Available online at www.uofk.edu/uofkej

UofKEJ Vol. 3 Issue 2 pp 1- 11 (August 2013)

UNIVERSITY OF KHARTOUM ENGINEERING JOURNAL (UOFKEJ)

Automatic Matching of Close-Range Video Images Using Parameter Space Clustering

Gamal H. Seedahmed

Department of Surveying Engineering, Faculty of Engineering, University of Khartoum
Khartoum, Sudan (*E-mail: Gamal.Seedahmed@gmail.com*)

Abstract: Image matching, which amounts to the automatic establishment of the correspondences between two images or more, is a fundamental problem in digital photogrammetry. It has a large number of applications such as image mosaicing and 3D surface reconstruction from images. The contributions of this paper are two folds. First, it presents a robust strategy for point features selection. Second, it presents a novel method for automatic point features matching for the images that were extracted from a moving video camera. The proposed matching methodology uses point features as matching entities and parameter space clustering as a matching method. The basic idea underpinning the parameter space clustering methodology is to pair each data element belonging to two overlapping images, with all other data in each image, through a mathematical transformation. The results of pairing are encoded and exploited in histogram-like arrays as clusters of votes in the parameter space defined by the transformation function. Due to the nature of video images the mathematical transformation that defines the parametric relationship between the two images is approximated by a 2D translation. As a consequence of this approximation, the matching problem is approached as an inexact-matching. The maximum consistent subset of votes in the parameter space is exploited to reveal the underlying correspondences between the two images. Successful and promising experimental results of matching video images are reported in this paper.

Keywords: *Parameter space clustering; Image matching; Video images;3D Surface reconstruction; 2D Translation.*

1. INTRODUCTION

Video imagery analysis is a well established research topic [1]. In this paper, the utility of video imagery or sequence will be motivated from two angles. First, video imagery is a rich source of visual information. Second, video imagery can provide an inexpensive source of information about the world. Video sequence or images is a much richer source of visual information than still images. This is primarily due to the capture of motion and the small time interval and distance between the images; while a single or still image provides a snapshot of a scene, a sequence of images register the dynamics of the scene. Motion carries a lot of information about the spatio-temporal relationships between image objects. This information can be used in such applications as traffic monitoring, for example to identify objects entering/leaving the scene or objects that just moved. Beside their richness, video images can provide an inexpensive source of information about the world. And once again, for many applications such as surveillance, situation assessment, activity recognition, navigation, road condition assessment, pipeline investigation, and landmarks identification and mapping, the utility of video images is increased if we are able to derive value-added products such as mosaics, panoramas, and 3D surfaces reconstruction. For all types of these applications, matching of video images, which is an essential element for motion detection and estimation, is a critical task to facilitate these applications. It is very import to stress that the goal of video sequence matching is to estimate the motion parameters between the images in this sequence. Therefore, there is a need for a mathematical model to estimate the motion between the video images. There are two essential models for motion estimation between images from a video sequence, namely, spatial motion models and temporal motion models [2], which are beyond the scope of this paper. In classical photogrammetric terms, the motion between two images will give rise to the air-base in aerial photogrammetry or image-base or stereo-base in close range photogrammetry.

The goal of spatial motion models is to estimate the motion of image points, i.e., the 2D motion or apparent motion. Such motion is induced by a combination of projections of the motion of objects in a 3D scene and of 3D camera motion in terms of its exterior orientation parameters (3D translation and 3D rotation angles). Where as the camera motion has a global impact on the image points in terms of matching. The motion of the 3D objects only affects a subset of image points that correspond to objects' projection in the image space. In general, the spatial motion between adjacent video images or frames can be modeled by a translation vector:

$$v(p) = \begin{pmatrix} b_x \\ b_y \end{pmatrix} \quad (1)$$

where: v: a motion vector in 2D.
p: an image point.
bx: motion parameters in the x and y directions.

This 2D translational model shown in equation (1) has proven to be very powerful in practice since it provides a good approximation for the underlying motion between the images in a video sequence. More complex models have been proposed as well, depending on the application, they do not always improve the accuracy of the motion parameters. In general, the higher the number of motion parameters, the more accurate the estimation of the motion parameters. Restricted motion models such as the one shown in equation (1) may limit the image matching into a particular region or regions of the images and not covers them entirely. In other words, the motion model is not applicable over the whole image. These regions are typically called the "support regions" for matching, or more precisely, the regions on which the motion model is valid.

This paper presents an integrated approach for image matching that combines some of the critical aspects of point features selection. Although there are a plethora of research papers that address the matching and registration of video images [3], none of them took a holistic approach in terms of addressing the matching, the point features extraction, and the point features selection in one unified approach. In other words, the image matching should be viewed as an integrated process or a system; and the contribution of each element in this system should be well understood and optimized to achieve the overall objective of image matching. The work presented in this paper can be considered as a precursor step for a comprehensive methodology for matching of video sequence as well as still images.

This paper is organized as follows. The next section reviewed the point features extraction process and presents a modified approach for point features selection. Section four presents the underlying principle of image matching by parameter space clustering. Section five presents the workflow for the matching video images or sequence. Section six presents the results and analysis. The last section concludes the paper.

2. MATERIALS AND METHODS

2.1 Point Features Extraction

In this work Moravec Operator [4], which is a classical algorithm for point features extraction, is used to provide image points for the matching process. This algorithm labels image pixels that have high contrast as point features. Yes indeed, a contrast threshold or value needs to be set for point features labeling or selection. Stepwise, Moravec Operator works as follows:

- For each pixel or image point (p) form a window over a (2N+1) x (2N+1) neighborhood (see Fig 1, N=1, 2, 3,.. n.

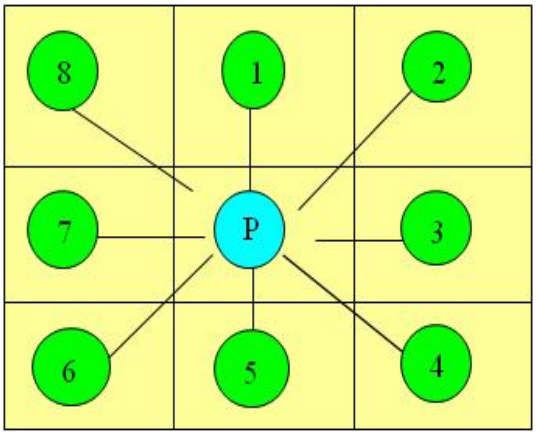

Fig. 1. An example of 8-points for an image point (p) neighborhood

- Compute the variances in the vertical, horizontal, and the two diagonals inside the window that was obtained in the previous step.
- Store the smallest variance with its associated image coordinates in a list (L).
- Repeat the previous steps for other image pixels.
- Sort the list (L) in a decreasing order.
- Set a threshold to classify the list (L) values.
- Use the classification result from the previous step to select the point features.

As mentioned, the original version of Moravec Operator is based on the computation of the variance of the intensity values in four different directions and within the neighborhood of an image pixel. The lowest variance is kept for further analysis by thresholding. Although the smallest variance is obtained from a specific direction in the image neighborhood, Moravec Operator can be considered as a non-directional filter since it does not use the directional information in any further analysis beyond the variance analysis and selection in the image neighborhood. From a photogrammetric and computational point of view, this operator lacks the following characteristics:

- It does not have an automatic capability for thresholding for point features labeling.
- It does not grantee a sufficient number of points or it may deliver a very large number of points, which may impact the computational time or complexity of image matching.
- It does not grantee a good distribution of point features over the image.

In light of the above shortcomings, the Moravec Operator is modified to satisfy the above requirements. In particular, the original version of the Moravec Operator is endowed with the following extra capabilities:

- A non-maximum-suppression procedure is added or adapted to the original version of Moravec Operator. The underlying spirit of this procedure is used in the

design of edge detection filters [5]. Its role in this research is to prevent image neighborhoods that have high variances or contrasts to contribute by more than one potential point feature candidate during the selection process. In other words, the non-maximum suppression minimizes the impact of local clustering of point features, which is undesirable feature along the value chain of obtaining usable information from the matching process such as the estimation of the relative orientation parameters between an image pair. As such, the non-maximum suppression should be regarded as a quality control mechanism. The working principle for this procedure is very simple. The maximum variance in an image neighborhood is kept and the rest are set zeros. The effect of the non-maximum-suppression procedure depends on the size of the image neighborhood in which the non-maximum-suppression is applied. A large image neighborhood will increase the computational time or complexity of the procedure; and a smaller one may limit its impact. Therefore, a balanced approach should be followed to select a practical size of an image neighborhood to deliver the promise of this procedure. In this research a size of 3 x 3 was used.

- It has a predetermined number of feature points (Np) to be delivered or requested from each image. This predetermination is very critical in terms of controlling the computational complexity of the matching method and even the correctness of the matches.

- The predetermined number of points (Np) is also acting as a symbolic threshold for the point selection or labeling as feature points. This is achieved by constructing a 1D histogram from the smallest variances and counting the frequencies of the highest bins in a back-order until their sum is equal-to or less-than the predetermined number of points. As such, the histogram is acting as a ranking mechanism for the information (here: refers to variances) that declare the coordinates of point features. In other words, the histogram schedules the priority for point selection and this is without the need for any direct sorting. The symbolic nature of this threshold freed the modified operator from setting a dependent threshold value. In other words, the threshold is become an image-independent value.

- A good distribution of points is ensured by dividing the image into four quadrants and let the modified operator to work over each quadrant independently. Each quadrant will deliver NP/4 points or less. In other words, the total number of the predetermined number of points will be extracted from the four quadrants. Indeed, this approach is equivalent to the setting of four different thresholds.

In light of the above modifications for the Moravec Operator, it can be said that a robust strategy for automatic point selection is developed during the course of this research.

2.2 Image Matching Using Parameter Space Clustering

The underlying principle of parameter space clustering was used by several researchers. For example, Stockman [6] developed an object recognition and localization approach via clustering. Seedahmed and Martucci [7,8] used a clustering approach for automatic registration of satellite images. The principle of parameter space cluster as related to this work can be explained by the following simulated example. Assume that we have two images (A and B). Image A has N points and image B has M points (see Fig 2). The information (here refers to points) between the two images are separated by translation values or motion along the x and y axis. As shown in Fig 2 the number of points in the two images does not have to be identical but some of them have to be shared between the two images. Mathematically, the translational motion between the two images can be expressed by:

$$x_T = x_{2i} - x_{1j} \quad (2)$$

$$y_T = y_{2i} - y_{1j} \quad (3)$$

where : xT: Translation along the x-axis.
yT: Translation along the y-axis.
x2i: x-coordinate that belongs to the second image.
x1j: x-coordinate that belongs to the first image.
y2i: y-coordinate that belongs to the second image.
y1j: y-coordinate that belongs to the first image.

In light of the parameter space clustering for image matching, the coordinates of the two images are compared, namely, subtracted from each other; and this is by using equations (2) and (3). This comparison has a combinatorial nature since each coordinates from the first image is compared with all other coordinates in the second image and the results of this comparison is encoded in a histogram-like structure (see Fig 3). The x-axis of this histogram represents the xT and the y-axis represents the yT. This process of comparison is repeated for all coordinates in the first image with the ones in the second image. The total number of comparison is M x N, where M is the number of points in the first image; and N is the number of points in the second image. As shown in Fig 3, the repeated or similar values of xT and yT will give rise to a peak. This peak is formed from a consistent subset of coordinates that belong to the two images. In other words, this subset of points that generates the peak are potential candidates for conjugate points in the classical sense of photogrammetry. More precisely, this peak can be understood from the following point of views:

- Perceptually, this peak is a placeholder for the matched points of the most consistent subset or structure between the two images. In particular, the loci of the peak in the parameter space should be found in order to extract or retrieve the matched point features.
- Statistically, this peak characterizes the highest relative frequency in the parameter space, which is the mode.
- Algebraically, this histogram-like structure tracks multiple solutions as cluster of votes and the most consistent one manifests itself in the peak.

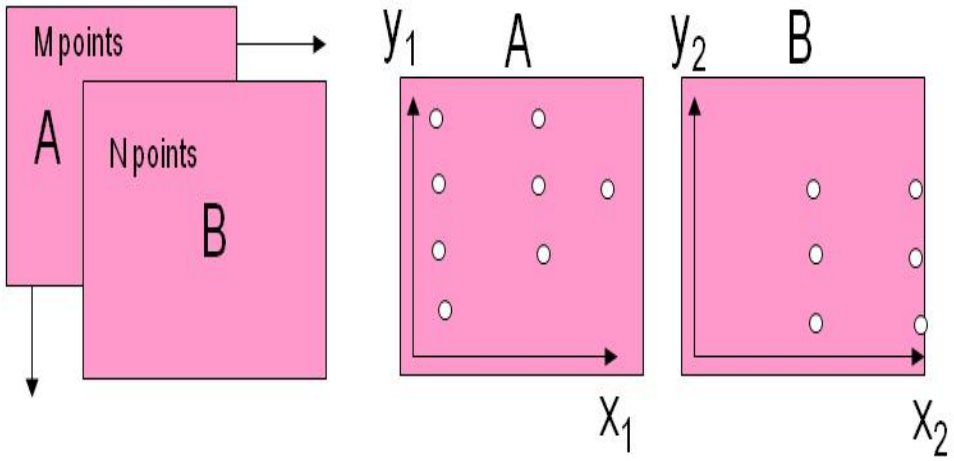

Fig. 2. Two images of the simulated example.

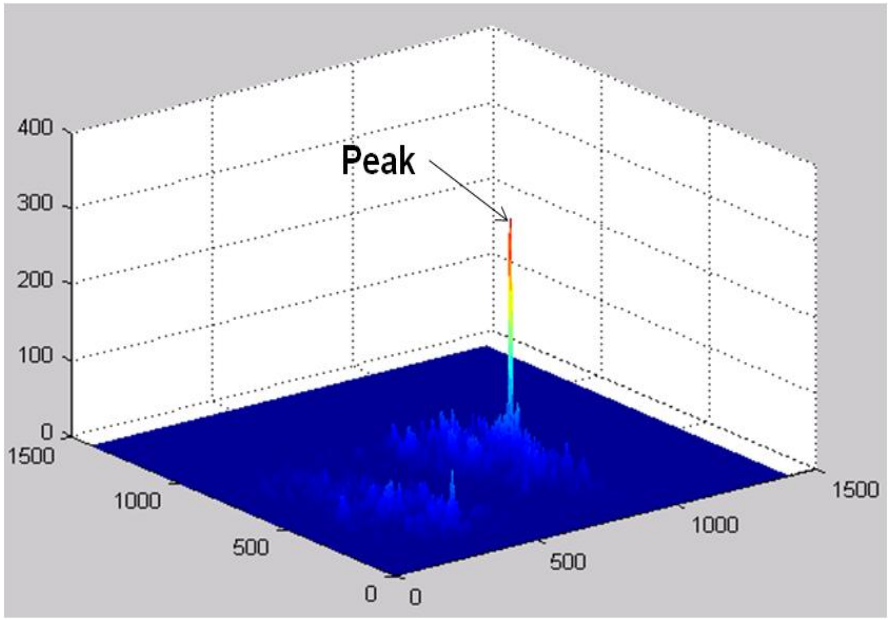

Fig. 3. A histogram-like structure for practical implementation of the parameter space clustering

2.3 The Workflow of Matching Video Images

The previous discussion paved the ground to present the steps of the proposed matching approach for video images. These steps can be summarized as follows:

- Sample a video sequence into still images.
- Extract point features from adjacent images in the video sequence using the modified Moravec Operator.
- Choose a cell size for the parameter space. Large sizes for the cell will allow us to realize the notion of inexact matching and it will increase the number of matches.
- Form the parameter space clustering.
- Identify the location of the peak in the parameter space.
- Find points that contribute to the formation of the peak. This step can be seen as a backtracking step.

3. RESULTS AND DISCUSSION

MATLAB-based prototype software was developed to implement the presented work in this paper. SONY DCR-SX85 digital video camera was used to collect the video sequences to test the proposed work. This video camera has a frame rate of 25 frames/seconds. In other words, 10 seconds of a video will generate 250 frames or images. In this research, the video sequences are sampled into still images every 5 frames. More precisely, the time interval between an image pair is 0.2 of a second. The size of the still image is 720 pixels x 576 pixels. Several experiments will be reported to test, to understand, and to demonstrate the critical elements of the point features selection and the parameter space clustering for image matching.

The first experiment demonstrates the full capabilities of the developed approach (see Table 1) over an image pair (see Fig 4) that was extracted from a video sequence. As shown in Table 1, the window size for Moravec Operator was set to 7 x 7 and this size was kept fixed for all experiments. The requested number of points that need to be extracted from the two images was set to 2,000 points per image. The extracted number of points from the first image is 1969 points and from the second image is 1974 points (see Fig 5). The non-maximum-suppression and points distribution over the 4 image quadrants are on. The cell size of the parameter space is set to 4. In other words, it is 4 times bigger than its original size. This size will allow more votes to populate the bins of the parameter space. The number of the matched points between the 2 images is 2099, which is greater than the number of the extracted points from either image (1969 and 1974 points).

Table 1. Specification of the first experiment

Specification	Value/Status
Windows size for the modified Moravec Opetator	7×7
Given number of predetermined point for the modified Moravec Operator	2000
Extracted number of point from the first image	1969
Extracted number of point from the second image	1974
Non-maximum suppression	Yes
Point distribution over the parameter space	Yes
Cell size for the parameter points	4
Number of matched point	2099
Execution time for matching	43 seconds

This large number of matched points can be explained by the fact of multiple matches and this is due to the large cell size or bin of the parameter space (here: cell size is 4 units). Therefore, this size of the cell of the parameter space renders the matching process as one-to-many but in the bound or the neighborhood of the cell size. Hence, this matching process can be viewed as an inexact-matching. Yes, indeed this inexact-matching will induce incorrect matches within the pixels neighborhood that will be defined by the cell size of the parameter space. On the other hand, these matches can be refined by other approaches such as area-based matching [9], which is beyond the scope of this paper. Fig 6 shows the parameter space of the matched points for the first experiment. A well defined peak is shown in Fig 6, which will be reflected in the quality of the match between the 2 images. Fig 7 shows the matched points between the 2 images, which is very

Fig. 4. An image pair for the first experiment

Fig. 5. Extracted point features from the image pair shown in Fig 4.

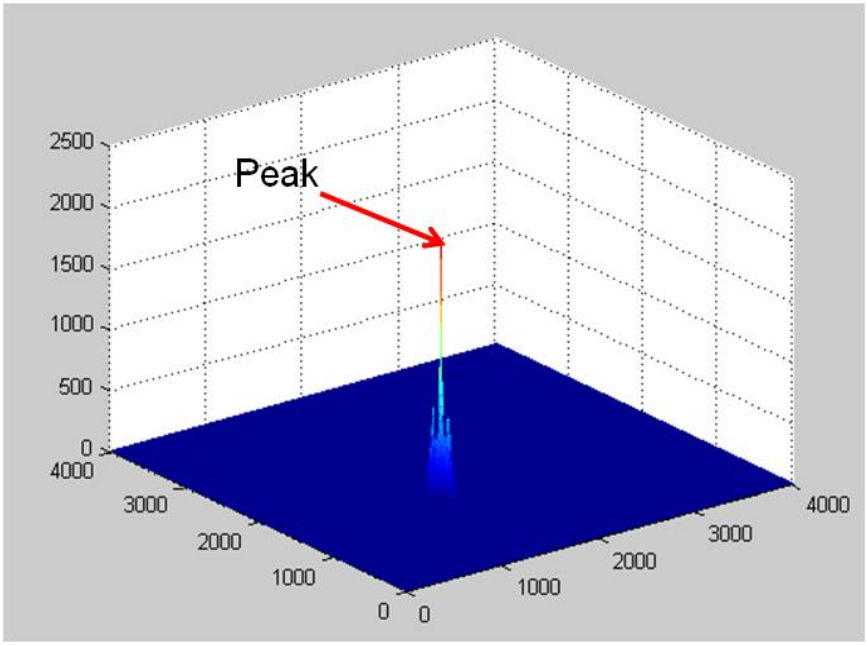

Fig. 6. The parameter space of the first experiment

satisfactory under the scope of this work. By comparing Fig 5 and 7, it is evident that the non-corresponding or conjugate points were not considered as matches (see yellow ellipses in Fig 5). The total execution of this experiment is 43 second. Now, the requested number of points that need to be extracted was set to 1,000 points from each image (see Fig 4). The total execution time for overall methodology went to 40 second. Then the requested number of points that need to be extracted was set to 4,000 points and the execution time went to 49 second. In view of this execution time, the algorithm is behaving very reasonably in terms of the computational complexity that will be induced by the number of points. Therefore, the requested number of points that need to be extracted should be application dependent. For example, small number of points can be extracted for the estimation of the relative orientation parameters. And on the contrary, more

points should be extracted for 3D surface reconstruction. By turning the non-maximum-suppression off and keeping the rest of the parameter as shown in Table 1, the execution time of the overall processes went to 13 second, which is dramatic reduction or improvement, but the number of matches is severely deteriorated (see Fig 8). Therefore, the gain from the non-maximum-suppression comes at a considerable amount of computational time, which is worth it. On the other hand, a considerable time saving can be gained and without turning the non-maximum-suppression off; and this is by executing the point features extraction and selection off-line.

Fig. 7. The matched point features for the first experiment

Fig. 8. Point features selection without non-maximum-suppression

Fig. 9. Point features extraction without 4-quadrants capability

By turning the 4-quadrants capability off and keeping the rest of the parameters as shown in Table 1, the execution time for the overall processes went from 43 second to 39 second. Therefore there is no considerable gain in the processing time. On the other hand, the quality of the extracted points was impacted. For example, the point that was outlined by the yellow ellipse in the right part of Fig 5 is disappeared from Fig 9. Although this point is not considered as a match as shown in Fig 7, the lack of good distribution will harm the overall value chain of the photogrammetric processes such as the estimation of the relative orientation parameters that requires a good distribution of matched points. Therefore, the existence of the 4-quadrants capability is very critical for the overall success of automated image matching.

By reducing the cell size of the parameter space from 4 to 2 and 1, the overall execution time went from 43 second to 69 second and 215 second respectively. And the number of the matched points went from 2099 points to 573 points and 176 points respectively. There is a dramatic increase in the processing time by reducing the cell size of the parameter and this is due to the increase in the search time for the maximum consistent subset or the peak in the parameter space.

The second experiment demonstrates the use of the proposed methodology on an image pair that was taken from a video sequence inside the Blue Nile Bridge (see Fig 10). The specification of this experiment is the same as the one shown in Table 1. Fig 11 shows the extracted points from the 2 images and Fig 12 shows the matched points between the 2 images. The yellow ellipses in Fig 12 outlined incorrect matches between the 2 images and this is because the speeds of the moving video camera and the minibus are not the same. On the other hand, there are good matches between the car in the 2 images as well as the structure of the bridge and this is for the following reasons. The structure of the bridge is not moving (zero speed) and the speed of the car and the moving camera is the same.

The last experiment demonstrates the use of the proposed approach of image matching on an image pair of a building (see Fig 13), which replicates a typical example of close range photogrammetric applications. The specifications of this experiment is the same as the one shown in Table 1 except that requested number of points that need to be extracted from each image was set to 4,000 points per image. Fig 14 shows the extracted points and Fig 15 shows the matched points. This example highlights an exciting possibility for using video sequence for fast, inexpensive, and automated capturing of 3D point clouds for 3D reconstruction of buildings and other structures.

4. CONCLUSIONS

Video images provide a rich and an inexpensive source of visual information. With the available computational power video images can be exploited automatically for 3D photogrammetric mapping, particularly, in close range applications.

Algorithmically, this paper presents a novel and holistic methodology for automated image matching that integrates the aspects of point features selection with the matching process in a unified approach. This integration reveals some of the hidden dependency between point features selection and matching. For example, the non-maximum-suppression comes at a high price of computational time or complexity but it can be offset by performing an off-line computation of the point features extraction and selection. At this stage, the developed approach is not meant to be a final solution for the image matching process. It can be viewed as a critical precursor step for developing a comprehensive framework for image matching.

Fig. 10. An image pair of the second experiment

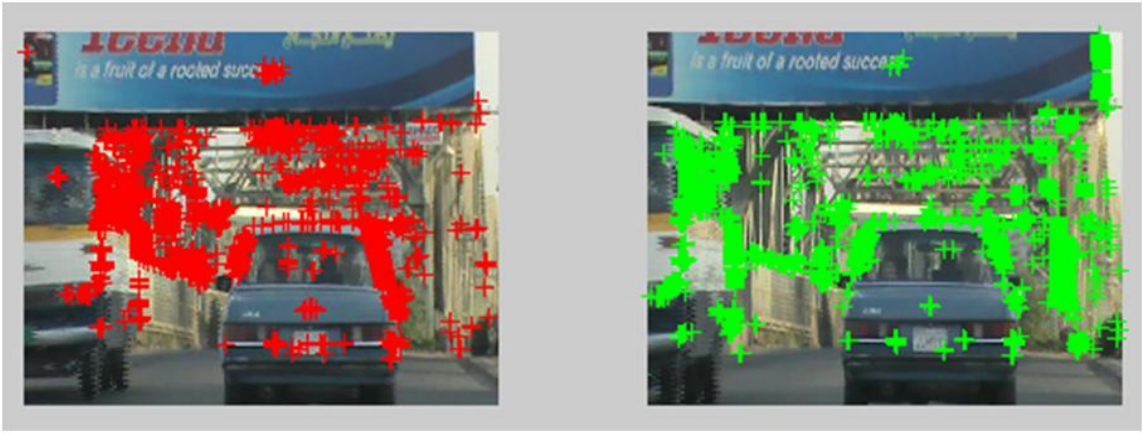

Fig. 11. Extracted point features for the second experiment

Fig. 12. Matched point features of the second experiment

Fig. 13. An image pair of a building

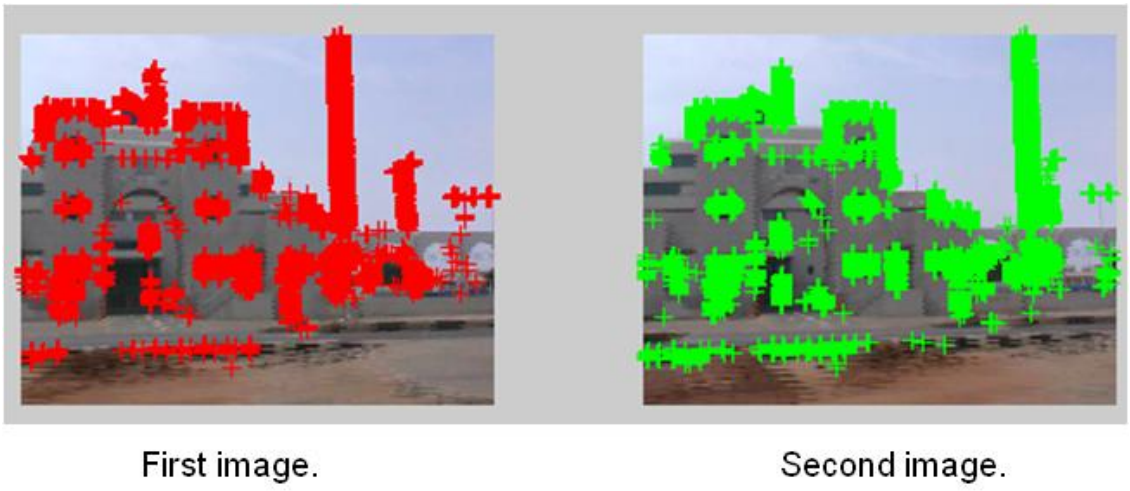

Fig. 14. Extracted point features from the image pair shown in Fig 13.

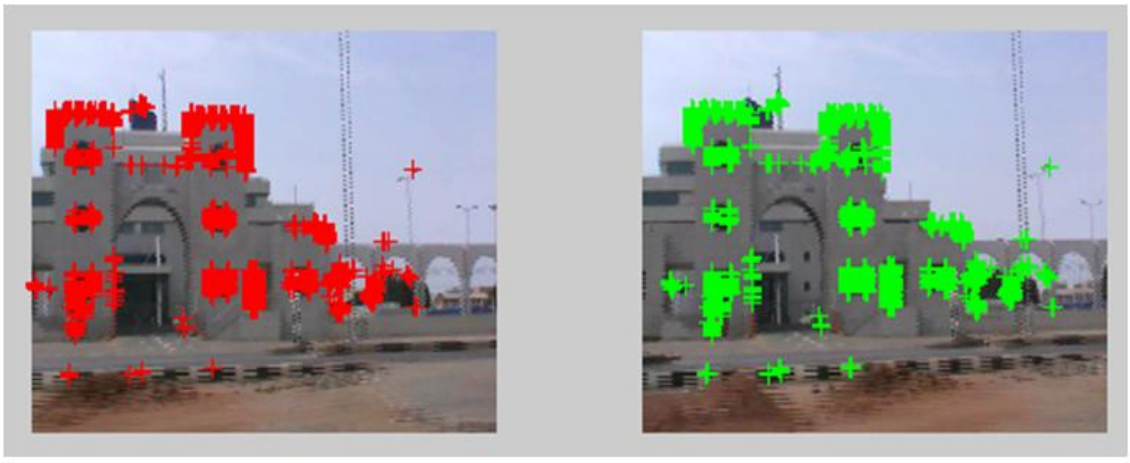

Fig. 15. Matched point features from the extracted points shown in Fig 14.

Indeed, this research can be extended in several directions such as subpixel point matching, automatic relative orientation of image pairs that will be obtained from video sequence, video-based photogrammetric triangulation, 3D surface reconstruction, and 3D reconstruction of human face.

REFERENCES

[1] Bovik, A. (Ed.), 2005. Handbook of Image and Video Processing. Elsevier Academic Press. 1372 pages.

[2] Trucco E. and A. Verri, 1998. Introductory Techniques for 3-D Computer Vision. Prentice Hall, Inc. Upper Saddle, NJ. 343 pages.

[3] Shah, M. and R. Kumar (Eds.), 2003. Video Registration. Kluwer Academic Press. 255 pages.

[4] Moravec, H., 1977. Towards Automatic Visual Obstacle Avoidance. Proc. 5th Int. Joint Conference on Artificial Intelligence, MIT, Cambridge, MA, USA, p. 584.

[5] Canny, J., 1989. A Computational Approach to Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679-698.

[6] Stockman, G., 1987. Object Recognition and Localization via pose clustering. Computer Vision, Graphics, and Image Processing. Volume (40):361-387.

[7] Seedahmed, G. and L. Martucci, 2002. Automated Image Registration Using Geometrically Invariant Parameter Space Clustering (GIPSC). In the Archives of the International Society of Photogrammetry and Remote Sensing, Commission III, Symposium 2002, Photogrammetric Computer Vision, September 9-13, 2002, Graz, Austria.

[8] Seedahmed G. and L. Martucci, 2004. Autonomous Conflation of Vector-to-Image. Proceedings of the First User Conference of Visual Learning Systems, Missoula, Montana September-15th-2004. 6 pages.

[9] Schenk, T., 1999. Digital Photogrammetry. Volume 1. Terra Science. 428 pages.