Image stitching and 2D to 3D Image Reconstruction for Abnormal Activity Detection

J.L. Mazher Iqbal, PhD
Professor,
Department of Electronics and Communication Engineering,
Madanapalle Institute of Technology & Science,
Angallu, Chittoor Dist, A.P, India.

M. Suriya Parveen
M.Tech Student
Department of Electronics and Communication Engineering,
Madanapalle Institute of Technology & Science,
Angallu, Chittoor Dist, A.P, India.

S. Arun
Assistant Professor
Department of Electronics and Communication Engineering,
Madanapalle Institute of Technology & Science,
Angallu, Chittoor Dist, A.P, India.

ABSTRACT
The basic idea of stirring the visual information is required to obtain 3D image. Over the time, various techniques have been evolved to enhance the visual information. There are several techniques for 2D to 3D conversion but it aim at creating a depth of vision using two images. The proposed method used multilayer information to get 3D information from 2D. The first step in the proposed work is to capture the video using web cam and then divide the captured information into frames and the images are registered. Features are extracted from the registered image such as edges and boundaries using scale invariant feature transform (SIFT). A series of images captured from different cameras are stitched by a geom etrically consistent mosaic either horizontally/vertically based on the image acquisition. Anaglyph method is applied to the stitched image for 3D reconstruction. In the proposed approach, the pictures taken from multiple viewpoints of the same scene are stitched and convert into 3D image from 2D, so that more informative representation of the scene is avail able for abnormal activity detection.

Keywords
Video capture, Image Acquisition, image registration, Stitching (SIFT), 3D Reconstruction.

1. INTRODUCTION
Visual surveillance is an active research topic in computer vision that tries to detect, recognize and track objects over a sequence of images and it also makes an attempt to un derstand and describe object behavior by replacing the aging old traditional method of monitoring cameras by human operators. A computer vision system, can monitor both immediate unauthorized behavior and long term suspicious behavior, and hence alerts the human operator for deeper investigation of the event. The video surveillance system can be manual, semi-automatic, or fully-automatic depending on the human intervention. In manual video surveillance system, human operator responsible for monitoring does the entire tas k while watching the visual information coming from the dif ferent camveras. It's a tedious and arduous job of an operator to watch the multiple screens and at the same time to be vigi lant from any unfortunate event. These systems are proving to be ineffective for busy large places as the number of cameras exceeds the capability of human experts. In the fully-autonom ous system there is no human intervention and the entire job is being done by the computer vision. These systems are intellig ent enough to track, classify, and identify the object. In addition, it reports and detects the suspicious behavior and d oes the activity recognition of the object. This paper proposes image stitching and 2D to 3D image reconstruction for abnor mal activity detection. Image stitching combines multiple pictures. The camera's field of view is always smaller than the human field of view, hence there is a need to combine pictures to get broader view of the picture so that the suspicious activities can be detected efficiently. The proposed system uses 3D reconstruction method by which an object is obtained within a virtual, three-dimensional (3D) space. This can be do ne in a number of ways, but usually involves the use of two-dimensional (2D) photographs or the scanning of the actual object as input data. Once this 3D reconstruction is created, then it can be manipulated or utilized for different application such as medical uses, law enforcement reconstructions, and even the creation of 3D graphics for film or television. The proposed system can be used efficiently in many applications for video surveillance applications such as security, traffic surveillance, crowd flux statistics and congestion analysis, person identification, detection of anomalous behavior, etc. Several literature paper are surveyed to develop the proposed system. Paper [1], discuss the different set-up's in the acquisition of images for generating panoramic images, the sti tching of the acquired images and the resultant panoramic im ages produced by stitcher. The method used is image registr ation and image merging. Paper [2] discuss the design a high performance software to perform geometrical and radio met rical corrections of two or more images without using any kind of special hardware. The work is on retina. In paper [3] a method is proposed for generating mosaic image using planar, cylindrical and general manifold. Manifold projection helps in the fast creation of low distorted panorama mosaics under very slight camera motions. In [4] pipe projection model is proposed. The model helps in removing parallax during co mplex motion. The proposed novel system performs image stitching, feature extraction using SIFT and 2D to 3D image reconstruction for abnormal activity detection.

2. PROPOSED METHOD
The proposed image stitching and 2D to 3D Image Recons truction system for Abnormal Activity Detection is shown in figure 1. The various blocks of the proposed is system is discussed below:

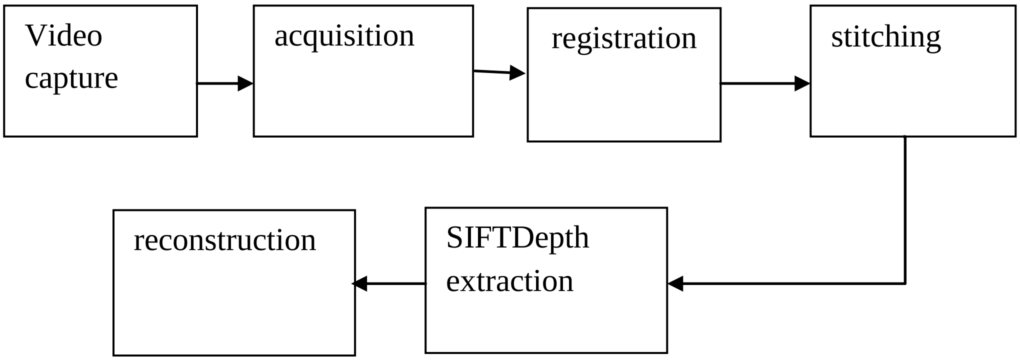

Fig. 1 Block diagram of proposed system

2.1. Image Acquisition

Image is capture using two webcam mounted on a tripod stan d. The camera angle is varied to take different overlapping sa mple images.

2.2. Image Registration

Image registration is a most important part of image mosaicing and 3D reconstruction. In accordance with certain i mage transform parameters such as edges and angles are generated by similarity measurement. Registration of multi-source images collects two or more images which are focused on the same target but produced from different perspectives, and different times. This process makes the multi-source imag es transformed into a same coordinate system and achieves the best match on pixel level. Using two two-dimensional arra ys,I_1(x,y) and I_2(x,y), to represent the reference image and the image to be matched, image registration can be stated as follows:

$$I_1(x,y)=g(I_2(f(x,y))) \quad (1)$$

f is a transform function for two-dimensional coordinates, g is global transform function.

It is the key point of image registration and can be described by global transformation model. The affine transformation function tells that a line in the first image can be mapped to a line in the second image and keeps balance through the affine transformation model.

2.3. Image Stitching using SIFT

Scale Invariant Feature Transform (SIFT) is used to extract features from images such as edges and boundary to help re liable matching between different views of the same object. The extracted features are invariant to scale and orientation, and are highly distinctive of the image. These are extracted in four steps. The first step computes the locations of potential interest points in the image by detecting the maxima and min ima of a set of Difference of Gaussian (DoG) filters applied at different scales all over the image. Then, these locations are refined by discarding points of low contrast. An orientation is then assigned to each key point based on local image features. Finally, a local feature descriptor is computed at each key po int. This descriptor is based on the local image gradient, tra nsformed according to the orientation of the key point to pro vide orientation invariance. Every feature is a vector of dim ension 128 distinctively identifying the neighborhood around t he key point.

2.3.1 Key pointDetection:

We adopt the Difference of Gaussians (DoGs), a Scale-invar iant key point detector. The scale space of a image is defined as a function, L(x,y,σ), with an input image, I(x,y):

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (2)$$

Where ∗is the convolution operation in x and y, and

$$G(x,y,\sigma) = \frac{1}{2\pi x\sigma^2} e^{-(x^2+y^2)/2x\sigma^2} \quad (3)$$

To efficiently detect stable key point locations in scale space, using scale space extreme in the difference of Gaussian function convolved with the image, D(x,y,σ), which can be co mputed from the difference of two nearby scales separated by a constant multiplicative factor k:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma). \quad (4)$$

There are a number of reasons for choosing this function. First, it is a particularly efficient function to compute, as the smoothed images, L, need to be computed in any case for sca le space feature description, and D can therefore be computed by simple image subtraction. Fig 3. Shows difference of Gaus sian for each octave of scale space. The initial image is rep eatedly convolved with Gaussiansto produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sam pled by a factor of 2, and the process repeated.

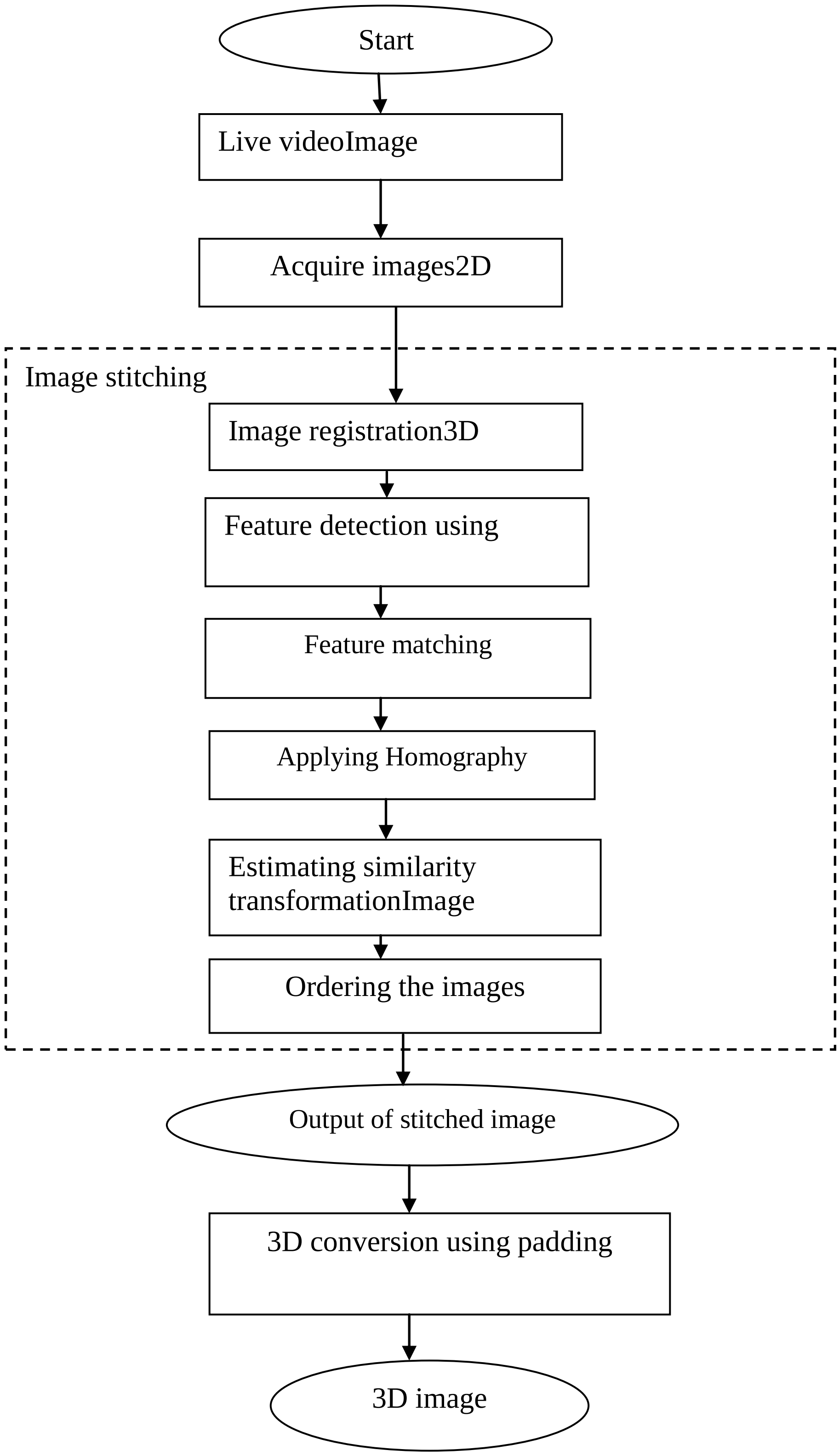

Fig.2 Flow chart of the proposed method

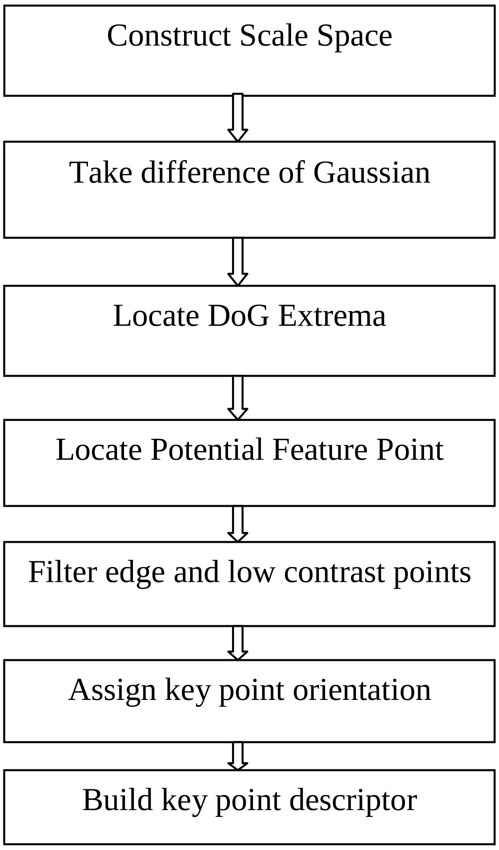

Fig.3 Image stitching mythology

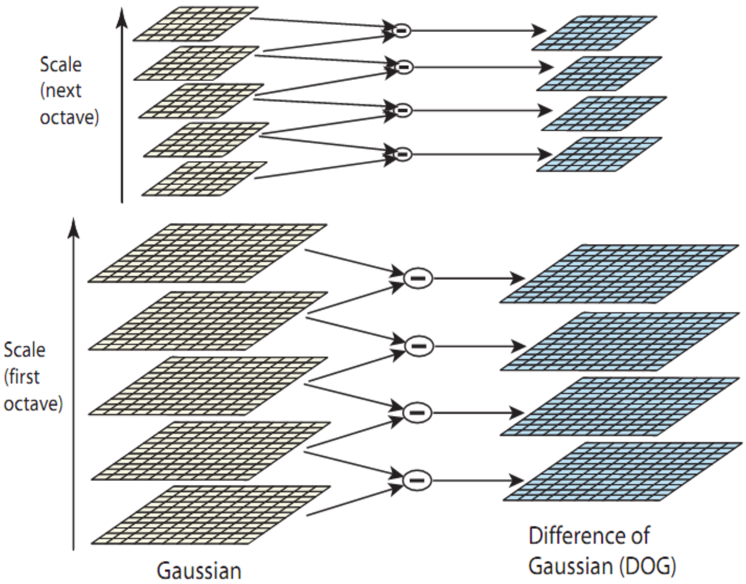

Fig.4 Ddifference of Gaussian for each octave of scale space.

2.3.2 Accurate key point localization

Once a key point candidate has been found by comparing a pixel to its neighbors, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points to be rejected that have low contrast (and aretherefore sensitive to noise. His approach uses theTaylor expansion (up to the quadratic terms) of the scale-space function, $D(x, y, \sigma)$, shifted so that the or igin is at the sample point:

$$D(x) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial x^2} X \qquad (5)$$

Where D and its derivatives are evaluated at the sample point and $x = (x, y, \sigma)^T$ is the offset from this point. The location of the extreme, $\hat{X}$, is determined by taking the derivative of this f unction with respect to x and setting it to zero, giving

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial x} \quad (6)$$

2.3.3 Eliminating Edge responses

For stability, it is not sufficient to reject keypoints with low co ntrast. The difference-of-Gaussian function will have a strong response along edges, even if the location along the edge is poorly determined and therefore unstable to small amounts of noise. The derivatives are estimated by taking differences of neighboring sample points. This is very efficient to compute, with less than 20 floating point operations required to test each key point. The experiments in this paper use a value of r = 10, which eliminates key points that have a ratio between the principal curvatures greater than 10.

2.3.4 The local image descriptor:

The previous operations have assigned an image location, sc ale, and orientation to each keypoint. These parameters impos e a repeatable local 2D coordinate system in which to describe the local image region, and therefore provide invariance to these parameters. The next step is to compute a descriptor for the local image region that is highly distinctive yet is as in var iant as possible to remaining variations, such as change in illumination or 3D viewpoint. One obvious approach would be to sample the local image intensities around the key point at the appropriate scale, and to match these using a normal ized correlation measure.

2.3.5 Key point matching

The best candidate match for each key point is found by identifying its nearest neighbor in the database of key points. The nearest neighbor is defined as the key point with m inimum Euclidean distance for the invariant descriptor vector. A more effective measure is obtained by comparing the dista nce of the closest neighbor to that of the second-closest neig hbor. If there are multiple training images of the same object, then we define the second-closest neighbor as being the clo sest neighbor that is known to come from a different object than the first, such as by only using images known to contain different objects. For false matches, there will likely be a num ber of other false matches within similar distances due to the h igh dimensionality of the feature space. We can think of the second-closest match asproviding an estimate of the density of false matches within this portion of the feature spaceand at the same time identifying specific instances of feature ambiguity.

2.3.6 Homography computation

Homography computation is the mapping between two spaces which is often used to represent the correspondence between t wo images of the same scene.The steps for Homography Det ection Algorithm using RANdom Sample Consensus (RANSAC) scheme is given below

Step 1: Firstly, corners are detected in both images.

Step 2: Variance normalized correlation is applied between corners, and pairs with a sufficiently high correlation score are collected to form a set of candidate matches.

Step 3: Four points are selected from the set of candidate matc hes, and a homography is computed.

Step 4: Pairs agreeing with the homography are selected. A p air (p, q), is considered to agree with a homography H, if for some threshold: Dist(Hp, q) < €.

Step 5: Steps 3 and 4 are repeated until a sufficient number of pairs are consistent with the computed homography.

Step 6: Using all consistent correspondences, the homography is recomputed by solving step 4.

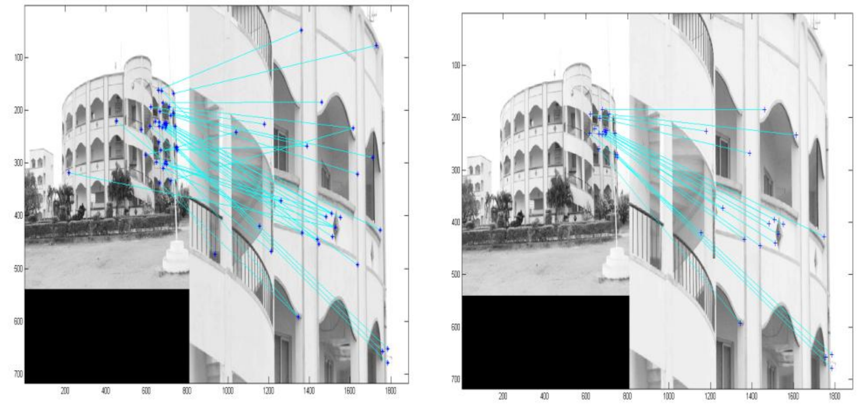

Fig.5 RANSAC

The left side of the image shows the key point matching in which some incorrect matches has occurred those are removed by homography shown in right side. Then each layer of the acquired stitched image is then given to a read image dialog b ox which converts the data of the image into a 3-D array. In th e 3-D array, x co-ordinates and y co-ordinates are pixel spac ing in x and y directions and z is the layer thickness. Then we find the size of each layer, after finding size of the layers pad ding method is applied to make the sizes of layers equal. Later we assign each layer with RGB.

3. EXPERIMENTAL RESULTS

The experimental results of the proposed system are shown in Fig.6, Fig.7 and Fig.8. Fig (6) shows the input images of the s cenes. SIFT technique has been applied on Fig.6 to match the feature in both images shown in Fig.7 then stitching is done and is shown in Fig.8. Finally an anaglyph method is applied to the stitched image to get 3D view of the scene.

Fig.6 2D Input images

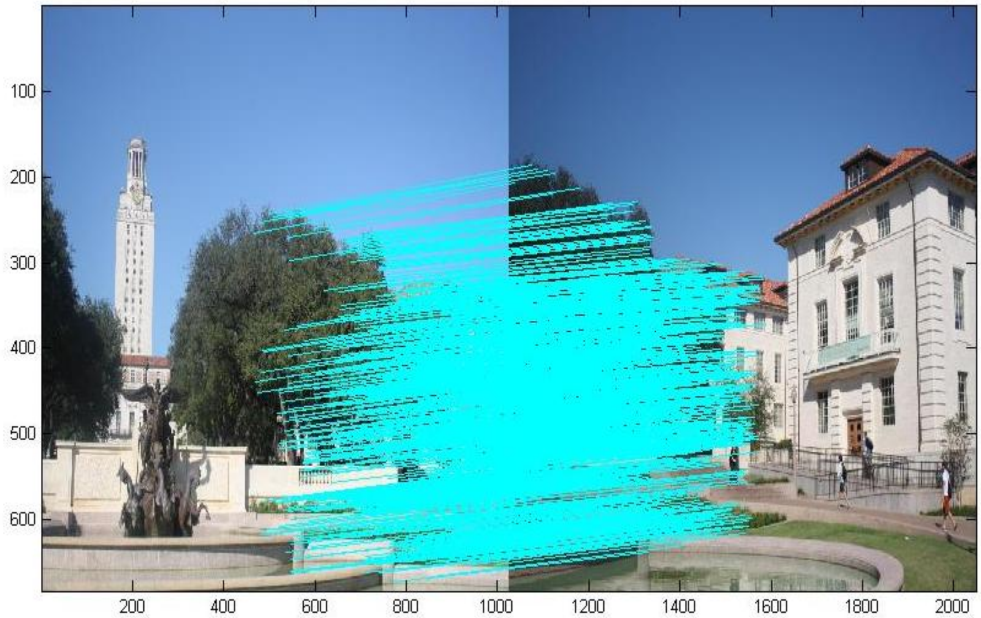

Fig.7 Feature Matching

Fig.8 Stitched Image

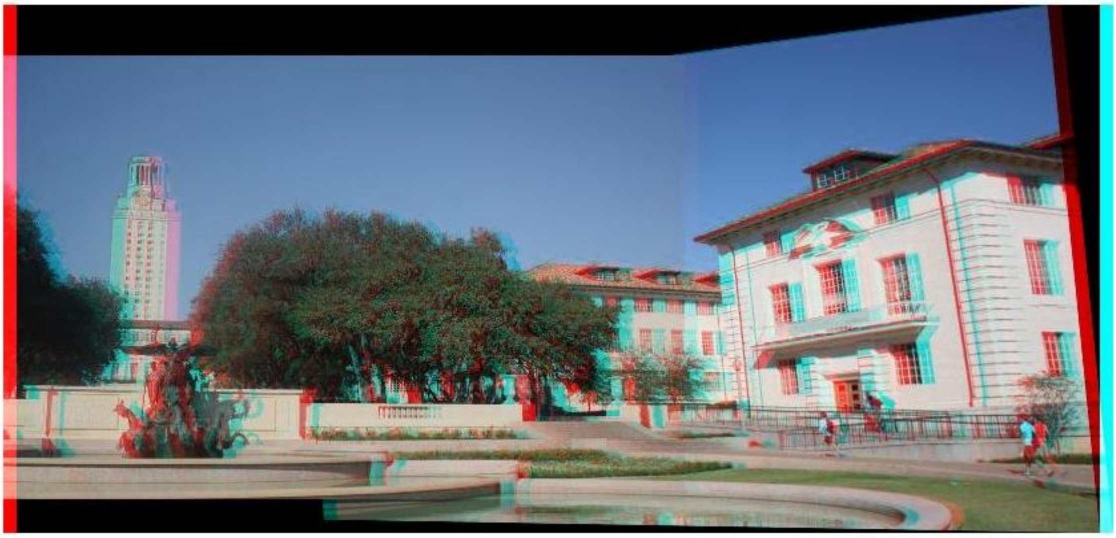

Fig.9 3D image

4. CONCLUSION

The proposed system is an automatic image mosaic method. It takes two input images and finds out the Harris corners in bot h the images, removing out the false corners in both the imag es and then use homography to matched corner pair to obtain output as mosaic. Finally a multi-layer method is applied to the stitched image to get 3D image. The proposed project uses Harris corner detector for corner and edge detection. In this paper Harris corner is used to detect corner points, which are nothing but the interest points. Interest points are used for fin ding the match between two images having some common fie ld of view. This corner detector is efficient, cheap and rotation invariant but it can't handle the images which suffer scale variation. It extracts highly distinctive features of the set of im ages and helps in helps in the correct object identification with low probability of mismatch. The proposed system performs Image stitching and 2D to 3D Image Reconstruction with high accuracy for Abnormal Activity Detection.

5. ACKNOWLEDGEMENT

This work is funded by UGC, Major Research Project, MRP-MAJOR-ELEC-2013-19701.

6. REFERENCES

[1] Lowe, David. "SIFT Keypoint Detector." http://www.cs.ubc.ca/~lowe/keypoints/

[2] R. B. Inampudi, ‖Image mosaicing‖, Proc. of Intern ational Conference on Geoscience and Remote Sensing Symposium, Seattle, pp.2363-2365, 1998.

[3] S. Peleg and J. Herman, ‖Panoramic mosaics by manifold projection‖ Proc. of IEEE Computer Society conference on Computer Vision and Pattern Recognition, San Juan, pp.338-343, 1997.

[4] B. Rousso, S. Peleg, I. Finci and R. Acha,‖ Universal mosaicing using pipe projection‖, Proc. of Sixth Inter national Conference on Computer Vision, pp.945-950,1998.

[5] J. Wang and Y. Li, ―Image mosaicing algorithm based on salient region and MVSC‖, Proc. of International Conference on Multimedia and Signal Processing, pp. 207-211, 2011.

[6] D. Ghosh, S. Park, N. Kaabouch and W. Semke,‖ Quantum evaluation of image mosaicing in multiple scene categories ‖, Proc. of IEEE Conference on Electro/Information Technology, pp.1-6, 2012.

[7] J. W. Hsieh, H. Y. M. Liao, K. C. Fan, M. T. Ko and Y. P. Hung, "Image Registration Using a New Edge-Based Approach, " Computer Vision and Image Understanding, Vol. 67, No. 2, pp. 112-130, August 1997.

[8] L. G. Brown, A survey of image registration techniques, ACM Comput. Surv. 24, No. 4, 1992, 325–376.

[9] Q. Zheng and R. Chellappa, A computational vision approaches to image registration, IEEE Trans. Image Pro cess. 2, No. 3, 1993, 311–326.