Review of Motion Estimation and Video Stabilization techniques For hand held mobile video

Paresh Rawat [1] , Jyoti Singhai [2]

Electronics & communication Deptt., TRUBA I.E.I.T Bhopal [1]
parrawat@gmail.com [1]
Electronics Deptt , MAINT BHOPAL [2]
j.singhai@gmail.com [2]

ABSTRACT

Video stabilization is a video processing technique to enhance the quality of input video by removing the undesired camera motions. There are various approaches used for stabilizing the captured videos. Most of the existing methods are either very complex or does not perform well for slow and smooth motion of hand held mobile videos. Hence it is desired to synthesis a new stabilized video sequence, by removing the undesired motion between the successive frames of the hand held mobile video. Various 2D and 3D motion models used for the motion estimation and stabilization. The paper presents the review of the various motion models, motion estimation methods and the smoothening techniques. Paper also describes the direct pixel based and feature based methods of estimating the inter frame error. Some of the results of the differential motion estimation are also presented. Finally it closes with a open discussion of research problems in the area of motion estimation and stabilization.

KEYWORS:

Video Stabilization, motion models, Interframe error, motion estimation.

1. INTRODUCTION

The videos taken from hand held mobile cameras suffer from different undesired and slow motions like track, boom or pan, these affect the quality of output video significantly. Stabilization is achieved by synthesizing the new stabilized video sequence; by estimating and removing the undesired inter frame motion between the successive frames. Generally the inter frame motion in mobile videos are slow and smooth. Video stabilization techniques can be broadly classified as mechanical stabilization, optical stabilization and image post processing stabilization. Mechanical stabilization systems based on vibration feedback through sensors like gyros accelerometers etc. have been developed in the early stage of camcorders [13]. Optical image stabilization, which has been developed after mechanical image stabilization, employs a prism or moveable lens assembly that variably adjusts the path length of the light as it travels through the camera's lens system [14]. It is not suited for small camera modules embedded in mobile phones due to lack of compactness and also due to the associated cost.

The digital image stabilization tries to smooth and compensate the undesired motion by means of digital video processing. In the image post processing algorithm, there are typically three major stages constituting a video stabilization process viz. camera motion estimation, motion smoothing or motion compensation, and mage warping. There are various algorithms proposed

DOI : 10.5121/sipij.2011.2213

for stabilizing videos taken under different environment from different camera system by modifying these three stages. Some electronic stabilization algorithms have been proposed based on the estimation of global motion vectors from local motion vectors generated by block matching of sub images. Few other stabilization techniques are there which involves global motion estimation on predefined region, [6, 15, and 16]. In this paper a modified video stabilization algorithm for hand held camera videos is proposed. The proposed algorithm uses hierarchical differential global motion estimation with Taylor series expansion

1.1 Mechanical image Stabilization

Mechanical image stabilization involves stabilizing the entire camera, not just the image. This type of stabilization uses a device called “Gyros”. Gyros consist of a gyroscope with two perpendicular spinning wheels and a battery pack.

Figure 1 Gyroscopic Stabilizer [13]

Gyroscopes are motion sensors. When the gyroscopes sense movement, a signal is sent to the motors to move the wheels to maintain stability. The gyro attaches to the camera’s tripod socket and acts like an "invisible tripod" [13]. Fig.1 shows a picture of a gyroscopic stabilizer. The vibration gyro sensors were improved by employing a tuning fork structure and a vibration amplitude feedback control. They are heavy, consume more power, and are not suitable for energy sensitive and payload constrained imaging applications as in hand held mobile cameras.

1.2 Optical Image Stabilization

The Optical Image Stabilization (OIS) system, unlike the DIS system, manipulates the image before it gets to the CCD. When the lens moves, the light rays from the subject are bent relative to the optical axis, resulting in an unsteady image because the light rays are deflected. By shifting the IS lens group on a plane perpendicular to the optical axis to counter the degree of image vibration, the light rays reaching the image plane can be steadied [17]. Since image vibration occurs in both horizontal and vertical directions, two vibration-detecting sensors for yaw and pitch are used to detect the angle and speed of movement then the actuator moves the IS lens group horizontally and vertically thus counteracting the image vibration and maintaining the stable picture.

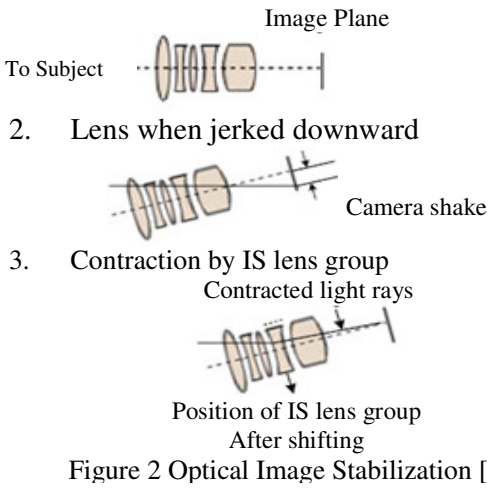

Figure 2 Optical Image Stabilization [17]

The Shift-IS component is located within the lens groups and is most effective for lower frequency movements caused by platform vibration or wind effect without increasing the overall size and weight of the master lens. Figure 2 shows an illustration of this type of image stabilization

1.3 Digital video Stabilizer

Digital stabilization method is the image post processing method or digital video stabilizer (DVS). The video stabilization can be achieved by three stages viz. motion estimation, motion smoothing or motion compensation, and mage warping. Various techniques have been proposed for stabilizing videos taken under different environment from different camera system by modifying these three stages. Various 2D stabilization algorithms are presented in [18, and 19]. Hansen *et al.* [19] describe the implementation of an image stabilization system based on a mosaic-based registration technique. Burt *et al*. [18] describe a system which uses a multi-resolution, iterative process that estimates affine motion parameters between levels of Laplacian pyramid of images. From course to fine levels, the optical flow of local patches of the image is computed using a cross-correlation scheme. The motion parameters are then computed by fitting an affine motion model to the flow.

Matsushita *et.al* [11], in 2006 proposed the direct pixel based full frame video stabilization approach using hierarchical differential motion estimation with Gauss Newton minimization. The Gausses of error function are minimized iteratively to find the optimized motion parameters. After motion estimation, motion inpainting is used to generate full frame video. This method performed well in most videos except in those cases when large portion of video frame is covered by a moving object, because this large motion makes the global motion estimation unstable. R. Szeliski, [9] in 2006 presented a survey on image alignment to explain the various motion models, and also presented a good comparison of pixel based direct and feature based methods of motion estimation. The efficiency of the feature based methods depends upon the feature point's selection [6]. Rong Hu, *et al* [12] in 2007 proposed an algorithm to estimate the global camera motion with SIFT features. These SIFT features have been proved to be affine invariant and used to remove the intentional camera motions. Derek Pang *et al* [21] in 2010 proposed the video stabilization using Dual-Tree complex wavelet

transform (DT-CWT). This method uses the relationship between the phase changes of DT-CWT and the shift invariant feature displacement in spatial domain to perform the motion estimation. Optimal Gaussian kernel filtering is used to smoothen out the motion jitters. This phase based method is immune to illumination changes between images, but this algorithm is computationally complex.

The feature-based approach, are although faster than direct pixel based approaches, but they are more prone to local effects and there efficiency depends upon the feature points selection. The direct pixel based approach makes optimal use of the information available in motion estimation and image alignment, since they measure the contribution of every pixel in the video frame. Hence for aligning the sequence of the frames in video direct pixel based approaches can be used. To further improve the stabilization efficiency hierarchical motion estimation can be used [9].

In the section 2, paper describes the various kinds of motions and 2D motion model available for the video stabilization. In the next section review of the available global motion estimation techniques are discussed. Finally some of the expected results of video stabilization are presented and paper concluded with the problems for the future.

2.0 MOTION IN MOBILE VODEOS

The person using camera are untrained, hence there are various kinds of motions present in hand held cameras. Motion in video images is caused by either the object motion or the camera movement.

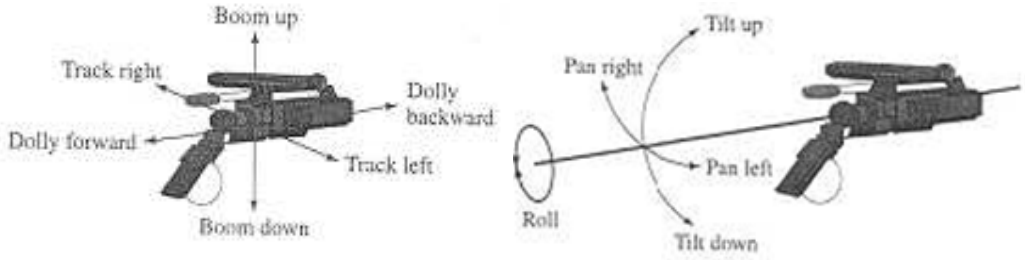

Figure 3 Typical camera motions [6].

The definition of seven main motions is extracted from [6] as shown in Fig. 3.

Track: It is defined as translation in X direction or Horizontal translation.
Boom: It is defined as translation in Y direction or Vertical translation.
Dolly: Translation in the direction of the optical camera axis.
Pan: It is defined as the Turning of camera around the vertical axis of the camera.
Tilt: It is defined as the Turning of camera around the horizontal axis of the camera.
Roll: It is defined as the Rotation around the optical axis of the camera.
Zoom: It is defined as the changes in the focal length of the camera.

These camera motions can be classified as; Unintentional or undesired camera motions viz. Track, Boom, and Pan, and Intentional or desired camera motions viz. Zoom, Roll, and Doll., That is why it is necessary to estimate first camera movement, then to compensate it and finally to generate the stabilized video by considering the estimated camera motion.

2.1 Motion models

Motions can be described either by a 2-D motion model or by a 3-D motion model. Two-Dimensional motion estimation is an important part of any video processing system. 2-D motion estimation is often the prerequisite step for 3-D structure and motion estimation. Motion estimation methods can be very different according to the desired application. The various transformations occur in the 2D plane are Translation, Euclidean or rotation, Similarity, Affine and projective as illustrated in Figure 4.

a. Translation: A 2D translation can be written as x' = x + t. or

$$x' = [\, I \quad t \,]\, x\hat{} \tag{1}$$

Where *I* is the 2x2 identity matrix and x^ is (x, y, 1) homogeneous or projective 2D coordinates. In the translation orientation is preserved.

b. Rotation +Translation: This is also known as 2D rigid body motion or 2D Euclidean transformation and can be written as

$$x' = Rx + t \quad \text{or} \quad x' = [\, R \quad t \,]\, x\hat{} \tag{2}$$

Where *R* is given as,

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3}$$

c. Scaled Rotation: Also known as similarity transform, and can be expressed as

$$x' = SRx + t \tag{4}$$

Where S is an arbitrary scale factor and given as

$$x' = [\, SR \quad t \,]\, x\hat{} = \begin{pmatrix} a & -b & t_x \\ b & a & t_y \end{pmatrix} \tag{5}$$

It is not require that $a^2 + b^2 = 1$. It preserves the angle between lines

d. Affine Transformation: It is written as

$$x' = A\, x\hat{} \tag{6}$$

Where A is the arbitrary 2x3 matrix given as

$$\mathbf{X'} = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{pmatrix} \mathbf{x\hat{}} \tag{7}$$

The parallel lines remain parallel in affine transformation. That means parallel lines are preserved.

e. Projective Transformation: This is also known as a prospective transform or homography, operates on homogeneous coordinates X^ and X'

$$X' \sim HX\hat{} \tag{8}$$

Where ~ denotes equality up to scale and *H* is an arbitrary 3x3 matrix .Note that *H* is itself homogeneous. The resulting homogeneous coordinate *X'* must be normalized in order to obtain an in homogeneous coordinate *X^*.

$$x' = \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}Y + h_{22}}, \quad y' = \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + h_{22}} \tag{9}$$

The projective transformations are the combination of affine transformation and projective warps. It uses the 8 parameters for estimation, and is commonly used with homogeneous coordinate systems. On the other hand affine transform uses 6 parameters and also have property to preserve lines, and parallel lines. Ratios are also preserved in affine transformation. Hence affine transformation is commonly used for the motion estimation problem.

2.2 Inter Frame Error

The inter frame error can be defined as the brightness difference between the two consecutive frames, or the motion between two sequential frames. The inter frame error can be modelled as either sum of square differences (SSD) or the sum of absolute differences (SAD) error between the current frame I and the motion compensated previous frame I_0. The SSD error is given by equation 10 as..

$$SSD = \sum_{x,y \in R} \left(I(x,y) - I'(x',y') \right)^2 \quad (10)$$

The SAD error is given by equation 11 as.

$$SAD = \sum_{x,y \in R} |I(x,y) - I'(x',y')| \quad (11)$$

Generally the motion is modelled by 6 parameter affine transform, and then inter frame error is minimized to find the affine parameters.

3. MOTION ESTIMATION

Motion estimation is another field of research in video processing. By using different motion estimation techniques, it is possible to estimate object motion or camera motion observed in video sequence. Object motion is defined as the local motion of the scene, and camera motion is defined as the global motion. The motion estimation techniques can be classified as feature based approaches [2, 9, 12, and 14] or direct pixel based approaches [1, 4, 8, 10, and 13].

These feature-based approaches are although faster than direct pixel based approaches, but they are more prone to local effects and there efficiency depends upon the feature point's selection. Hence they have limited performance for the unintentional motion.

The direct pixel based approach makes optimal use of the information available in motion estimation and image alignment, since they measure the contribution of every pixel. The simplest technique is to pick the search algorithm and try all possible matches, that means do the full search. But this method is very lengthy and slow. Hence to make computation faster, image pyramid based hierarchical techniques are used as alternatives. Alternatively To get sub-pixel precision in the alignment, incremental methods based on a Taylor series expansion of the image function are often used. These can also be applied to parametric motion models.

The differential motion estimation has proven to be highly effective for computing inter frame error. In this the derivatives of the inter frame error is equated to zero and then differential equation is solved to get the parameters. Differential global motion estimation is commonly used with image pyramid based hierarchical video sequence stabilization.

Hence in this paper the motion between two sequential frames, f(x, y, t) and f(x, y, t −1) is modelled with a 6-parameter affine transform. The motion vectors between two images can be described by a single affine transformation as given by eq. (1). Where m_1, m_2, m_3, m_4 represents the 2 × 2 affine rotation matrix A, and m5 and m6 the translation vector $\overline{T}$.

$$f(x,y,t) = f(\mathrm{m_1x} + \mathrm{m_2y} + \mathrm{m_5}\,,\ \mathrm{m_3x} + \mathrm{m_4y} + \mathrm{m_6}\,, t-1) \quad (12)$$

Where $$A = \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix} \ and \quad \overline{T} = \begin{pmatrix} m_5 \\ m_6 \end{pmatrix} \quad (13)$$

In order to estimate the affine parameters, we define the following quadratic error function to be minimized.

$$E(m) = \sum_{x,y\in\Omega} [f(x,y,t) - f(m_1x + m_2y + m_5\,, m_3x + m_4y + m_6\,, t-1]^2 \quad (14)$$

Where Ω denotes a user specified region of interest here it is the entire frame. Since this error function is non-linear its affine parameters m, cannot be minimized analytically. To simplify the minimization, this error function is approximated by using a first-order truncated Taylor series expansion. The quadratic error function is now linear in its unknowns, *m* and can therefore be minimized analytically by differentiating with respect to *m* as shown in eq.15.

$$\frac{dE(m)}{dm} = \sum_{\Omega} 2C[k - C^T m] \quad (15)$$

The temporal derivatives can be find using Separable filters as mentioned in [1]**.** Finally L-level Gaussian pyramid is built for each frame, f(x, y, t) and f(x, y, t − 1). The motion estimated at pyramid level L is used to warp the frame at the next higher level L − 1, until the finest level of the pyramid is reached (the full resolution frame at L = 1). Large motions are estimated at coarse level by warping using bicubic interpolation and refining iteratively at each pyramid level.

4 EXPECTED RESULTS

The performance of the proposed video stabilization algorithm is tested on various real time video sequences with the resolution of 176 x 144. The motion in input videos between each pair of frames is stabilized using differential global motion estimation. The inter frame error between original input frames are compared with, inter frame error after motion estimation with Bicubic interpolation and spline interpolation methods. The some of the expected results of the differential motion estimation technique are given here in Fig. 5 for a college video taken from mobile phone camera. The results of two different frames are shown. The stabilized frame has slight blurring effect due to error compensation. Comparison of the Mean square error and Signal to noise ration are presented in the Fig. 6 and 7 for the 10 consecutive frames having the maximum motion in the given two videos. It is clear from the Fig.6 that bicubic interpolation gives better stabilized results. From Fig.7 it can be evaluated that with proposed algorithm using a Bicubic interpolation MSE and SNR are more stabilized while with simple mean, median filters and Spline interpolation variation in MSE are very large.

Figure. 5 Video stabilization (a) previous frame (b) current frame (c) Stabilized Frame

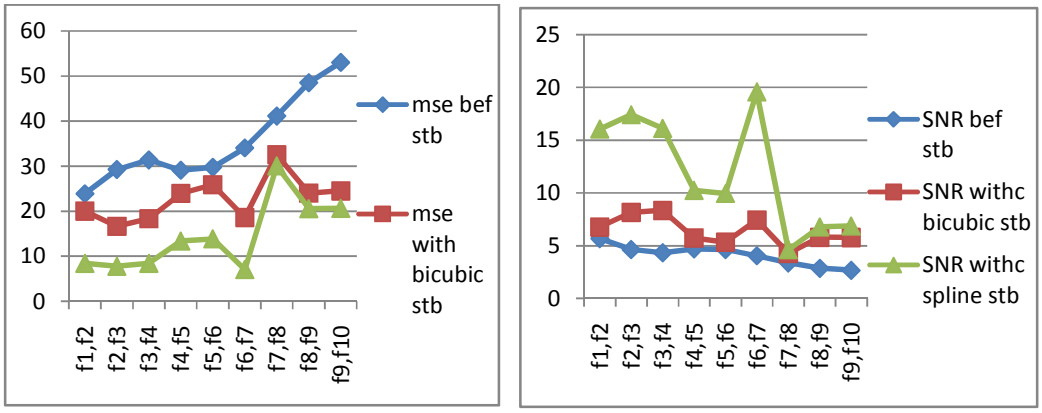

Figure. 6 Motion estimation results a) MSE Comparision b) SNR compairision for coaridoor video (before stabilization, with bicubic interpolation and with spline interpolation)

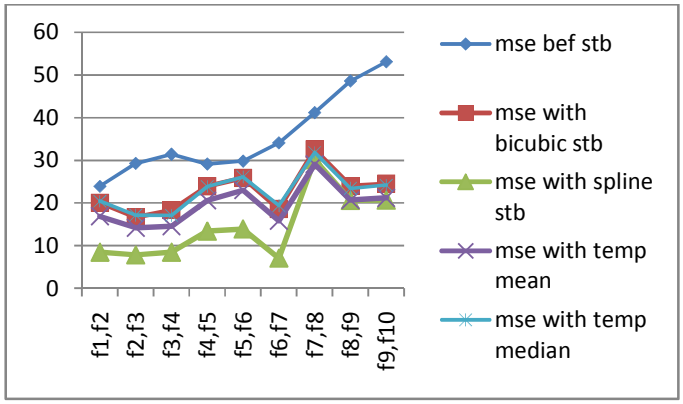

Figure. 7 compairison of results for motion estimation with tempral mean and tempral median filters a MSE comparision

5. DISCUSSION

Review of the various motion estimation and stabilization techniques are presented in the paper. In section three good comparison of the direct and feature based method are presented it is clear that the feature based approaches are suitable for the task of image alignment and stitching, Where features are important and the local effects are prone. Whereas direct method is suitable for stabilizing the global camera motions. Hence in this paper a video stabilization algorithm for hand held camera videos is proposed using differential motion estimation. The results obtained with the proposed algorithm shows the stabilized motion in n after motion estimation. The use of Taylor series improves the convergence rate and increases the efficiency of the motion estimation. Some of results are presented for 10 consecutive frames with maximum motion. The method gives best stabilization with bicubic interpolation; it is found that peak to peak variation in MSE is reduced from 30 to 17. Due to accumulation error and wrapping results after motion estimation having blurring effects and some missing image areas. Motion smoothening is the scope for the future the computation cost can also be reduced to improve the efficiency of the estimation and stabilization in future work.

6.REFRENCES

[1] Hany Farid and Jeffrey B. Woodward, " Video stabilization & Enhancement" TR2007- 605, Dartmouth College, Computer Science, 1997.

[2] C. Schmid and R. Mohr.," Local gray value invariants for image retrieval", *IEEE Trans. on Pattern analysis and Machine Intelligence*, vol. 19 No.5, : pages 530 -535, May 1997.

[3] F. Dufaux and Janusz Konrad ,' Efficient robust and fast global motion estimation for video coding," IEEE Trans. on Image Processing , vol.9, No. 3, pages 497-501 March 2000.

[4] C. Buehler, M. Bosse, and L. McMillian. "Non-metric image based rendering for video stabilization". *Proc. Computer Vision and Pattern Recog.*, vol.2: page 609–614, 2001

[5] J. S. Jin, Z. Zhu, and G. Xu., "Digital video sequence stabilization based on 2.5d motion estimation and inertial motion filtering", *.Real-Time Imaging, vol.7 No. 4: pages 357– 365, August 2001.*

[6] Olivier Adda., N. Cottineau, M. Kadoura, "A Tool for Global Motion Estimation and Compensation for Video Processing " LEC/COEN 490, Concordia University , May 5, 2003.

[7] D .G. Lowe, "Distinctive image feature from scale invariant key points", *Int. Journal of Computer Vision*, vol. 60 No.2: pages 91–110, 2004.

[8] Y. Wexler, E. Shechtman, and M., Irani, "Space-time video completion. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1 pages 120–127, 2004

[9] R. Szeliski, "Image Alignment and Stitching: A Tutorial," Technical Report MSR-TR- 2004-92, Microsoft Corp., 2004.

[10] H.-C. Chang, S.-H. Lai, and K.-R. Lu. "A robust and efficient video stabilization algorithm. *ICME '04: Int. Conf. on Multimedia and Expo*, vol. 1: pages 29–32, June 2004

[11] Y. Matsushita, E. Ofek, W.Ge, XTang, and H.Y.Shum,." Full frame video Stabilization with motion inpainting." *IEEE Transactions on Pattern Analysis* and *Machine Intellig, vol.* 28 No. 7: pages 1150-1163, July 2006.

[12] Rong Hu1, Rongjie Shi1, I-fan Shen1, Wenbin Chen2 "Video Stabilization Using Scale Invariant Features".*11th Int. Conf. Information Visualization* IV'07 IEEE 2007.

[13] Multimedia, "Use Image Stabiliza. For Gyroscopic Stabilizer", [online], URLhttp://www.websiteoptimization.com/speed/tweak/stabilizer. Access 13-January- 2009].

[14] C. Morimoto, R. Chellappa, "Fast electronic digitl image stabilization" Proc. 13th International Conference on Pattern Recognition, vol. 3, pages .284-288, 25-29 August 1996

[15] J.Y. Chang, W. F. Hu, , M.H. Chang and BS Chang, "Digital Image Trans. and Rotational Motion Stabilization using Optical Flow Technique", *IEEE Trans.. On Consumer Electronics,* vol. 48. No. 1, pages. 108- 115, 2002.

[16] C. Morimato and R. Chellappa, "Evaluation of Image Stabilization Algorithms", *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Proc,* Vol. 5, pages. 2789-2792, 1998.

[17] Canon Digisuper 100xs, Product Manual, [online] URL http:// www.canon.com/bctvproducts/pdf/ 1 00xs.pdf [Accesses 20 May 2009]

[18] M. Hansen, P. Anandan, K. Dana, G. van der Wal, and P.J. Burt. "Real time scene stabilization and mosaic construction". *Int. Proc. DARPA Image understanding Worksop,* pages 457-465, Monterey, CA, November 1994.

[19] Feng Liu, Michael Gleicher, Hailin Jin, **Aseem Agarwala,** "Content-Preserving Warps for 3D Video Stabilization," *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2009),* vol. 28 No. 3,, pages. 44:1-44:9, 2009.

[20] J.M. Wang, H.P. Chou, S.W. Chen, and C.S. Fuh, "Video stabilization for a hand-held camera based on 3d motion model," in IEEE Int. Conf. on Image Processing (ICIP), pages. 3477 –3481, 6th Nov. 2009,.

[21] Derek Pang, Huizhong Chen and Sherif Halawa, "Efficient Video Stabilization with Dual-Tree Complex Wavelet Transform", EE368 Project Report, Spring 2010.

Authors

1. Paresh rawat have doen BE In Electronics and Communication, In 2000 and M.Tech. in Digital Communication From MANIT Bhopal in 2007 and Currently Doing Ph.D From MANIT Bhopal is Working as Astt. Prof. In TRUBA. I.E.I.T. Bhopal

2. Dr. Jyoti Singhai (Ph.D.)
Currently Working as Prof. In MANIT Department of Electronics and Communication Engineering