International Journal of Computer Vision and Signal Processing, 2(1), 8-16(2013)

A Real-time Scheme of Video Stabilization for Mobile Surveillance Robot

Saksham Keshri*

National Institute of Technology Karnataka, Surathkal

S.N. Omkar

Dept. of Aerospace Engineering,
IISc, Bangalore

Amarjot Singh

Research Assistant,
NUS, Singapore

Vinay Jeengar, Maneesh Kumar Yadav

National Institute of Technology Karnataka, Surathkal

Abstract

The purpose of this research is to develop a mobile surveillance robot capable of capturing and transmitting video on rough terrains. Recorded video is affected by jitters resulting into significant error between the desired and captured video flow. Image registration with a contrario RANSAC variant has been used to minimize the error between present and desired output video as it has proved to be a fast algorithm for video stabilization as compared to the conventional stabilization methods. This is the first paper which makes use of this method to design mobile wireless robot for surveillance applications. The video captured by the robot is stabilized and transmitted to the controller in the control room. Once the video is stabilized the controller moves the objects from one place to another with the help of robotic arm mounted to the robot using a wireless transmitter and receiver. The surveillance capabilities of the system are also tested in low illumination situations as spying in dark is an important requirement of todays advanced surveillance systems.

Keywords: RANSAC, Image registration, Video stabilization.

© *2012, IJCVSP, CNSER. All Rights Reserved*

IJCVSP
International Journal of Computer Vision and Signal Processing

ISSN: 2186-1390 (Online)
http://www.ijcvsp.com

Article History:
Received: 1 July 2012
Revised: 1 September 2012
Accepted: 20 September 2012
Published Online: 25 September 2012

1. INTRODUCTION

The development of intelligent systems based on mobile sensors has been aroused by the increasing need for automated surveillance of environments such as airports [1], warehouse [2], production plants [3] etc. The latent of surveillance systems is remarkably amplified by the use to communicate in different environments with humans or with other robots for complex cooperative actions [4] to active surveillance. Mobile robots are still in their initial stage of progress and are developed according to the needs of the requirements unlike accustomed non-mobile surveillance devices.

A number of mobile security platforms have been introduced in the past for multiple applications. Mobile Detection Assessment and Response System (MDARS) [2] is a multi-robot system used to inspect warehouses and storage sites, identifying anomalous situations, such as fire, detecting intruders, and finally determining the status of inventoried objects using specified RF transponders. The Airport Night Surveillance Expert Robot (ANSER) [1]an Unmanned Ground Vehicle (UGV) makes use of non-differential GPS unit for night patrols in civilian airports communicating with a fixed supervision station under control of a hu-

*Corresponding author
Email addresses: sakfor09@gmail.com (Saksham Keshri), omkar@aero.iisc.ernet.in (Dr. S.N. Omkar), amarjotsingh@ieee.org (Amarjot Singh), vnayjeengar8@gmail.com (Vinay Jeengar), maneeshyadav01@gmail.com (Maneesh Kumar Yadav)

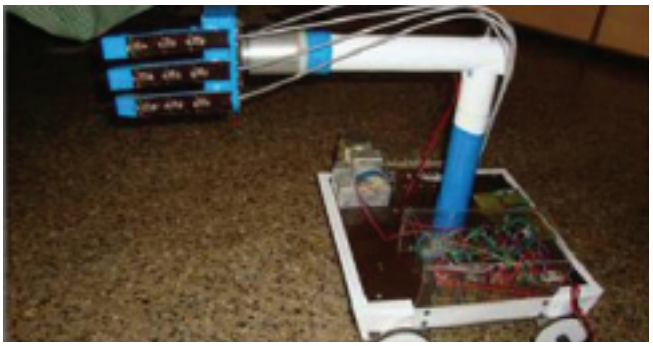


Figure 1: Camera on top of robot

man operator. At Learning Systems Laboratory of AASS, the rivet of a research project has been a Robot Security Guard [3] for remote surveillance of indoor environments aimed at developing a mobile robot platform. The system is able to patrol a given environment, acquire and update maps, keep watch over valuable objects, recognize people, discriminate intruders from known persons, and provide remote human operators with a detailed sensory analysis. The autonomous robots also have been used for delivering crucial information on the different spects like location of fissure [5] target locations after dynamite explosions [6] and accessing specific areas where survival is hard.

Automated surveillance systems find great application in defense based applications [6][7] Mobile robots have been used to stream video of intruders working in far environments useful to keep track of their activities. It has been observed that camera mounted on the robot produces jitters in video to be streamed due to the movement of the robot on uneven random terrains. Complications in the stabilized video streaming may also arise in rough terrains due to geography, obstructions like slopes, steps and turnings etc. These problems cause the camera to shake severely and induces disturbance in video flow. To get best information from the video streaming, a stabilized video is required.

Video stabilization aims to compensate the disturbing motions in video frames. It is an important technique employed in case of autonomous robot to reduce translational and rotational distortions. The method searches for the object in specified dimensions of frames and reduces the displacement by fixing the view on the object. The results into a new stabilized video sequence by removing the jitters from the original video flow. A number of methods have been proposed with the motivation to reduce the computation complexity and to improve the accuracy of the motion estimation. The two approaches been used for motion estimation include: feature based approaches [8][9][10] or pixel based approaches.[11][12][13][14] A feature tracking approach based on optical flow was presented by Chang et al. [9] considering a fixed grid of points in the video. But this approach was particular for motion model. The direct pixel based approach measure the contribution of every pixel in the video frame, making optimal use of the information available in motion estimation and image alignment. The motion between video frames as a global affine transform and parameters, estimated by hierarchical differential motion techniques was modeled by Hany Farid and J.B. Woodward in 1997. [14] To this stabilized video, temporal mean and median filters were applied for enhancing the video quality. Olivier Adda, et al [12] in 2003 suggested the use of hierarchical motion estimation with gradient descent search for converging the parameters. But the method was slow and complex to be used for real time applications. Image based rendering technique was proposed by Buehler et al. [15] to stabilize video sequence in which non-metric algorithm was used for camera motion and image-based rendering was applied to smoothed camera motion. However, this algorithm performs well only with simple and slow camera motion videos and so was not useful in application to mobile surveillance robot. Erturk [16] applied the Kalman filter to remove short-term image fluctuations with retained smooth gross movements. The Kalman filter based on constant velocity model was used to estimate intentional motion and was applied for the real-time video stabilization for the rescue robot. [17] This method produced very accurate results in most of the cases, but it required tuning of camera motion model parameters to match with the type of camera motion in the video and also proved to be computationally intensive. In 2006, Matsushita et al. [13] proposed an improved method called Motion inpaintng for reconstructing undefined regions and to smooth the camera motion Gaussian kernel filtering was used. Though this method had good results in most of the cases, but its performance relies on the accuracy of global motion estimation.

This SIFT feature has been proved to be affine invariant, is used to remove the intentional camera motions. Further, this feature-based approach is faster than global intensity alignment approach. A mobile surveillance robot is designed which can be used to record information from different environments. The robot is used to capture and transmit the video in three different environments to the control center. We performed the experiment in three different terrains with small objects placed around as target. Real time scheme is used to stabilize the undesired fluctuation in position of the surveillance video due to jitters produced robots motion on rough terrain. In the first step, SIFT algorithm [8] is applied in each pair of consecutive frames to extract the feature points. In the next step, RANSAC [18] algorithm is applied to estimate the best perspective transform for the pair of images from the matched feature points. The perspective transformation is calculated from randomly chosen 8 random correspondences from the matched feature points using the RANSAC process. We then apply the RANSAC process for certain random number of times, say 2000 times and evaluate the result by applying it to the matched pairs. The best approximation is the transformation that is correct for largest number of matched pairs. This is an iterative step which

results into matched pair of frames.The one which was correct for largest number of matched pairs was regarded as the best approximation. The motivation to use this method is the high computational speed and efficiency as compared to the conventional video stabilization methodologies especially for smaller targets. This paper first time makes use of this method to design a mobile wireless robot for surveillance applications. Once the video is stabilized, the controller uses the wireless robotic arm to pick and place objects from one place to another.

The paper is further divided into following sections. Section 2 explains the video stabilization algorithm while section 3 focuses on the design of the robot used in the paper. Section **??** elaborates the results and finally section **??** presents a brief summary of the paper.

2. VIDEO STABILIZATION ALGORITHM

The paper makes use of feature-based image registration method for video stabilization. Image features or land marks are determined using local image information in order to recognize slowly varying geometric differences between images. The method makes use of SIFT [8] algorithm to extract unique features, landmarks, from the images. Ransac method is further used to perform image registration [18] in order to match features between two frames. The method is briefly described below:

2.1. SIFT Descriptors

The algorithm given by David G. Lowe [8] is used to extract prominent features at consecutive frames of the video clip from video frames. In this algorithm, a difference of Gaussians operator is applied to input image at multiple scales using equation:

$$D_\sigma(x,y) = (G_{k\sigma}(x,y) - G_\sigma(x,y)) * I(x,y) \qquad (1)$$

Where is a circular 2D Gaussian filter with standard deviation. Sub pixel is determined by local extrema. An orientation depending on the local gradient information and a scale depending on are assigned to each interest point. SIFT descriptors are extracted from a neighborhood around each interest point, where the neighborhood is defined by the scale and orientation of the interest point. The descriptors (128 values for each interest point) provide a coarse characterization of the gradient orientations in the neighborhood.

2.2. Tentative Correspondences

To make tentative correspondence SIFT descriptors from one frame are matched with the SIFT descriptors with another frame. The Euclidean distance between the 128-dimensional descriptors is calculated. The two descriptors are declared as matched case if the ratio of the distance from the key point to the closet neighbor to the distance from the key point to the second-closet neighbor 0.8. This algorithm is however, a brut force techniqu whose complexity is, where represents the number of features in image. This brute force approach can be replaced with more efficient indexing schemes such as the approximate nearest neighbor method developed by Beis and Lowe [19]. In addition, the spatial locations of the SIFT descriptors can be utilized to lessen the number of potential matches to be evaluated given an initial approximate registration function. (For example, as recovered from ephemeris and pointing information).

2.3. RANSAC

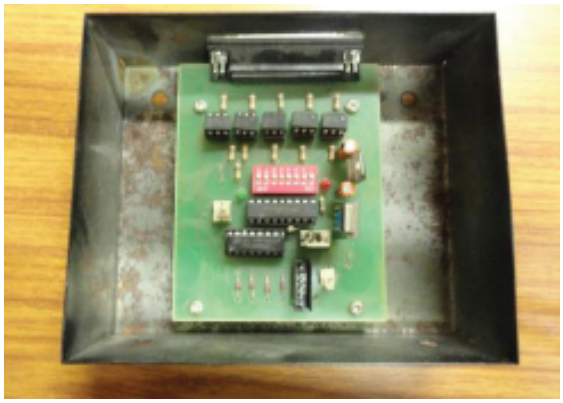
Once the tentative correspondence is computed, geometrical transformation parameters from one image to the other are estimated. False correspondences may remain even after SIFT descriptors has largely reduced the number of false matches. In such cases where outliers, are present, RANSAC is used to estimate transformation parameters [18]. Outliers are basically the data whose distribution cannot be explained by model parameters. A minimal set of correspondences among the tentative correspondences is randomly chosen. The main idea is to use this set to estimate transformation parameters and then determine if the estimated transformation works for the entire set of points. Points (or correspondences) for which the estimated transformation works well are labeled as inliers. The transformation can be re-estimated using all the available data after the inliers have been determined. This usually improves the results over the transformation estimated with only a minimal set. We use minimal set consisting of points one, three, four and seven respectively for transformation classes including translation, affine, homographic and fundamental matrices. To ensure at least one of the randomly selected minimal sets comprise of valid correspondences, the process of random selection of a minimal set is repeated several times. The number of trails, ntrials, depends on the size of the minimal set k, the fraction of tentative correspondences that are truly valid, and the probability (set by the user)describing the possibility that RANSAC finds at least one outliner-free minimal set.

$$n_{trails} \geq \frac{\log(1-p)}{\log(1-\alpha^k)} \qquad (2)$$

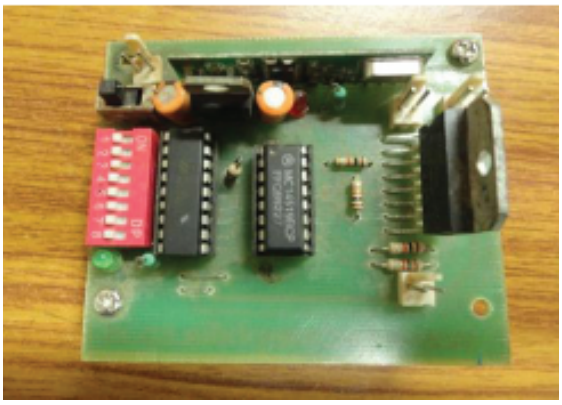
3. EVALUATION METRICS

To quantify the registration accuracy of the method, root-mean-squared color difference (RMSCD) between registered images is used [18].

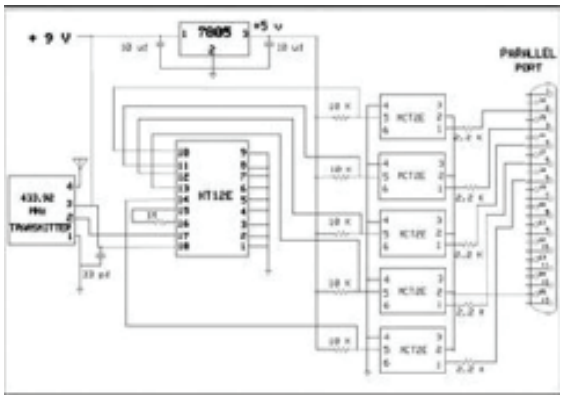
$$RMSCD_k = \Big\{ \frac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} (R_{k+1}[i,j] - R_k[i,j])^2 + (G_{k+1}[i,j] - G_k[i,j])^2 + (B_{k+1}[i,j] - B_k[i,j])^2 \Big\}^{\frac{1}{2}} \qquad (3)$$



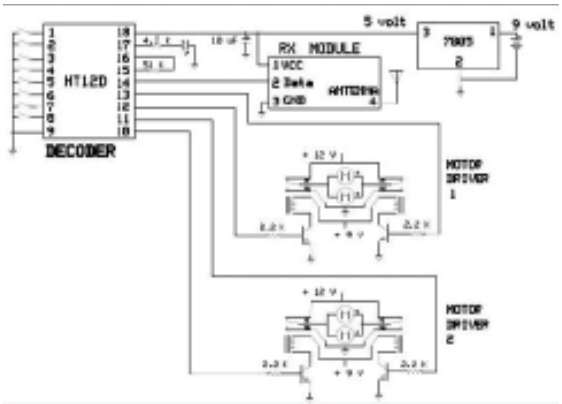
(a) Transmitter



(b) Receiver



(c) Circuit Diagram: Transmitter



(d) Circuit Diagram: Receiver

Figure 2

Here, Rk, Gk, Bk are the red, green and blue components of frame k, Vk. However, the root-mean-squared intensity difference between the images is used to quantify the registration accuracy, if the images to be registered are in gray-scale. In general, smaller RMSCD or the registration is more accurate when the area covered by the moving targets is much smaller than the background area. Further, we denote Vk as the reference image and Vk+1 as the test image. It should be noted that Vk and Vk+1 need not to be consecutive frames, rather frame Vk+1 may appear even after Vk in the sequence. The standard deviation of the RMSCD over several registrations is calculated to ascertain the variability of the registration method. The registration will be more stable if the variability is smaller. So the behavior of a stable registration is predictable and is enviable. Now, the registered images are visually examined in order to determine the registration reliability. The one incorrectly registered is identified. The metric to quantify reliability is the number of correctly registered images over the number of registrations tried in a video. A reliability of 1 is required for successful tracking.

4. ROBOT DESIGN

The mobile surveillance robot developed has 16cm by 16cm chassis with a camera and robotic arm mounted on it. The robot is controlled using a PC controlled transmitter and receiver pair. The transmitter is connected to the personal computer while the receiver is placed on the robot as shown in Fig. 1. Once the target object is recognized, video is transmitted to the control room. The controller controls the robot using the transmitter. The signal is transmitted using a transmitter Fig. (1) HT12E encoder which encodes the signal to be transmitted to the receiver Fig. (1). The transmitter is connected to the PC using MCT2E integrated circuit chips. The information to be sent is encoded on a 12 bit signal sent over a serial channel. The 12 bit signal is a combination of 8 address bits and 4 data bits. The information sent by the transmitter is received by the HT 12Dreceiver. The information is analyzed as 8 address bits and 4 data bits. If the address bits of the receiver and the address bits sent by the transmitter are same then the data bits are further processed which leads to the movement of the receiver. The detailed circuit diagram of the transmitter as well as receiver is shown in Fig. 1 and Fig. 1.

Figure 3: Video taken on different terrain (a) Indoor smooth ground, (b) outdoor terrain with normal illumination, (c) Outdoor terrain with low illumination

5. EXPERIMENTAL RESULTS

The results obtained from the simulation enable us to evaluate the capability of the four wheeled surveillance robot developed to be used on random terrains. The capability of the system is tested on three different rough terrains as shown in fig. 3. The first terrain is of indoor smooth ground, the second is an outdoor terrain with normal illumination while the third is outdoor terrain with low illumination. The system is used at 30 frames per sec with size of a resolution of 640*480 pixels. The jitters produced in the video moving on random terrain are stabilized by image registration algorithm mentioned in the paper. In the first step, SIFT [8] algorithm is applied with scales between 1.5 and 2.5 to find the descriptors or landmarks. In the next step, RANSAC is used to ascertain the correspondence between these points with projective transformation [18] using a distance tolerance of 2 pixels. After alignment, the corresponding landmarks in the image fall within a distance of 2 pixels and the range of scale of detected landmarks is taken within the scale of 1.5 to 2.5 pixels. This is because points with the scale smaller than 1.5 are influenced by noise and points with scales larger than 2.5 are not well positioned. Fig. 1 shows the result of registration when applied to two consecutive frames with varied parameters.

Table 1: Registration Results

Video No.	Average RMSCD	Variability	Reliability
1	9.3	.99	.9
2	9.2	.95	.97
3	7.6	2.8	.95
Average	8.7	1.516	.97

The result obtained is summarized as shown in Table 1. The 1st column index represents the video to be analyzed, average RMSCD between registered images is represented

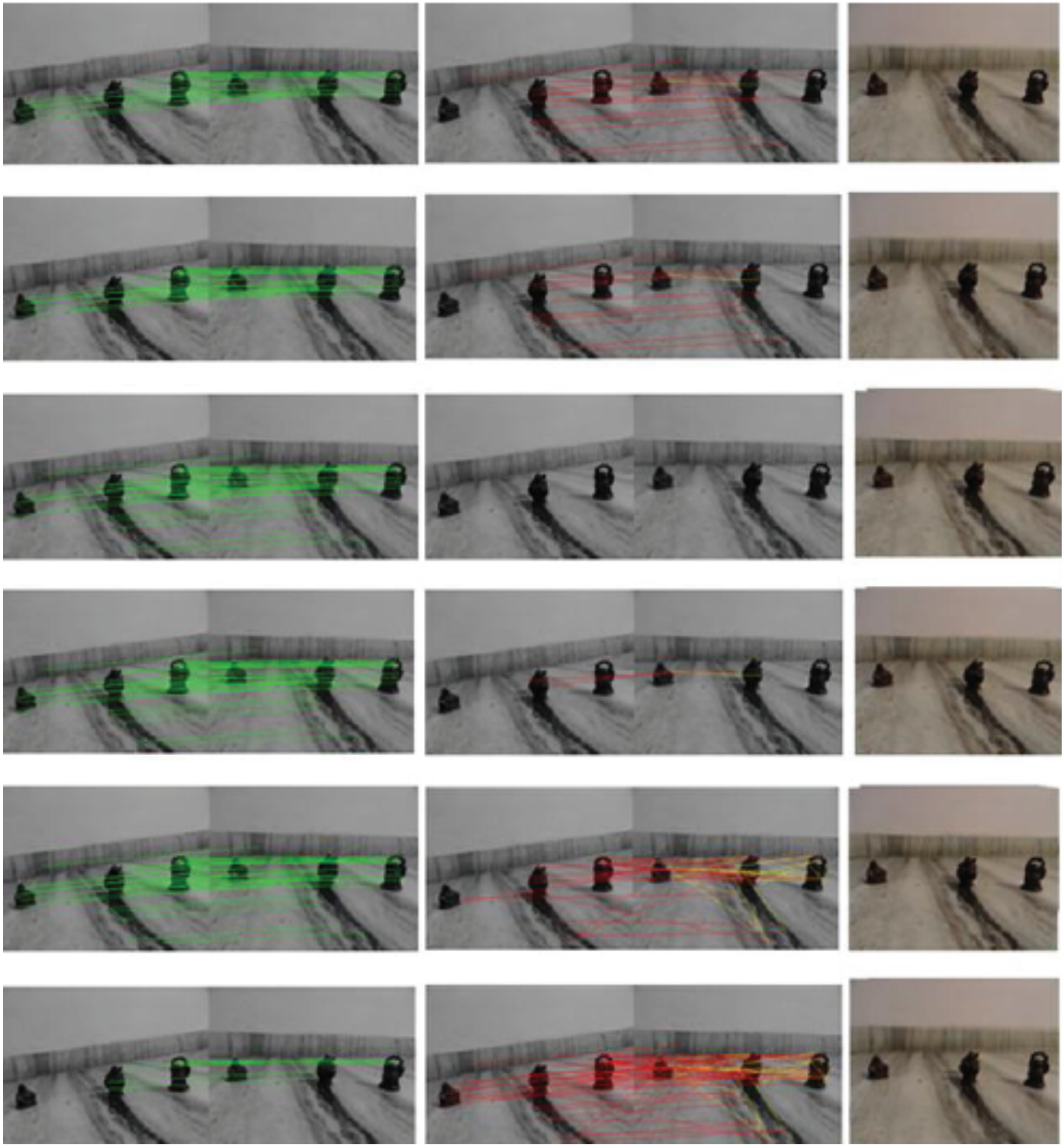


Figure 4: Registered images in case of input taken as well illuminated indoor image. (a): Precision= 1, Sift ratio= 0.8, (b): Precision= 2, Sift ratio= 0.8, (c): Precision= 5, Sift ratio= 0.6, (d): Precision= 5, Sift ratio= 0.8, (e): Precision= 5, Sift ratio= 1, (f): Precision= 0.5, Sift ratio= 1

by column 2nd and the standard deviation of the RMSCD which quantifies the variability of the method is shown by 3rd column. Further this variability will be small, if a method always produces the same RMSCD, however if a method produces varied RMSCD from one frame to the next, the variability will be higher. The variability needs to be low as the process is not stable in case of high variability hence registration cant be predicted. The reliability

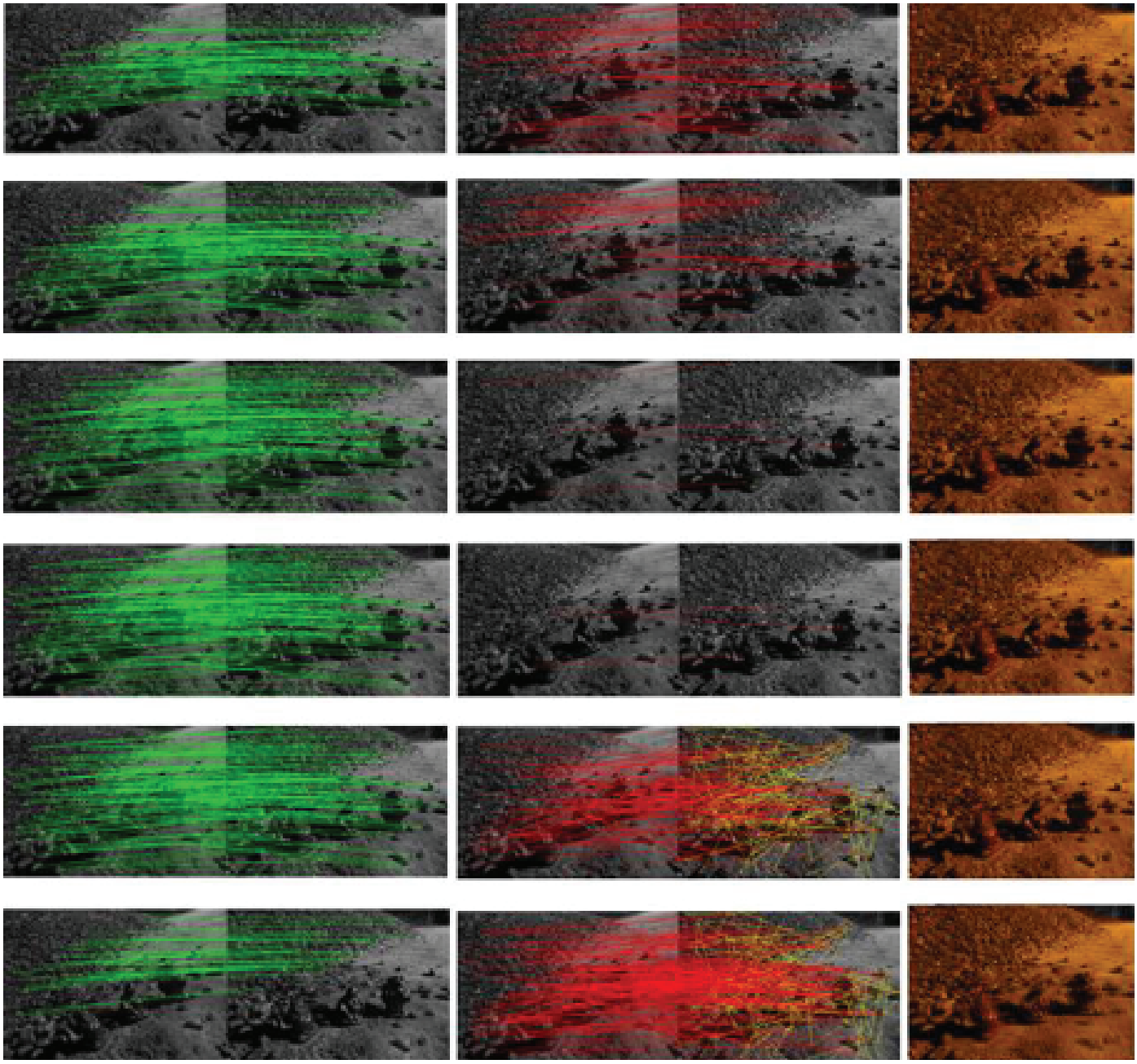


Figure 5: Registered images in case of input taken as well illuminated outdoor image. (a): Precision= 1, Sift ratio= 0.8, (b): Precision= 2, Sift ratio= 0.8, (c): Precision= 5, Sift ratio= 0.6, (d): Precision= 5, Sift ratio= 0.8, (e): Precision= 5, Sift ratio= 1, (f): Precision= 0.5, Sift ratio= 1

given by the ratio of correctly registered cases to the total number of registration attempt is shown by the column fourth. The reliability of the method is shown by this metric. Finally, the bottom row in the table summarizes the result by calculating the mean of each registration metric over the three videos captured. We performed the experiment over three terrain, the difficulty of the terrain was measured in terms of the energy consumed by the robot in traveling 10 m. The energy consumed in 1st, 2nd and 3rd video were respectively 1.5 Watt, 2.3 Watt and 3.5 Watt for 10 m which justifies the difficulty of the terrain in ascending order. In the first experiment, the robot was moved indoor with well illuminated surroundings and on smooth terrain as shown in Fig. 1. In this case, the average root-mean- squared color difference (RMSCD) was found to be 9.3 and standard deviation of RMSCD that is, variability resulted to .9. This low value is the proof for non-variability of RMSCD values of consecutive frames of video. The method turned with around cent percent success as the reliability was found to be .99. In the second

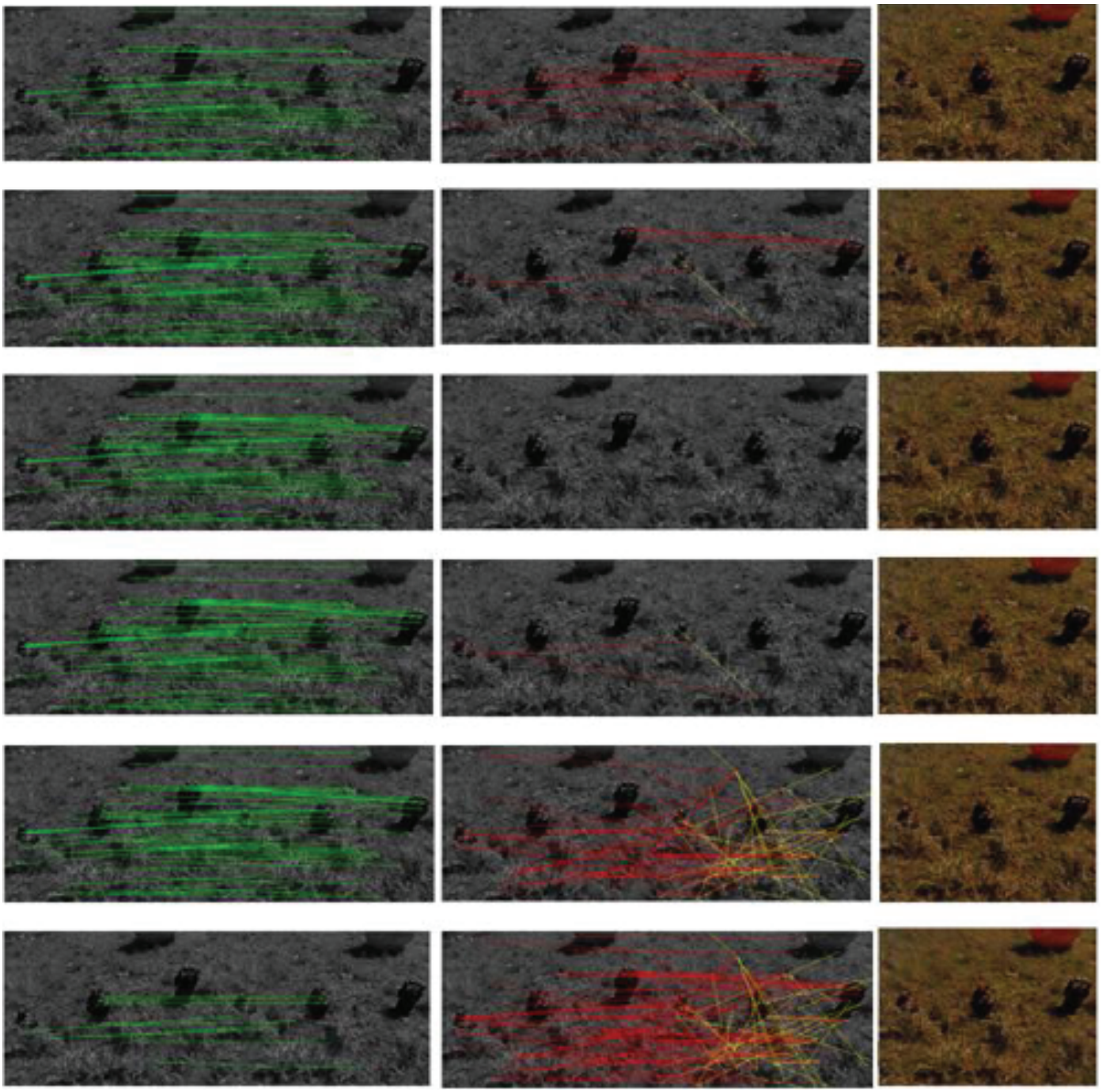


Figure 6: Registered images in case of input taken as poorly illuminated outdoor image. (a): Precision= 1, Sift ratio= 0.8, (b): Precision= 2, Sift ratio= 0.8, (c): Precision= 5, Sift ratio= 0.6, (d): Precision= 5, Sift ratio= 0.8, (e): Precision= 5, Sif t ratio= 1, (f): Precision= 0.5, Sift ratio= 1

experiment, the robot is moved on an outdoor terrain in a playground with good illumination as shown in Fig. 1. Avg. RMSCD and its standard deviation were respectively found to be 9.2 and .95 in this case. The success score for this video was 97In the last experiment, the performance of the system was tested on outdoor terrain with low illumination conditions as shown in Fig. 1. Because of the non- favorable condition, the Avg. RMSCD, Variability and Reliability were respectively calculated to 7.6, 2.8 and .95 respectively. To overview the result we calculated the average of these three values which were respectively found in three cases. The average of average RMSCD, average of variability and average of reliability was respectively calculated to 8.7 , 1.516 and .97 . Once the video is stabilized, it is transmitted to the controller in the control room. The controller further uses pick and place to move objects from

one place to another using the robotic arm connected to the robot. The robot is controlled manually by the controller using the transmitter receiver circuit. The transmitter is connected to the computer while the receiver is placed on the robot. The signal command 111100111111 was transmitted to the wireless robot directing to move forward. The command consisted of first 8 bits as address bits while the last 4 bits as data bits. Once the address bits were matched by the receiver HT12D, then only the command corresponding to the data bits is executed. Once the robot reached the target Buddha statutes as shown in Fig. 1 111100111110 commands were given by the controller to pick the object using the robotic arm. Buddha was moved from one place to other with the help of robotic arm.

6. CONCLUSIONS

The system developed works effectively on different terrains with varied illumination changes. The algorithm used for video stabilization has proved to be effective in reducing the jitters in video recorded at random terrains. The result is also found satisfactory, when a video is captured in low illumination. Jitters are decreased to a great extent by adopting the image registration solution. Once the video is stabilized the controller successfully moves objects from one place to another using the robotic arm fixed on the robot. The system can be effectively used for defense purposes to defuse bombs and in multiple other applications.

References

[1] F. Capezio, A. Sgorbissa, R. Zaccaria, GPS based localization for a surveillance UGV in outdoor areas, in: Proceedings of the Fifth International Workshop on Robot Motion and Control(RoMoCo'05), Dymaczewo, Poland, 2005, pp. 157–167.

[2] H. Everett, Gage, D. W, From laboratory to warehouse:security robots meet the real world, International Journal of Robots Research, Special Issue on Field and Service Robotics 18 (7) (1999) 760–768.

[3] T. Duckett, G. Cielniak, H. Andreasson, L. Jun, A. Lilienthal, P. Biber, T. Martinez, Robotic security guard- autonomous surveillance and remote perception, in: Proceedings of IEEE International Workshop on Safety, Security and Rescue Robotics, Bonn, Germeny, 2004.

[4] W. Burgard, Collaborative multi-robot exploration, in: Proc. IEEE International Conference on Robotics and Automotion (ICRA), 2000.

[5] D. Yorger, A. Bradley, B. B. Walden, M. Cormier, W. F. Ryan, Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot robotics and automotion, in: Proceedings, ICRA IEEE International Conference, Vol. 2, 2000, pp. 1787–1797.

[6] B. Bhanu, Automatic target recognition: state of ther art survey, IEEE Trans. Aerosp.Electron Syst. 22 (4) (1986) 364–379.

[7] J. A. Racher, C. P. Walters, R. G. Buser, B. D. Guenther, Aided and automatic target recognition based sensory inputs from image forming system, in: IEEE Trans. Patt. Anal. And Marc. Intell, Vol. 19, 1997, pp. 1004–1019.

[8] D. Lowe, Distincive features from scale-invariant key-points, Int'l J. Computer Vision 60 (2) (2004) 91–110.

[9] H. C. Chang, S. H. Lai, K. R. Lu, A robust and efficient video stabilization algorithm, in: ICME'04: International Conference on Multimedia and Expo, Vol. 1, 2004, pp. 29–32.

[10] R. Hul, R. Shil, I. fan Shenl, wenbin Chen2, Video stabilization using scale-invariant features, in: !!th International Conference Information Visualization (IV'07), IEEE, 2007.

[11] F. D. nad Janusz Konard, Efficient robust and fast global motion estimation for video coding, IEEE Transactions on Image Processing 9.

[12] O. Adda, N. Cottineau, M. karoura, A tool for global motion estimation and compensation for video processing, Concordia University, 2003.

[13] Y. Matsushita, E. Ofek, W. GE, X. Tang, H. Y. Shum, Full frame video stabilization with motion inpainting, IEEE Transactions on Pattern Analysis and Machine Intelligence (2006) 1163.

[14] H. Farid, J. B. Woodland, Video stabilization and enhancement, 1997.

[15] C. Buehler, M. Bosse, Leonard mcmillan mit laboratory for computer science cambrige, ma 02 139.

[16] S. Erturk, Digital image stabilization with sub-image phase correlation based global motion estimation, IEEE Trans. on Consumer Electronics 49 (4) (2003) 1320–1325.

[17] J.-P. Hsiaol, C.-C. Hsul, T.-C. Shihl, P.-L. Hsul, S.-S. Y. 2, B.-C. Wang3, The real-time video stabilization for the rescue robot, in: ICCAS-SICE, 2009.

[18] M. Fisher, R. Bolles, Random sample consensus: A paradigm for modeling fitting with applications to image analysis and automated cryptography., Comm. ACM 24 (6) (1981) 381–395.

[19] J. S. Beis, D. G. Lowe, Shape indexing using appropiate nearest neighbor search in high-dimentional spaces, in: IEEE Comp. Soc.Conf. on Computer vision and patter recognition, 1997, pp. 1000–1006.

[20] A. Telea, An image impaniting technique based on the fast marching method, J. Graphics Tools 9 (1) (2004) 23–34.

[21] A. Litvin, J. Konrad, W. Karl, Probabilistic video stabilization using kalman filtering and mosaicking, in: Prof. OF IS&T/SPIE Synposium on Electronic Imaging, Image and Video Comm., Vol. 1, 2003, pp. 663–674.