MULTICAMERA VIDEO STITCHING SURVEILLANCE SYSTEM

Ahsan Ali1, Asghar Ali Shah*2, Kashif Aftab*3, Muhammad Usman*4
Department of Computer Science, Bahria University, Lahore, Pakistan
ahsanali3812@gmail.com

ABSTRACT: *CCTV cameras are commonly used for security issues. Pan-tilt-zoom (PTZ) cameras are mostly used for this purpose. To stitch two or more video streams from different cameras is much cheaper than PTZ solution. There are three stages of video stitching. Feature identification is the first stage of video stitching. To scale the invariant features like rotation, scaling and noise etc. Direct and feature base identification has basically two types of feature identification. Shifting and warping the images purpose to identify how these features are agreeing with each other is the main concern for direct base identification. While feature identification rely on extracting the features and then perform matching among them on the base of features. Calibration is the second stage of the video stitching. The images are stitch in panoramic way in calibration depending upon alignment among them. Blending is the last stage of video stitching where numerous videos are display in single panoramic way. Any blending algorithm is used to blend the pixels together and for final view.*
Keywords: Video stitching, Surveillance system, Computer vision, Image processing.

1. **Introduction.** This template, modified in MS Word 2007 and saved as a "Word 97-2003 Document" for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers. (2) Automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products. (3) Conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in, examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multileveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

As there are many kinds of processing like image processing, signal processing etc. in this process some input is given to process the image. There are different types of image processing which digital and analog image are processing. By using the electrical source alteration of images is the concern of analog image processing, like television image. The digital computers are used in digital image processing purpose to process the images. Image processing plays a very important role in video stitching perspectives. Video stitching process totally depends upon the image processing. Video stitching is process to combine multiple videos to a single view. Today there are lot of cameras place at every point but the main problem is that we need to observe each camera. So, it is lot of time consuming. There is a solution that all videos can combine in single view. Video stitching give the best option to see all camera result in a single view. If we look all cameras result then there is chance that some space remains in another camera. So due to this reason we cannot see where problem exist. So, video stitching covers all areas and gives a full panoramic view. CCTV cameras are used for security issues everywhere. Demand of the CCTV cameras is rapidly increasing in security and monitoring perspectives. Pan-tilt- zoom (PTZ) cameras are used mostly for its better quality. Market deserve the cheaper solution rather than PTZ cameras. PTZ camera solution is costly way and especially its monitoring software's are required to buy [1]. To stitch two or more cameras. Video is much better Solution to review the market demand and cheaper solution the video stitching technique is proposed for CCTV cameras.

There are many techniques used for the video stitching. In this paper we will discuss the pixel-based identification and temporal synchronization. The different algorithms used to compare the transformations. The section 2 consist previous work, section 3 defines the process of video stitching. A desktop application will be

developed which will combine multiples videos in a single view for resolve after meeting all the stages of video stitching. Following are the main purpose of application.

- 360-degree view of whole room.
- Panoramic view of image
- Total view of all images that captured by camera.

Video stitching system is the system which detect point where problem exist. Video stitching system is used for controlling and monitoring the system in security places.

2. Literature Review. It has been a hot topic since last decade [2]. Real time preview is also created [3]. Panoramic video of accepted standard at real time is also created [4] but it can combine only two cameras. Three video streams are also combined [5]. Four video streams are also combined [6] but it does not focus on visual quality. Real time panorama video with good performance is also proposed [7] but it only works with two cameras and low-resolution image. Good visual quality and high-resolution image is also proposed [8, 9] but it requires expensive and specialized hardware. The previous work has various issues such as Low visual quality [6]. Require expensive and specialized hardware [8, 9]. Stitches only a few videos [4]. Other issues such as Monitoring blind areas. Removing artifacts (ghosting like effects). Video stitching related with all the processes. The last step after blending is to monitor and assistance of the panoramic video. Image stitching mean combine two or more images to form a single view. In [10] there are main steps of stitching the images into single view which are image calibration, image registration and image blending. The purpose of calibration is that it minimizes difference between ideal lens and camera lens. The difference between images can be resulted by optical defects such as distortion and exposure. In [11] there are two type of camera parameter which is extrinsic parameter and intrinsic parameter. In [12] image registration discussed as alignment of images taken by multiple points. Image registration focuses to create geometric representation among the images.

In [11] image blending concern stitch images to form a single view. There are two ways of blending which are alpha feathering and Gaussian pyramid. Alpha feathering works well when images pixel are well aligned and intensity represent the difference between two images. Gaussian pyramid work on the frequency of image it merges the image on the basis of frequency and then filter them accordingly. There are two approaches of image stitching. Direct technique basically compares the pixel intensities of images. It minimizes the overlapping difference between images. The actual benefit of direct technique is that it provides more useful detail for image alignment. There is also disadvantage that there is limited coverage range. In feature base technique feature point in one image and other image which need to compare can be find by local descriptor. Feature base method start by establishing a corresponding connection between points, lines, edge and other geometric entities.

A. Feature detection and description

There are two types of feature descriptor. One is vector descriptor and second is binary descriptor. ORB, BRIEF are binary descriptors while Scale Invariant Feature Transform and Speed up Robust Feature are direction descriptors. Further concern study is about feature detectors.

a) SIFT : SIFT (scale invariant features transform) extracts the invariant features and partially illuminate the changes like scaling and rotation. A descriptor also called vector (dimension 128) is related with every feature. The feature matching among the image pairs is allowed by the descriptors (vector) to satisfy the geometric transformations. The piece filtering approach is used for feature detection. The imaginary and tentative outcomes of line berry show that Gaussian function that provides possible weighbridge of different scale image space. It is the new method for scale-invariant key point's detection. As the Lundeberg algorithms motivations that key points are extract by SIFT as the minimum difference of Gaussian method, that could be calculated by differentiation one or more neighbors gages that are disconnected to persistent multiplication element k. the computation cost is the main advantage of the LOW'S detector but it reduced when compare with this method. By using the SIFT the matching between key points is extracted. Describing indigenous appearance area nearby basic ideas is goal of this matching. SIFT descriptor links to the distribution base descriptor (DBD) class. The histograms use to represent these techniques and its appearance. The distribution of pixel intensities represented by histogram is the

concern of simple descriptor. The robustness against the shifts in geometry is the great advantage of these operators.

b) SURF : It is the extension of SIFT based on multi-scale space theory but work with different method to extract the features. The fast approximation of HESSIAN matrix speeds up its commutations.

c) FAST: It is applicable feature detector for actual applications. It is known as high speeded detector. The circle of 16 pixels measured in this algorithm that are around the candidate corner p. It often happen when there is pixels that involved in circle are optimistic [13].

d) HARRIS: Suggested by Harris and Stephens [14]. It is junction indicator based on moravee method. Fluctuating the frame in a bit in multiple track is resulted to determining the variation in intensity. The window center point is extracted as corner point.

e) ORB: Ruble develop the ORB technique et al [15], which is combination of FAST and BRIEF descriptors that describe the input image features in binary string rather than vector [16].

f) OPENCV: Source library for computer vision and image processing that contain many inbuilt functions for real-time image processing [17]. Video stitching is post processing of image stitching. Image stitching consists of image acquisition that has three processes the 1. Geometric transformation, 2. Visual quality alignment and 3. Temporal synchronization. Video stitching surveillance system consists of real time video stitching in which more than two or more cameras are combine in single panoramic view. Video stitching solution is much cheaper than PTZ camera solution [15]. Post processing like tracking and panning used to improve the cinematic color, contrast and white balance. Such algorithms that provide best quality are preferred to increase the video quality.

B. Preprocessing

Video stitching pre-processing consists of transformation of images. Transformations focus on rotation and scaling of the images. The next is to improve the visual quality, for this purpose many algorithms are available in computer vision. After the temporal, synchronization is necessary that focus on frame to frame adjustments of the images. So, the mainly preprocessing has three steps.

     a)   Geometric transformation
     b)   Visual quality Alignment
     c)   Temporal synchronization

a) Geometric Transformation : The image processing is the preprocessing for video stitching. The homography helps to detect the overlapping pairs of images. The optimal values of the elements in the 2D homography are estimated by minimizing the total transformation error [thesis]. The homography approximation method proposed for the proper image transformation [18]. It converts one plane to another based on feature points matching intended by match pairs. SIFT [19], SURF [20], SFOP [21], EBR [22] commonly recycled feature finders. The surf is the best algorithm as compared with the other algorithms due to its best performance. After feature point extraction and approximation matching is needed for this purpose the matching algorithms used.

b) Visual Quality Alignment : The visual quality in video stitching is achieved by the better alignment of the images. Essentially it is programmatic technique in which we adjust bundle. Resolution of this phase is to reduce miss-registration between all pairs of images. Least-squares algorithm [12] is used to get the optimal value of 3D reconstruction of image scene and camera position. Once completed the global alignment then there may be some errors due to local miss-registration like double images, images may be blur, so for these problems there is need to perform local adjustment like parallax removal. Different layouts use on which image stitching use identical rectilinear projection. On this layout stitched image can be viewed on 2D dimensional level overlapping the panosphere in a single point. Different lines of images show their direction on 2D plane. The disparity calculation algorithm is used to measure the disparity among the images and better for visual quality alignment.

c) Temporal synchronization: Temporal synchronization is achieved by recovering temporal alignment. It determines the sequence of the frames which first frame is given to the next corresponding frame. The cameras should be rigidly joined together and there should be non-uniform motion between the cameras. This process consists to assess the errors. For cylindrical panorama images captured with camera mounted on level of tripod. Each image can be warped on their cylindrical coordinates if the camera focal length known. There are two types of cylinder-shaped changing

- Forwarding changing
- Inverse changing

In forwarding changing image source is mapped onto the cylindrical apparent but there is problem that it has hovels in the endpoint of images since some pixels not mapped on the cylindrical surface. So that reason we use inverse warping because in this warping each pixel in endpoint image clearly mapped onto the basis image. The following bilinear exclamation secondhand to evaluate colors of pixels at end point [23]. Video stitching is the post processing of image stitching. We combine the videos taken by multiple cameras in single panorama shape [24]. The overlapping regions are handled and the viewers can view the panorama video. The video stitching Process consist the three stages

3. Proposed Model. The image stitching process is consisting of four stages. The user will select input overlapping images first, which are to be stitched. Image merging and registration will provide the final results. Image registration consists of feature matching which is done by feature descriptor technique and creating feature detector and. combination of key point dataset created by SURF, SIFT, MSER feature detector algorithms will included in Feature key points. The proposed technique below is a panorama image stitching system that focusing on image registration part. We match the features of images in image registration part by using some feature detection technique and output will be generated. The image registration's output will provide you higher quality of image stitching i.e. your final image. Feature matching provide key point detectors and key point descriptors. We can transform or overlay 2 images to efficient manner Based on large number of matches. Feature abstraction and equivalent are Entire process of stitching. So, feature abstraction part of image stitching technique is the main focus in our study. Our system involves four stages.
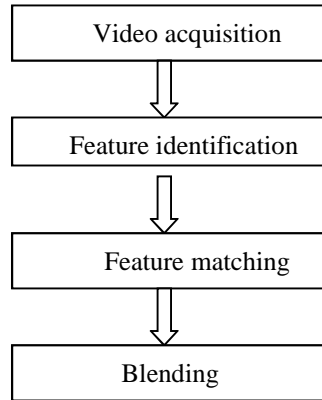

Fig 1. Proposed Video Stitching Model

Table 1. Process and Technique

Process Name	Technique	Best Technique
Video Acquisition	Physically Adjustment of cameras	Adjustment of cameras accordingly camera's FOV (field of view)

Feature Identification	1.SIFT (Scale Invariant Feature Technique) 2. SURF (Speed up Robust Feature) 3.ORIENTED FAST (feature from accelerated segment test) 4.ROTATED BRIEF (binary robust independent elementary feature) 5.ORB (oriented fast and rotated brief)	ORB detector (oriented fast and rotated brief)
Feature matching	1.HARRIS CORNER PAIR algorithm 2.RANSAC (random sample consensus) 3.NORM-HAMMING ALGORITHM 4.NORM-HAMMING ALGORITHM2	RANSAC (random sample consensus)
Blending	1.Alpha Feathering, Gaussian pyramid 2.Multi band blending technique	Multi band blending technique

In this section we will identify audiovisual by physically adjustment of the CCTV cameras. Here camera station, angle or field of view is the highly concentrated. No overlapping regions in cameras field of view. There should not be any area that uncovered by cameras FOV. Following are different category for image pre-processing. These methods are used for new pixel intensity calculation.

- Pixel intensity transformation
- Linear transformation
- Preprocessing method that use a local neighborhood of processed pixel

The video acquisition is process of getting the videos from multiple cameras and performs actions to next. In this section there are preprocess that are needed to perform for further process.

A. Feature identification

It is used to find interesting features or interesting points in the overlapping part of the image. To recognize the features of the images is the first stage where we citation the features of the images and match the features by using matching algorithms. Following are feature indicator technique like Harris, Scale Invariant Feature Transformation (SIFT) [25], Speed up Robust Feature (SURF) and ORB [26], feature from accelerated segment (FAST). So additional algorithm which is SURF try to improve calculation stretch of Scale Invariant Feature Transformation via use about uncomplicated image to profligate limited ascent computation upon image. Harris corner detector most useful because it detects the interest point (common point) between images. These points have large intensity. When point detect then there is need to match them.

ORB: It is the feature detector that is the combination of basically two detectors

a)   Oriented fast

b)   Rotated brief

ORB detector has the object with function that have number of Optional Parameters. It selects two points at a time and identify its features.

B. Feature Matching After identify the features the next stage is matching. Here points are match by normalized cross correlation which generate pairs. After that these pair need to be match if anyone is wrong or not. In [9] there is also Harris corner algorithm in which it applies multiple constraints to remove incorrectly match pair of images. Basically, this algorithm work on frame by frame because it need to identify which image is on left or right side. In [15] there is solution of Harris corner that instead of using frame by frame stitching there is need to find the interest point of previous of image by using optical flow and area of overlapping. They found that there is 85% time consume in matching interest point of images. By using OPENCV after completing the feature detection then matching algorithm use for matching. The NORM-HAMMING distance algorithm use for matching which takes 3 or 4 points to produce brief descriptor then NORM-HAMMING2 distance is used.

C. Blending: Feature image blending technique is basically used in computer graphics. It is simple method in which pixel value in merged regions have average value over weighted overlapping regions. Due to presence of exposure this technique does not work well. Another approach is multi-band image blending. In this technique image gradient form by each source is copied and matches to another. If match then these gradients are reconstructed. Approach image pyramid is actually demonstration of image .it mean hierarchical representation of images at different resolution. It provide useful properties for different application like noise reduction, image exploration and image development etc. it is the last stage the viewer can enjoy the panoramic video of the multiple cameras.

4. Discussion. There are different techniques used in video stitching process. The best technique in process is Speed up Robust Feature (SURF). Because this algorithm is best as compare to other algorithm due to its best performance. All other algorithm is also best but best is only speed up robust feature. In RANSAC basically find picture to its neighbor picture which match it. In feature identification interesting point is overlapping part of identified image. Next phase is feature matching where points are matched by normalized cross correlation which generate pairs. In blending process alpha featuring is Gaussian pyramid techniques are used. Basically, blending technique applied on image so that all image stitch should be seamless.]

5. Conclusion. The video stitching is today's burning topic. There are many algorithms used for video stitching. The image stitching is the preprocessing while the video stitching is the post processing. Preprocessing consist geometric transformation, visual quality Alignments and temporal synchronization. Geometric transformation performs by homography matrix, visual quality is achieved by the better alignment of images and temporal synchronization is achieved by the proper bundle adjustment. Feature extraction is the identification of features between the images and then matching with each other by using matching algorithms. Blending is the last stage of the video stitching where stitched video is displayed on the screen without any artifacts.

REFERENCES

[1] "Marquard, Stephen"; "Matterhorn 2014 Unconference: Ideas for automated post-recording video Handling," 2014. 19 March 2014.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Brown, M., & Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. International journal of computer vision, 74(1), 59-73.

[3] Baudisch, P., Tan, D., Steedly, D., Rudolph, E., Uyttendaele, M., Pal, C., & Szeliski, R. (2005). Panoramic viewfinder: Shooting panoramic pictures with the help of a real-time preview. In Proc. UIST.

[4] Abood, S. S., & Karim, A. A. (2014). Fast generation of video panorama. terminology, 3(6).

[5] Jiang, W., & Gu, J. (2015). Video stitching with spatial-temporal content-preserving warping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 42-48).

[6] Tang, W. K., Wong, T. T., & Heng, P. A. (2005). A system for real-time panorama generation and display in tele-immersive applications. IEEE Transactions on Multimedia, 7(2), 280-292.

[7]   Adam, M., Jung, C., Roth, S., & Brunnett, G. (2009, November). Real-time Stereo-Image Stitching using GPU-based Belief Propagation. In VMV (pp. 215-224).

[8]   Kulkarni, A. V., Jagtap, J. S., & Harpale, V. K. (2013). Object recognition with ORB and its Implementation on FPGA. International Journal of Advanced Computer Research, 3(3), 164.

[9]   Kaiser, R., Thaler, M., Kriechbaum, A., Fassold, H., Bailer, W., & Rosner, J. (2011, November). Real-time person tracking in high-resolution panoramic video for automated broadcast production. In Visual Media Production (CVMP), 2011 Conference for (pp. 21-29). IEEE.

[10]   "Shashank, K., Sivachaitanya, N., Mainkanta, G., Balaji, C., & Murthy, V. V. S. A. (2014). A survey and review over image alignment and stitching methods. International Journal of Electronics & Communication technology, 5, 50-52.

[11]   Rosten, E., & Drummond, T. (2006, May). Machine learning for high-speed corner detection. In European conference on computer vision (pp. 430-443). Springer, Berlin, Heidelberg.

[12]   Patel, U. B., & Mewada, H. N. (2013). Review of image mosaic based on feature technique. International Journal of Engineering and Innovative Technology, 2(11).

[13]   Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence, 22.

[14]   Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In Computer Vision (ICCV), 2011 IEEE international conference on (pp. 2564-2571). IEEE.

[15]   Agarwal, A., Jawahar, C. V., & Narayanan, P. J. (2005). A survey of planar homography estimation techniques. Centre for Visual Information Technology, Tech. Rep. IIIT/TR/2005/12.

[16]   Karthik, R., AnnisFathima, A., & Vaidehi, V. (2013, July). Panoramic view creation using invariant momentsand SURF features. In Recent Trends in Information Technology (ICRTIT), 2013 International Conference on (pp. 376-382). IEEE.

[17]   Deng, Y., & Zhang, T. (2003, November). Generating panorama photos. In Internet Multimedia Management Systems IV (Vol. 5242, pp. 270-280). International Society for Optics and Photonics.

[18]   Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In Computer Vision (ICCV), 2011 IEEE international conference on (pp. 2564-2571). IEEE.

[19]   Brown, M., & Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. International journal of computer vision, 74(1), 59-73.

[20]   Tennoe, M., Helgedagsrud, E., Næss, M., Alstad, H. K., Stensland, H. K., Gaddam, V. R., ... & Halvorsen, P. (2013, December). Efficient implementation and processing of a real-time panorama video pipeline. In Multimedia (ISM), 2013 IEEE International Symposium on (pp. 76-83). IEEE.

[21]   Kim, C. W., Sung, H. S., & Yoon, H. J. (2015). Camera image stitching using feature points and geometrical features. International Journal of Multimedia and Ubiquitous Engineering, 10(12), 289-296.

[22]   El-Saban, M., Izz, M., & Kaheel, A. (2010, September). Fast stitching of videos captured from freely moving devices by exploiting temporal redundancy. In Image Processing (ICIP), 2010 17th IEEE International Conference on (pp. 1193-1196). IEEE.

[23]   Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381-395.

[24]   Karthik, R., AnnisFathima, A., & Vaidehi, V. (2013, July). Panoramic view creation using invariant momentsand SURF features. In Recent Trends in Information Technology (ICRTIT), 2013 International Conference on (pp. 376-382). IEEE.

[25]   El-Saban, M., Izz, M., & Kaheel, A. (2010, September). Fast stitching of videos captured from freely moving devices by exploiting temporal redundancy. In Image Processing (ICIP), 2010 17th IEEE International Conference on (pp. 1193-1196). IEEE.

[26]   Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., ... & Van Gool, L. (2005). A comparison of affine region detectors. International journal of computer vision, 65(1-2), 43-72.