produce the wrong output. Also it takes more time to detect the human face and to match it. So to avoid these problems three algorithms are proposed in this project to increase the speed of the process and for perfect matching. The main aim of the project is to collect the short videos, preprocessing it and to produce a long shot video which contains individual person videos alone. So it would be easy for users to collect their videos from many short videos. Also this can be used by any user at a time. i.e many number of users can use it at a time. User login also added to provide security for the users to keep their videos secretly.

REFERENCES

[1]. AhangerG., “Automatic composition techniques for video production,” IEEE Trans. Knowl. Data Eng., vol. 10, no. 6, pp. 967–987, Nov. 1998.

[2]. AxelrodmA., CaspiY., A. Gamliel and Y. Matsushita, “Dynamic stills and clip trailers,” Visual Comput., vol. 22, no. 9, pp. 642–652, Sep. 2006.

[3]. Barnes C, Goldman D, Shechtman E, and Finkelstein A, “Video tapestries with continuous temporal zoom,” in Proc. SIGGRAPH, 2010.

[4]. Bennett E, “Computational time-lapse video,” ACM Trans. Graph., vol. 26, no. 102, Jul. 2007.

[5]. Bhat K S, Hodgins J, Khosla P, Seitz S, “Flow-based video synthesis and editing,” ACM Trans. Graph., vol. 23, no. 3, pp. 360–363, Aug. 2004.

[6]. Chiu P, Girgensohn A, and Liu Q, “Stained-glass visualization for highly condensed video summaries,” in Proc. ICME, 2004.

[7]. Cootes T, Taylor C, and Cooper D, “Active shape models-their training and application,” Comput. Vision Image Understand., vol. 61, no. 1, pp. 38–59, Jan. 1995.

[8]. Cutting J.E, “Representing motion in a static image: Constraints and parallels in art, science, and popular culture,” Perception, 2002.

[9]. Everingham M, Gool L.V, Williams, J. Winn, and Zisserman A, “The Pascal visual object classes (VOC) challenge,” Int. J. Comput. Vision, vol. 88, pp. 303–338, 2010.

[10]. NikolaidisC. C, “Video shot detection and condensed representation. A review,” IEEE Signal Process. Mag., vol. 23, no. 2, pp. 28–37, Mar. 2006.

[11]. PhilbinO. C, “Near duplicate image detection: min-hash and tf-idf weighting,” in Proc. BMVC, 2008.