Video Stabilization, Camera Motion Pattern Recognition and Motion Tracking Using Spatiotemporal Regularity Flow

Karthik Dinesh and Sumana Gupta
Indian Institute of Technology Kanpur/ Electrical, Kanpur, INDIA
Email: karthik.dinesh17@gmail.com, sumana@iitk.ac.in

***Abstract*—In this paper we propose a different approach based on a spatio-temporal feature called the Spatio Temporal Regularity Flow (SPREF) to stabilize unwanted camera motions in a video, recognize the camera motion patterns between consecutive frames and Group of Frames(GOF) and track the motion of an object in a video with the background subtracted. The method for stabilization based on Camera Motion uses the Translational Regularity flow vectors (TSPREF). In this method we fit the TSPREF vectors into parametric model to calculate the unstabilized global motion. An adaptive Gaussian smoothing method is used to smoothen the global motion followed by motion compensation to produce a stabilized sequence. Experimental results are provided and the stabilization achieved is validated using the qualitative measure Interframe Transform Fidelity (ITF).In camera motion pattern recognition we make use of TSPREF vectors to recognize the cognizant camera motion patterns. This is done for consecutive frames as well as Group of Frames(GOF)of different video sequences. In motion tracking we use the TSPREF vectors to track the moving object present in a video. The test videos taken have a still background with one or two moving objects. In all the cases we have the background subtracted from the moving object.**

***Index Terms*—video stabilization, camera motion pattern, global motion estimation, gaussian smoothing, motion compensation, motion tracking, background subtraction, spline, regularity flow**

I. Introduction

The video acquisition process is often confronted by various degradations like jitter caused as a result of hand motion or platform vibrations. The aim of video stabilization is to reduce the unwanted camera motion. The three basic steps involved in the video stabilization are 1) Global Motion Estimation 2) Global Motion Smoothing 3) Global Motion Compensation.

Many approaches are reported in the literature to solve the stabilization problem. We can broadly classify them as stabilization based on using either image features or that based on block motion vectors. Tsung-Han Tsai *et al.* [1] have used features like corner points to calculate the global motion and stabilize the sequences. Homer H. Chen *et al.* [2] have used block motion vectors to find the global motion. Hung-Chang Chang *et al.* [3] have used the optical flow to estimate the global motion.

In this paper our main objective is to develop a different approach to the problem of video stabilization. We use the Spatiotemporal Regularity Flow (SPREF) [4] vectors to find the global motion of an unstabilized frame. In the proposed method, the TSPREF vectors are fitted into a 4-parameter affine motion model to obtain the global motion parameters. We smoothen it using a Gaussian kernel [5] in an adaptive manner. Adaptiveness is brought in due to the use of the temporal correlation between two consecutive frames. It is this adaptive nature that helps us to experiment with different correlation coefficient values which in turn gives us different window lengths in single run of the algorithm.

Camera motion characterization is used in applications like video indexing, shot detection, scene change detection etc. Camera motion characterization has two parts viz. the camera motion representation and the camera motion pattern recognition. We concentrate upon the camera motion pattern recognition. In rest of paper pattern recognition implies camera motion pattern recognition. In camera motion pattern recognition we adopt a new approach using TSPREF vectors to find the motion patterns. Previously block motion vectors have been used to do the same for consecutive frames. Ngo *et al.* [6] have analyzed the spatiotemporal slices in XT and YT domains and used a tensor histogram approach to classify various motions patterns of camera. In this paper we use a nonparametric approach and find the camera motion patterns between consecutive frames and group of frames (GOF).

Motion tracking has been used in various instances. For e.g. the surveillance cameras used in the airports and offices have the motion tracking software. Other areas like sports, traffic systems etc. have also utilized it. The problem of motion tracking has been approached by using various methods. Ssu-Wei Chen *et al.* [7] have used Three Temporal Difference and Gaussian Mixture Model for motion tracking purpose. Saravanakumar. S *et al.* [8] has developed multiple human tracking approach based on the motion estimation and detection, shadow removal, occlusion detection. In this paper, as mentioned earlier

Manuscript received February 15, 2014; revised May 9, 2014.

doi: 10.12720/joig.2.1.33-40

we use the TSPREF vectors to track the motion of the object. TSPREF gives the direction in which a video is regular. So in case of videos with still background, we do not have any motion in that area which implies that TSPREF is zero. We obtain the TSPREF vectors for a region having motion. In this way we will be able to differentiate the background from the objects in motion in a video. Hence the use of TSPREF is the novel approach in solving various applications discussed above.

The rest of paper is organized as follows. Section 2 introduces the concept of SPREF (specifically TSPREF). Section 3 discusses the method of stabilization using camera motion based approach. Section 4 discusses the algorithm for motion tracking in a video with subtracted background. Section 5 discusses the camera motion pattern recognition. Section 6 gives results for video stabilization, pattern recognition and motion tracking. Section 7 concludes the paper.

II. Spatiotemporal Regularity Flow

Spatiotemporal Regularity Flow defined in [4] gives the direction, in which a video is regular, i.e. the directions in which pixel intensities vary the least. Here we try to find the directions in which sum of directional gradients is minimum. This is represented by a continuous energy equation given as

$$\int_{\Omega}\left|\frac{\partial(I*H)(x,y,t)}{\partial F(x,y,t)}\right|^2 dxdydt \quad (1)$$

Since the video is discrete in nature, we discretize the above energy equation to get,

$$E(F_t)=\sum_{\Omega}\left|(I*\frac{\partial H}{\partial x})c_1(t)+(I*\frac{\partial H}{\partial y})c_2(t)+(I*\frac{\partial H}{\partial t})\right|^2 \quad (2)$$

Here $[c_1(t), c_2(t), 1]$ is the flow vector. We find the flow vector that minimizes $E(F_t)$. Since the flow vector is a function of time we refer it as Translational Regularity Flow (TSPREF)vector. The advantage of the regularity flow is that we can find the flow vectors for the whole video cube at one time instant. The regularity flow vector components are expressed using the spline approximation [4] method. Box spline of degree one has been used that is shifted and scaled. The regularity flow vector components are given as

$$c_m(u)=\sum_n \alpha_n^m b(2^{-l}u-n)=\sum_n \alpha_n^m b_n^l(u) \quad (3)$$

α_n^m represents the spline coefficient, l represents the scaling factor that varies from 1...k, where $k=log_2\,(totalframes)$ and $n=2^l$.

The TSPREF is global in nature. Video however, also has local motions. So in order to calculate optimal TSPREF, oct tree segmentation is used i.e. the video is segmented into smallest possible levels and TSPREF is calculated for all the segmented video cubes. Using the TSPREF vectors video is reconstructed and the reconstruction error is calculated using,

$$Error=\sum_{x,y,t}|I_{main}(x,y,t)-I_{recons}(x,y,t)|^2 \quad (4)$$

where $I_{main}(x, y, t)$ is the original video cube and $I_{recons}(x, y, t)$ is the reconstructed video cube formed by shifting the original video cube by an amount equal to the flow vectors calculated from previous equations. Now optimal video is got by using an error criterion as follows

$$Error_i<\sum_{q=1}^{8} Error_{i,q} \quad (5)$$

where $Error_i$ is the error of the parent video and $Error_{i,q}$ is the error of a child video. Since it is oct-tree segmentation, each parent video is divided into eight child videos. Hence q varies from 1 to 8.

Evaluation of optimal TSPREF vectors for the unstabilized video is an important step prior to applying the stabilization method.

III. Stabilization Using the Camera Motion Estimation

- Global TSPREF Estimation: - Here our aim is to find the global motion for the frame. So the TSPREF is fit in to a 4-parameter affine motion model [3]. The affine motion model is represented as

$$\begin{pmatrix} x' \\ y' \end{pmatrix}=\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}+\begin{pmatrix} c \\ d \end{pmatrix} \quad (6)$$

here 'a' represents scaling, 'b' represents rotation, 'c' represents the horizontal motion and 'd' represents the vertical motion. Now, consider we have N TSPREF's for a frame. x' and y' in the next frame is got by adding the TSPREF to the pixels x and y in the present frame. So (6) can be transformed into over constrained linear system as shown below [3].

$$\begin{pmatrix} x_1 & -y_1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & -y_n & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_n & x_n & 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}=\begin{pmatrix} x_1+u_1 \\ \vdots \\ x_1+u_n \\ y_1+v_1 \\ \vdots \\ y_1+v_n \end{pmatrix} \quad (7)$$

Since this system is susceptible to the outliers, which may have occurred during the reconstruction of video during the oct tree segmentation, they are calculated as

$$U_n'=ax_n-by_n+c-x_n' \quad (8)$$

We collect the statistics and calculate the mean and standard deviations. The data points with large deviation are considered as outliers. The remaining data points are once again introduced into (7) and process is repeated till the outliers are few. Thus we get the global TSPREF estimate for the frame.

$$V_n'=ay_n+bx_n+d-y_n' \quad (9)$$

- Global TSPREF Smoothing:-In this step we create a transformation matrix from the camera motion parameters estimated in the previous step which is given as

$$T_i=\begin{pmatrix} a_i & -b_i & c_i \\ b_i & a_i & d_i \\ 0 & 0 & 1 \end{pmatrix} \qquad (10)$$

Now our aim is smooth this global transformation matrix T_i for a frame. This transformation matrix is smoothed using a Gaussian kernel [5]. The smoothed transformation matrix S_i is given as,

$$S_i=\sum_{i\in N_t} T_i * G(k) \qquad (11)$$

where $G(k)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{k^2}{2\sigma^2}}$ and $\sigma=\sqrt{k}$. N_t is defined as $N_t=\{j: t\text{-}k\leq t\leq t+k\}$. The range specified is considered as window. This window is made adaptive i.e. the choice of the length of the window is made dependent on the temporal correlation [9] between the successive frames. The temporal correlation between two frames is given by the formula,

$$C_{t,t+l}=\frac{\Sigma_{i,j}\left(\left(p_t(i,j)-\bar{p}_t\right)\cdot\left(p_{t+l}(i,j)-\overline{p_{t+l}}\right)\right)}{\sqrt{\Sigma_{i,j}\left(\left(p_t(i,j)-\bar{p}_t\right)^2\cdot\left(p_{t+l}(i,j)-\overline{p_{t+l}}\right)^2\right)}} \qquad (12)$$

where $l=1$ for correlation between the successive frames. 't' is the frame number. $p_t(i, j)$ represent pixel at (i, j) in frame 'k'. $C_{t,t+l}$ represents correlation coefficient between t th and $t+l$ th frame. $\overline{p_t}$ and $\overline{p_{t+l}}$ represent mean value of frames at time instants t and $t+l$.

Depending on the value of temporal correlation we can change the window length. The logic we have used is to have window of larger length if the temporal correlation is less and a window of smaller length if temporal correlation is high. The parameters viz. the window length '*k*' and correlation coefficient '*C*' can be adaptively chosen as per the requirements of the video.

- Global TSPREF Compensation:-The difference between the smoothed global motion and the unstabilized global motion is used to motion compensate a frame. It is given as

$$d=g^s-g^m \qquad (13)$$

d gives us a value by which we compensate a frame.

After the completion of this step we obtain a stabilized frame. This has to be performed for all the frames and thus the video gets stabilized.

IV. Motion Tracking Using TSPREF

The algorithm for motion tracking with subtracted background is described below:

1) At first we calculate the TSPREF for a test video cube.
2) After calculating the TSPREF for the cube we rearrange it framewise.
3) We look for TSPREFs that are zero. Zero TSPREF implies no motion. The corresponding points in the video are made zero.
4) Initially we may not get the correct background subtract and tracked video which can attributed to the fact that at the edges of the moving object there may be some pixels in certain frames, which belongs to region with no motion, that may not have been zero. Since this is not a problem in every frame, we need to sum the video (which we have got from step 3) over all the frames and then decide some threshold which varies for different videos. The pixels above the threshold are kept intact and the pixels below the threshold are made zero.
5) Step 4 is iterative and needs to be performed until we get the required output video. Sometimes we may also have to interpolate the information from either previous or the subsequent frames. The results are shown in next section.

V. Camera Motion Pattern Recognition

In any video scene there are two kinds of motion viz. the camera motion and the object motion. The camera motion is referred to as the global motion and the object motion is referred to as the local motion. So we have different camera motions like camera horizontal left, camera horizontal right, camera vertical up, camera vertical down, still camera, camera zoom. In this paper by using the TSPREF we will be able to recognize various kinds of the above mentioned camera motion.

At first we consider a group of frames and calculate the optimal TSPREF of the sequence of frames. The TSPREF has both x and y components. Re-arrange the SPREF framewise and calculate the magnitude and the direction of the SPREF between two frames. Plot the magnitude and the angular histogram. Same process applies for finding patterns between the group of frames. The equations are indicated as follows:

$$SPREF_{mag}=\sqrt{(SPREF_x)^2+(SPREF_y)^2} \qquad (14)$$

$$SPREF_{dir}=\arctan\left(\frac{SPREF_y}{SPREF_x}\right) \qquad (15)$$

We divide the angular histogram into different sectors as shown in Fig. 1. The typical magnitude and the angle histograms for six different camera motion patterns are tabulated in Table II (first 3 columns).

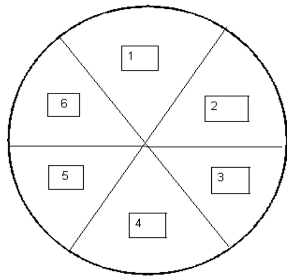

Figure 1. Polar angle histogram

VI. Results

We have tested our video stabilization algorithm on video named OUTDOOR. We have divided this video into three parts and named as OUTDOOR1, OUTDOOR2, and OUTDOOR3. Two parts have 256 frames and final one with only 128 frames. The motion plots of the sequences are shown in Fig. 2 to Fig. 7.

Fig. 2 and Fig. 3 show the unstabilized and the smoothed global motion in the x and y directions obtained by stabilizing the video sequence using the camera motion estimation method. The run time for the method is 0.9sec/frame for the sequence OUTDOOR1. Fig. 4, Fig. 5 as well as Fig. 6, Fig. 7 show the same for video sequences OUTDOOR2 and OUTDOOR3respectively. We have run the algorithm in Intel Core i5 2.3 GHz 64-bit operating system.

A qualitative measure called the Interframe Transform Fidelity (ITF) that measures PSNR between successive frames is evaluated to find the stabilization performance of the proposed algorithm. ITF is given as

$$\mathrm{ITF}=\frac{1}{\mathrm{N}-1}\sum_{\mathrm{k}-1}^{\mathrm{N}-1}\mathrm{PSNR}(\mathrm{i_k},\mathrm{i_{k+1}}) \qquad (16)$$

where N is total number of frames in video and $PSNR$ is peak signal to noise ratio between two frames which is given as

$$\mathrm{PSNR}(\mathrm{i_k},\mathrm{i_{k+1}}) = 10\log_{10}\frac{\mathrm{i}^2_{\mathrm{max}}}{\mathrm{MSE}(\mathrm{i_k},\mathrm{i_{k+1}})} \qquad (17)$$

where i_{max} is maximum intensity value of a pixel and MSE is the mean square error between the two frames which is given as

$$\mathrm{MSE}(\mathrm{i_k},\mathrm{i_{k+1}})=\frac{1}{\mathrm{w}\times\mathrm{h}}\sum_{\mathrm{x}=1}^{\mathrm{h}}\sum_{\mathrm{y}=1}^{\mathrm{w}}\left(\mathrm{i_k}(\mathrm{x},\mathrm{y})-\mathrm{i_{k+1}}(\mathrm{x},\mathrm{y})\right)^2 \qquad (18)$$

where w and h are width and the height of the frame. A stabilized sequence should have higher ITF than the unstabilised sequence. Table I shows the ITF for 3 different unstabilized and the stabilized video sequences obtained using the proposed method. We observe the improvement in the stabilized video compared to the unstabilized video for all the sequences. Since OUTDOOR1 has more jitter, we can see more improvement in it. OUTDOOR3 has least improvement.

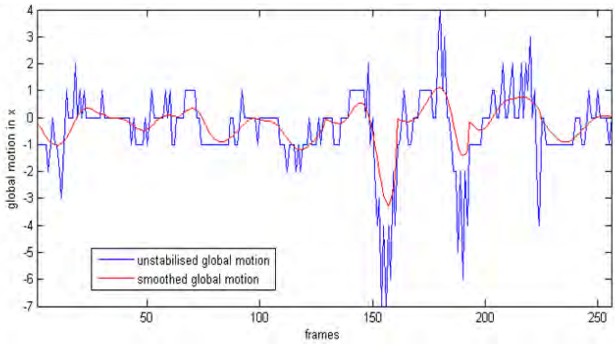

Figure 2. Unstabilized and smoothed global motion in x-direction using camera motion estimation method for OUTDOOR1

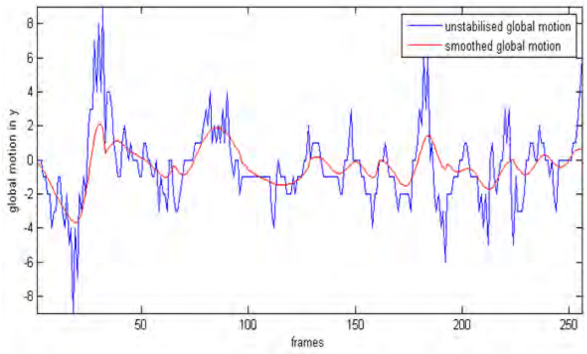

Figure 3. Unstabilized and smoothed global motion in y-direction using camera motion estimation method for OUTDOOR1

For the camera motion pattern recognition we have used the video sequences coastguard, husky and Stefan. We have taken 128 frames in each sequence. The concentric circles in the angular histogram plot indicate total number of motion vectors aligned in that particular direction. For example, in Fig. 8 we can see there are 15000 vectors which are aligned in the sector 2. The camera motion patterns (both polar and magnitude histograms) for consecutive frames and group of frames are shown in Fig. 8 to Fig. 21, for different video sequences.

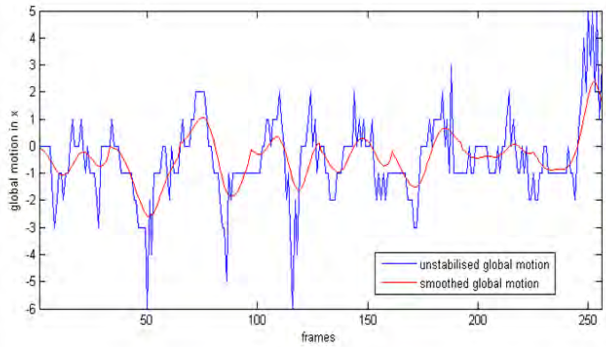

Figure 4. Unstabilized and smoothed global motion in x-direction using camera motion estimation method for OUTDOOR2

Between the frame 85 and 86 of the coastguard sequence (144*176) there is a horizontal movement to the right. Dominant angle bin is located in the sector 5. This is shown in Fig. 10.

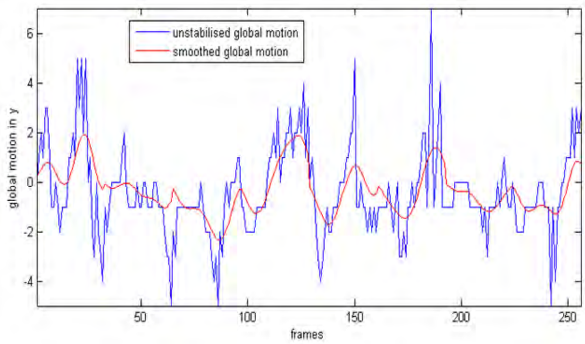

Figure 5. Unstabilized and smoothed global motion in y-direction using camera motion estimation method for OUTDOOR2

In same coastguard sequence (144*176), between frame 27 and 28 there is a horizontal left movement and it is shown in Fig. 8 with dominant bin in the sector 2. Between the frame 71 and 72 of the coastguard sequence there is a vertical up movement which is shown in the Fig. 12 with dominant bin in sector 1. In the husky video sequence (288*352), between the frames 57 and 58 there is camera vertical down movement which is shown in the Fig. 14 with a dominant bin in sector 4.

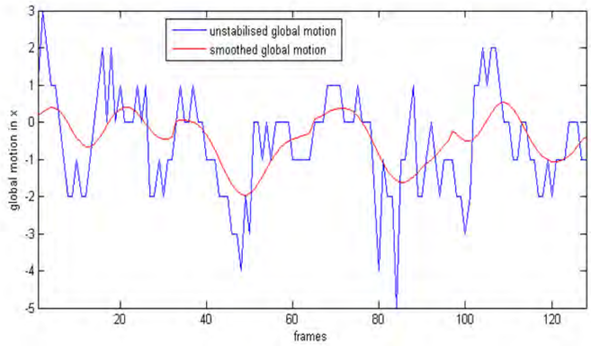

Figure 6. Unstabilized and smoothed global motion in x-direction using camera motion estimation method for OUTDOOR3

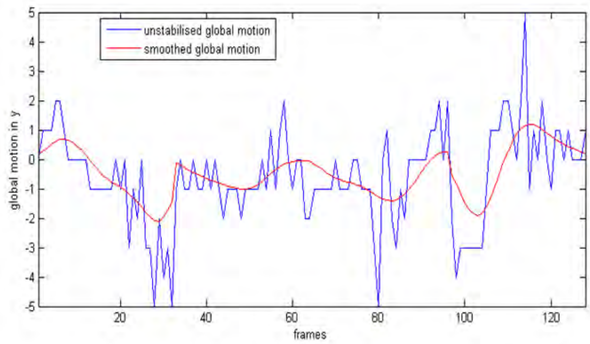

Figure 7. Unstabilized and smoothed global motion in y-direction using camera motion estimation method for OUTDOOR3

In Stefan sequence (240*320) between the frame 19 and 20 there is a zoom and this is shown in Fig. 16 where there are more than one dominant bin in sector 1 and 2.

TABLE I. ITF USING CAMERA MOTION ESTIMATION

Video Sequence	Original in dB	Stabilized in dB
OUTDOOR 1	23.5966	25.3575
OUTDOOR 2	23.0348	24.3468
OUTDOOR 3	23.3200	23.9143

In table tennis sequence, between frames 111 and 112 camera is still. This is shown in Fig. 18 with dominant bin in sector 2. But this may also represent a horizontal left movement. So to avoid the confusion we go for the magnitude histogram which is shown in the Fig. 19 with only one dominant bin positioned at zero. This holds good for all still sequences. Although we have represented the magnitude histograms (Fig. 9, Fig. 11, Fig. 13, Fig. 15, Fig. 17, Fig. 19) adjacent to the polar histograms of different video sequences, it is not of much importance as it does not give a clear idea about the pattern between the two frames except for the case where the camera is still. All the above mentioned results regarding the camera motion pattern recognition have been summarized in Table II.

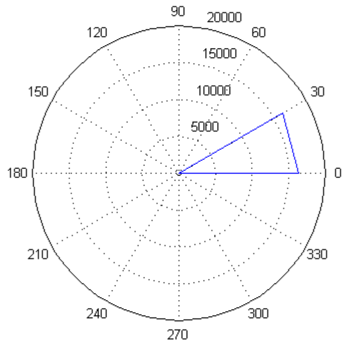

Figure 8. Horizontal left

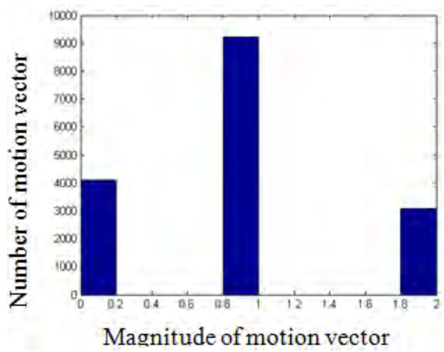

Figure 9. Horizontal left magnitude

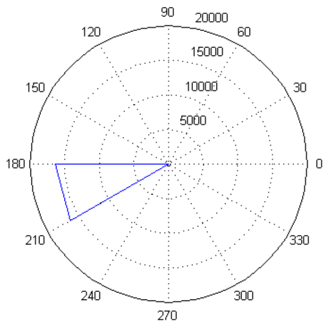

Figure 10. Horizontal right

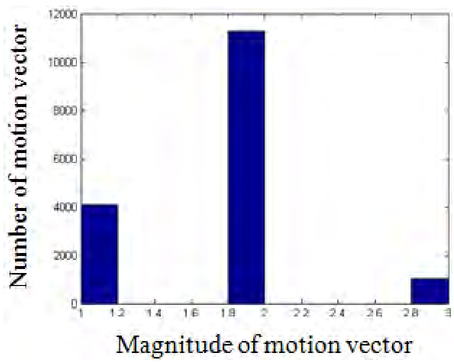

Figure 11. Horizontal right magnitude

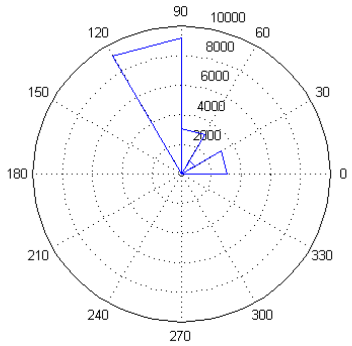

Figure 12. Vertical up

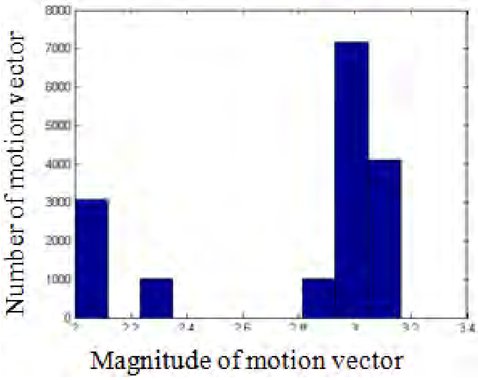

Figure 13. Vertical up magnitude

Next we look into camera motion pattern recognition between the group of frames. As already mentioned, it is the availability of TSPREF which helps us to look into group of frames. We have taken the coastguard sequence (144*176*40) from which frames from 48 to 88 i.e. a group of 40 frames are taken. These frames consist of three kinds of motion. First there is a horizontal left movement for 20 frames followed by vertical up movement for 8 frames and then horizontal right movement for 12 frames. This is shown in Fig. 20 where we can see there are bins found in 3 sectors with highest bin length for sector 2. This is because we have horizontal left for 20 frames, hence large number of TSPREF fall into this region. This is followed by horizontal right in sector 5 and vertical up in sector 1.

Figure 14. Vertical down

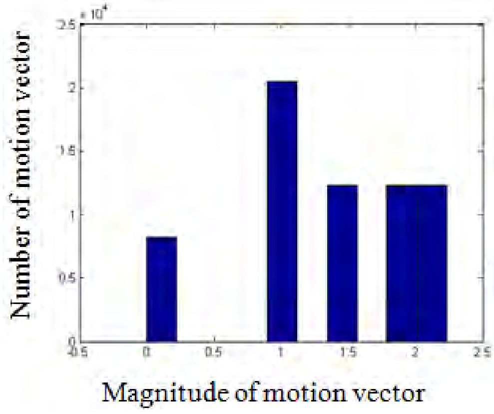

Figure 15. Vertical down magnitude

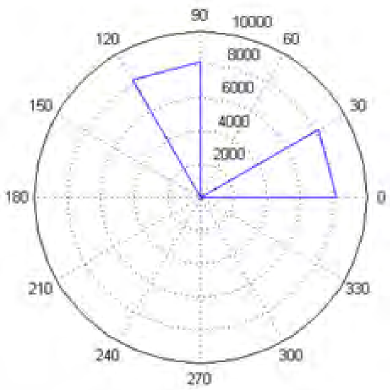

Figure 16. Camera zoom

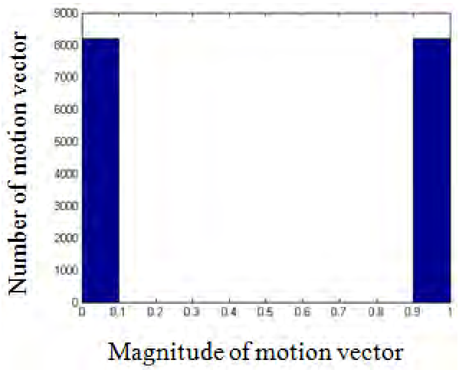

Figure 17. Camera zoom magnitude

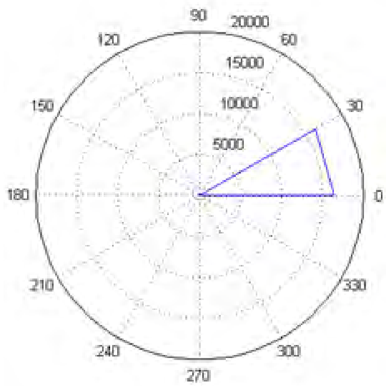

Figure 18. Camera still

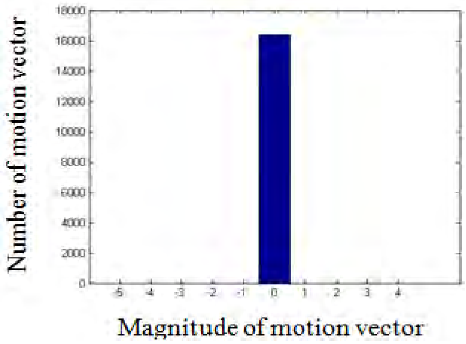

Figure 19. Camera still magnitude

We have taken another sequence named husky (288*352*112), from which group of 112 frames have been taken.

These frames initially consist of horizontal left movement for 40 frames followed by vertical down movement for 40 frames and then horizontal right movement for 32 frames. We can see this in Fig. 21 where 3 bins are present in 3 sectors.

Now we look into time taken to recognise the pattern between the two frames of video sequence. The Table III shows time taken to recognise a pattern between two frames of different video sequences.

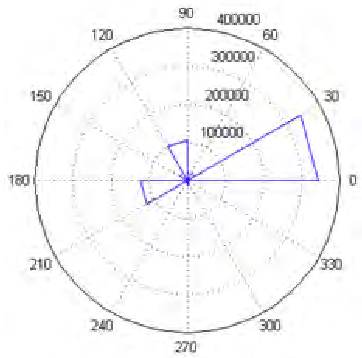

Figure 20. Coastguard GOF

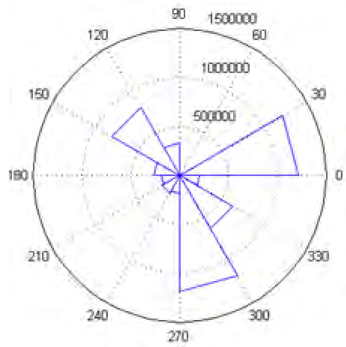

Figure 21. Husky GOF

For the motion tracking we have created two sequences named as CAR and TWO-CAR which has got single car and two cars moving with a still background. We have taken 64 frame sequences. We have shown two frames from each sequence and we can observe the motion of the cars being tracked from Fig. 22 to Fig. 25. The background has also been subtracted from them.

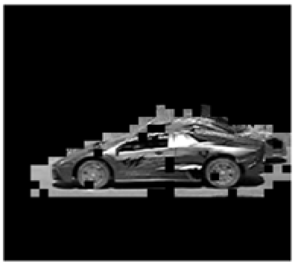

Figure 22. TWO-CAR frame 14

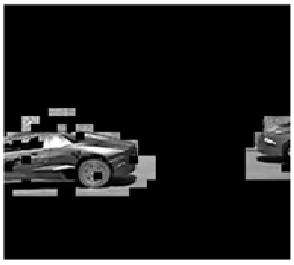

Figure 23. TWO-CAR frame 29

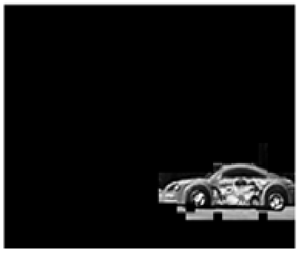

Figure 24. CAR frame 21

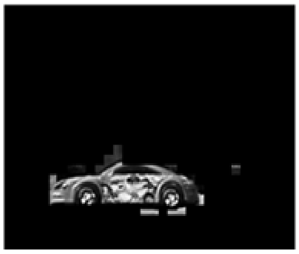

Figure 25. CAR frame 57

In all the applications, which have been discussed, the block motion vectors or the optical flow vectors have played very important role. We conducted a simple experiment where a cameraman image was shifted by [-2, -2] and a *64*64*8* sequence was formed. We computed the motion vectors using the block matching method with different search methods, optical flow, TSPREF and time taken by the respective methods were noted. Table IV shows the time taken to calculate the flow vectors for all the frames in the video by different methods. We clearly observe that TSPREF calculates the flow vector faster than other methods. TSPREF can be compared to the block motion vectors as flow vectors are function of time only. But the advantage lies in the fact that energy minimization is carried out on a video cube unlike the block matching or optical flow methods. The additional advantage of TSPREF over the optical flow is that when the spatiotemporal gradients are small, the optical flow estimate is not accurate.

TABLE II. CAMERA MOTION PATTERN RECOGNITION

Motion Pattern	Bin location in angle histogram (Theoretical)	Magnitude histogram (Theoretical and practical)	Sequence	Bin location (practical)
Horizontal left	sector 2 or 3	No peak	Coastguard (consecutive	sector 2
Horizontal right	sector 5 or 6	No peak	Coastguard (consecutive)	sector 5
Vertical up	sector 1	No peak	Coastguard (consecutive)	sector 1
Vertical down	sector 4	No peak	Husky (consecutive)	sector 4
Zoom	More than 1 sector	No peak	Stefan (consecutive)	sector 1 and 2
Vertical up	sector 2	peak	Tabletennis (consecutive)	sector 2
Left, Up, Right			Coastguard (GOF)	sector 2,1,5
Left, Down, Right			Husky (GOF)	sector 2,4,6

TABLE III. RUN TIME OF CAMERA MOTION PATTERN RECOGNITION BETWEEN 2 FRAMES

Video Sequence	Resolution	Time taken in sec
Stefan	240*320	1.5584
Flower garden	480*640	4.9491
Coastguard	144*176	1.0341
Husky	288*352	3.1351

However, in TSPREF since we are taking a group of frames this problem is avoidable.

TABLE IV. TIME TAKEN BY DIFFERENT METHODS TO CALCULATE MOTION VECTORS

Method	Time taken for all frames in sec
Block motion vector by exhaustive search	1.8
Block motion vector by diamond search	1.2
Optical flow by Horn-Schunk method	0.6
TSPREF	0.25

VII. CONCLUSION

In this paper we described a new approach using the concept of SPREF to handle different problems like video stabilization, recognition of camera motion patterns and motion tracking with subtracted background respectively.

Specifically we used TSPREF vectors in each of the above cases and obtained encouraging results. This work is currently being extended to take into account affine motion using ASPREF vectors.

REFERENCES

[1] T. H. Tsai, C. L. Fang, and H. M. Chuang, "Design and implementation of efficient video stabilization engine using maximum a posteriori estimation and motion energy smoothing approach," *IEEE Transactions Circuits and Systems for Video Technology*, vol. 22, no. 6, pp. 817-830, June 2012.

[2] H. H. Chen, C. K. Liang, Y. C. Peng, and H. A. Chang, "Integration of digital stabilizer with video codec for digital video cameras," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 7, pp. 801-813, July 2007.

[3] H. C. Chang, S. H. Lai, and K. R. Lu, "A robust real-time video stabilization algorithm," *Journal of Visual Communication and Image Representation*, vol. 17, no. 3, pp. 659-673, Mar. 2006.

[4] O. Alatas, P. Yan, and M. Shah, "Spatio-temporal regularity flow (SPREF): Its estimation and applications," *IEEE Transactions Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 584-589, May 2007.

[5] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H. Y. Shum, "Full-frame video stabilization with motion inpainting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1150-1163, July 2007.

[6] C. W. Ngo, T. C. Pong, and H. J. Zhang, "Motion analysis and segmentation through spatio-temporal slices processing," *IEEE Transactions on Image Processing*, vol. 12, no. 3, pp. 341-355, Mar. 2003.

[7] S. W. Chen, L. K. Wang, and J. H. Lan, "Moving object tracking based on background subtraction combined temporal difference," in *Proc. International Conference on Emerging Trends in Computer and Image Processing*, Bangkok, 2011.

[8] S. Saravanakumar, A. Vadivel, and C. S. Ahmed, "Multiple human object tracking using background subtraction and shadow removal techniques," in *Proc. International Conference on Signal and Image Processing*, 2010, pp. 79-84.

[9] A. Liu, W. S. Lin, M. Paul, F. Zhang, and C. Deng, "Optimal compression plane for efficient video coding," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2788-2799, Oct. 2011.

Karthik Dinesh received the B.E degree in Electronics and Communication from The National Institute of Engineering, Mysore, India and M.Tech. degree in Electrical Engineering from The Indian Institute of Technology, Kanpur, India. His current research interests are in video signal processing, image processing and DSP.

Sumana Gupta (M'90) received the B.E. and M.E. degrees in electrical communication engineering from Indian Institute of Science, Bangalore, India, and the Ph.D. degree in electrical engineering from the Imperial College of Science and Technology, London, U.K. She is currently a Professor in the Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India. Her current research interests are in video compression, video restoration, and image analysis.