Evaluation of Stereo Matching Algorithms for Temporary Structure Monitoring

Youyi Feng and Fei Dai

West Virginia University, Department of Civil and Environmental Engineering, P.O. Box 6103, Morgantown, WV 26506-6103; PH (304) 293-9940; FAX (304) 293-7109; email: yofeng@mix.wvu.edu; fei.dai@mail.wvu.edu

ABSTRACT

Temporary structures such as formwork support and scaffolding plays an important role in on-site production for construction projects. However, they are normally disregarded when considering the importance of qualifications and experience. In fact, most of the time temporary structures won't be considered until it is actually constructed. As a result, most tragic failures during construction are usually the result of improperly designed, constructed and/or maintained temporary structures. A real-time monitoring system enabled by video cameras may facilitate the monitoring of integrity and safety of construction temporary structures. It entails applying close-range photogrammetry and digital image correlation to measure the deformation of structural material shapes. In this system, stereo image matching algorithms and accuracies are considered to be one of the most significant elements. It is to seek the corresponding relationships between left and right-view images of the same scene to obtain the disparity images. Thus, this paper attempts to evaluate the image matching algorithms that exist in order to identify the most appropriate one for the temporary structure monitoring system. The principle of stereo matching is firstly introduced. This is followed by the comparison of three classical algorithms, semi-global matching (SGM), block matching (BM), and graph cuts (GC) method. The undertaking of the comparison includes statistical analyses of the time and disparity images quality estimation. The research concludes that the SGM algorithm possesses both relative high efficiency and disparity images with excellent qualities. The present work facilitates the investigation of the aforementioned monitoring system.

INTRODUCTION

Temporary structures, such as formwork support and scaffolding, are widely used in on-site production for construction projects. However, the safety of these temporary frames is not usually getting enough attention to secure the structure and construction workers. In fact, most tragic failures during construction are usually the result of improperly designed, constructed and/or maintained temporary structures. This failure will undoubtedly induce tremendous loss, delay, injuries, and causalities as the consequence (Choi et al. 2013).

A surveillance monitoring system that can real-time monitor the deformation and then measure dynamic responses of load bearing members may facilitate the monitoring of integrity and safety of construction temporary structures. And this will compensate the deficient design and construction of temporary structures. To realize such a system, it entails applying image sensors and imaging processing technique such as digital image correlation to measure the deformation of structural material shapes. This poses the challenges that the measurement needs to achieve both high level of accuracy and efficiency.

In this system, stereo image matching algorithms are considered to be one of the most significant elements. It is to seek the corresponding relationships between left and right-view images of the same scene to obtain the disparity image that is set as input data for 3-D models reconstruction. Sequentially, displacement and deformation of the target scene could be calculated based on two 3-D models reconstructed before and after the deformation occurrence (Feng et al. 2010). In these processes, quality of the disparity images plays an extremely important role in determining the accuracy of the displacement and deformation result.

There are many algorithms that can be used to implement the stereo matching process; however, these stereo matching algorithms have diverse performances to generate disparity images and which algorithm is most appropriate for the temporary structure monitoring is unknown. The aim of this paper is to evaluate the matching algorithms that exist in order to identify the most appropriate one for the temporary structure monitoring system.

Three typical image matching algorithms, including block matching, semi-global matching and graph cuts matching, have been chosen and implemented using C++ and the image handling open library OpenCV. Then, experiments are designed and conducted to test the practical performances of these selected algorithms. In the analysis of the experiment results, both quality of disparity images and the algorithm efficiencies were taken into account to evaluate those matching algorithms.

BACKGROUND

Temporary structure failures. It is crucial to guarantee the safety and integrity of temporary structures during the process of construction. Otherwise, failure of temporary structures may occur and lead to severe consequences to the project. Table 1 provides a few examples with the injuries and fatalities due to temporary structure failures in recent years.

Table 1. Temporary structure collapse accidents.

Time	Location	Injuries/Fatalities	Accident depiction
2/22/06	Beijing, China	16 injuries, 3 deaths	Collapse during temporary structure demolition
5/20/06	Dalian, China	3 injuries, 6 deaths	Formwork collapse during concrete pouring
1/12/10	Anhui, China	7 injuries. 8 deaths	The collapse occurred during concrete pouring
3/13/10	Shenzhen, China	1 injured, 9 deaths	Scaffolding collapse accident
4/15/10	Chengdu, China	8 injuries, 3 deaths	A under-construction wall broke down
5/8/10	Guangzhou, China	4 injuries, 4 deaths	Formwork support collapse

To reduce the painful casualties induced by temporary structure collapse, current practice to ensure the safety of temporary structures mainly relies on regular manual inspection, for instance, engineers periodically check the compliance of structures with specifications, or by applying total station to achieve the goal (Chang and Hu 2009). These methods can achieve a relatively high accuracy. However, their limitation is also quite obvious because of the sudden nature of an accident. It cannot ensure the safety of the whole construction process. Therefore, our goal here is to build upon a surveillance monitoring system by applying digital cameras that can real-time and continuously measure the displacement and strains in the whole construction process to avoid the accident to a maximum extend.

Monitoring method: Digital image correlation (DIC). Digital image correlation method is the process that computes the pixel similarities between left-view and right- view images of one stereo pair to determine the corresponding image points or target areas. Then, after matching those corresponding image points or areas, the output will be the disparity map that contains the depth values of the scene. Based on the disparity map obtained by the matching process, a 3-D model of the scene can be generated through re-project the 2-D coordinates to 3-D dimensional space. As a result, the structure displacement and deformation can simply calculated from the two 3-D models of the structure before and after deformation (Vaananen et al. 2013).

In above whole process, the quality of the disparity map is considered to be the most decisive factor that greatly influences the accuracy of the calculation of displacement and deformation. Besides, the quality of disparity map is directly determined by the algorithms chosen to implement the matching process. Therefore, the next section of this paper will concentrate to analyze and compare those image matching algorithms (Hirschmuller and Scharstein 2009).

Classification of image matching methods. According to the different optimization theory, image matching algorithms can be divided into local constraint methods and global constraint methods.

Local constraint methods are based on the neighboring information around the interest points to calculate the matching energy cost. Normally, these local methods have a relatively lower computation complexity and higher operation efficiency because of the less information taken into account. Whereas, in general these local methods are sensitive to the noises in images and the matching quality is usually not good in those non-textured, discontinuity of disparity regions and occlusion areas.

Global constraint methods are on the information of whole image to calculate the matching energy cost. These methods can achieve the global optimal solutions and convert the process of seeking corresponding matching points to computing the global optimization of energy functions.

In this paper, a quite representative local method (block matching), a typical global method (graph cuts based matching) and another novel image matching algorithm called semi-global matching were chosen to conduct the comparative experiment.

SAD based blocking matching (BM). The basic objective of block matching is to get the corresponding disparity value of the pixels in stereo pairs. firstly, we set a

3 by 3 target window based on the central pixel to be processed; then, search in the right image sequentially to get the 3 by 3 reference windows and simultaneously compute the sum of the absolute difference (SAD) of target windows and reference windows; next, set the reference window which has the minimal SAD (matching metric function) as the matching window to the target window and the central pixel as the matching pixel. By repeating these steps, we can get all corresponding matching pixels in the right image to the target image. After the points matching process, a disparity map will generate as a processing result which contains each pair of corresponding points' disparity value. Some other matching metric functions are sum of the square difference (SSD), correlation coefficient function and so on (Wang and Zheng 2013).

Graph cuts (GC) based image matching. The basic thought of Graph cuts based image matching is converting the traditional image matching issue into a process to calculate the energy function values instead. Based on the corresponding energy function to set up the appropriate images and compute the minimal cuts (max-flow) which converts the image matching process to calculating corresponding minimal cuts. Then, energy cost minimization to implement the image matching process can be achieved by applying these minimal cuts (Chen et al. 2012). The following is the flowchart of the GC algorithm.

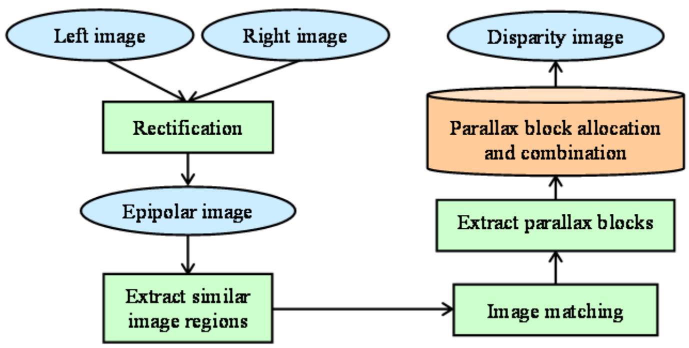

Figure 1. Graph-cuts flowchart.

Semi-Global Matching (SGM). A stereo algorithm uses a base image I_b and a match image I_m for calculating a disparity image D that corresponds to the base image. It is assumed that the epipolar geometry between the images is known. An epipolar line $e_{bm}(\mathrm{p}, d)$ in the match image is defined by the pixel p in the base image and the disparity d as line parameter. For rectified images $e_{bm}(\mathrm{p}, d) = [\mathrm{p}_x - d, \mathrm{p}_y]^T$. It is noteworthy that certain camera geometries (*e.g.*, push-broom cameras that do not move on a straight path) do not allow an exact rectification of the resulting images. Therefore, a general definition using arbitrarily defined epipolar lines is preferred.

The SGM method aims to determine the disparity image D, such that the global energy $E(D)$ is a minimum. The cost function can either be Birchfield and Tomasi's sampling insensitive intensity difference or Mutual Information (Hirschmuller 2008). The latter one has the advantage that it takes complex relations between corresponding intensities into account. This has been shown to be very robust against radiometric differences that often occur in practice. Finding the global minimum of $E(D)$ for the whole 2D image is known to be an NP-complete Problem.

SGM calculates $E(D)$ efficiently along 1D paths from either 8 or 16 directions towards each pixel. The cost to reach a pixel p at the disparity d from the direction r is defined. For each pixel p the disparity d is chosen that corresponds to the minimum cost. For sub-pixel estimation, a quadratic curve is fitted through the neighboring costs (*i.e.*, at the next higher and lower disparity) and the position of the minimum is calculated. The result is a disparity image D_b that corresponds to the base image I_b.

EXPERIMENT DESIGN AND IMPLEMENTATION

In this experiment, OpenCV and Visual Studio 2010 were used to implement the above mentioned BM, GC and SGM image matching algorithms. Following the implementation process, two sets of stereo image collections, namely standard and self-taken stereo pairs, were tested with the implemented algorithms. The output was the corresponding disparity images and compared against the ground truths available or collected.

Standard stereo pairs test. The input standard stereo pairs and ground truth images are from Middlebury College computer vision lab database (data source: http://vision.middlebury.edu/stereo/). The raw, resulting, and ground truth data are shown in Figure 2.

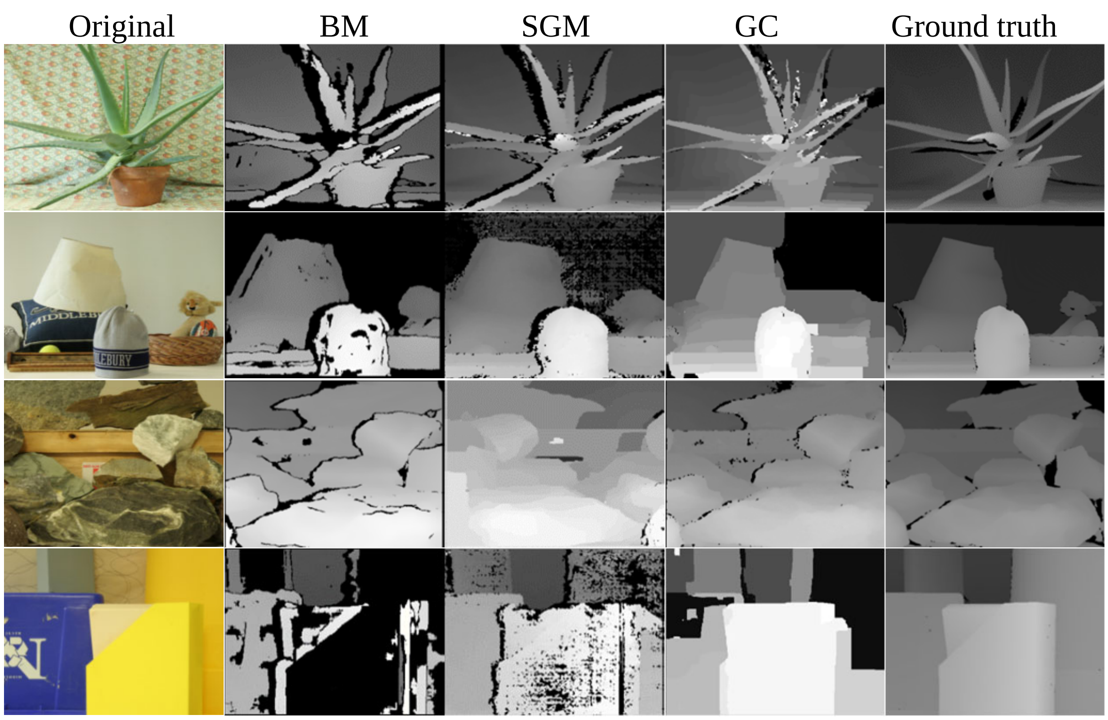

Figure 2. Standard pair testing results.

Experiment of civil structure scenes. In addition to testing on standard stereo image pairs, we also test the implemented algorithms on the images collected at the WVU concrete laboratory. The input stereo pairs were captured using a Canon 5D Mark III digital camera. The ground truths were generated by the time-of-flight camera MESA SR4000. Figure 3 shows the collected original, processed, and ground truth images.

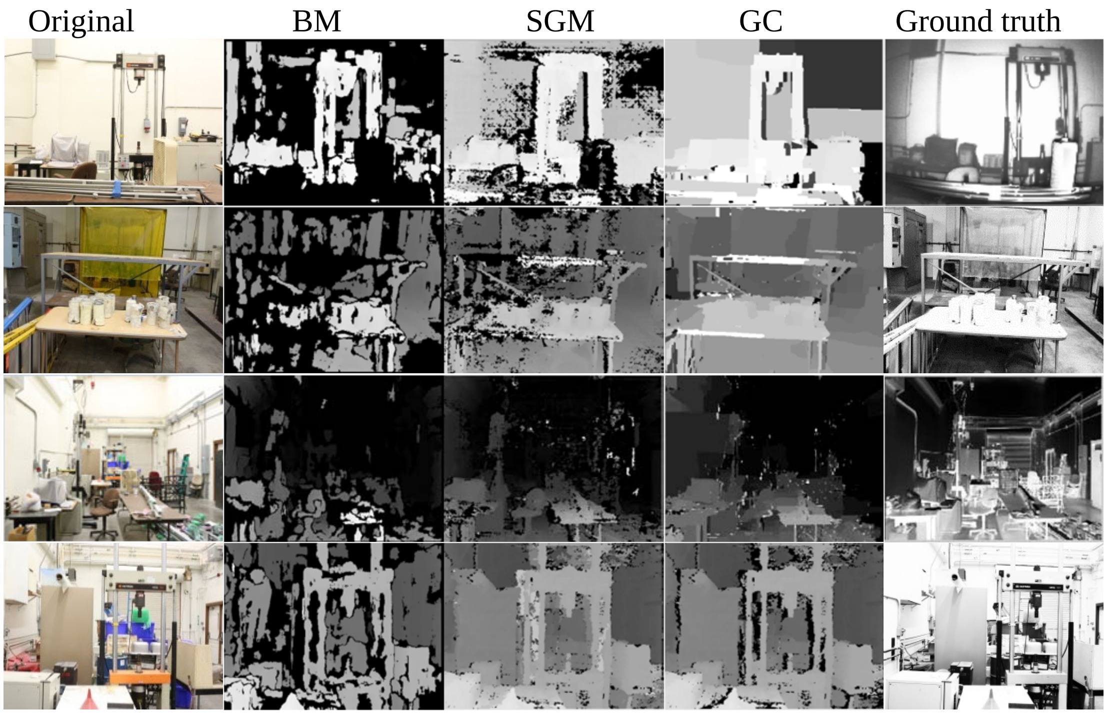

Figure 3. Concrete lab stereo pair testing result.

COMPARATIVE ANALYSIS AND DISCUSSION

Quality performance comparison. To compare the disparity images produced by different matching algorithms in terms of quality performance, two criteria have been generated in this experiment. The first one is the mean absolute percentage error (MAPE) and the other one is mean square error (MSE) as shown in Eq. 1 and Eq. 2.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(ResultImage_i - GroundTruth_i)^2 \quad (1)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{ResultImage_i - GroundImage_i}{GroudTruth_i}\right| \quad (2)$$

In Eqs. 1 and 2, n is the total pixel number of the images. MSE reflects gray level discrepancy. MAPE is a value of percentage. Table 2 shows the analysis result.

From the statistics of Table 2, GC matching algorithm has the minimum MSE and MAPE values. The SGM algorithm is in the second order. The BM algorithm has the largest values. More specifically, MSE(GC) < MSE(SGM) < MSE(BM), and MAPE(GC) < MAPE(SGM) < MAPE(BM). As a result, the quality of disparity images produced by the three algorithms is in the order of GC > SGM > BM.

Table 2. MSE and MAPE results.

Image	Aloe			Midd			Platic			Rocks		
Metric	**BM**	**SGM**	**GC**	**BM**	**SGM**	**GC**	**BM**	**SGM**	**GC**	**BM**	**SGM**	**GC**
MAPE	0.08	0.07	0.06	0.05	0.01	0.01	0.30	0.001	0.001	0.10	0.08	0.08
MSE	16.92	14.76	13.92	17.41	13.53	12.85	25.48	13.76	13.15	15.81	15.52	15.21
Image	**Scene1**			**Scene2**			**Scene3**			**Scene4**		
Metric	**BM**	**SGM**	**GC**	**BM**	**SGM**	**GC**	**BM**	**SGM**	**GC**	**BM**	**SGM**	**GC**
MAPE	0.088	0.02	0.01	0.14	0.07	0.05	0.17	0.14	0.12	0.02	0.01	0.01
MSE	13.74	8.46	6.95	18.69	13.36	10.38	25.11	23.59	19.29	13.67	10.92	9.33

Through comparing the MSE and MAPE values of SGM with the values of GC in Table 2, it is obvious that quality performance of SGM is quite close to that of GC. This means that the SGM algorithm can achieve a comparable performance with the GC algorithm. This can also be confirmed by comparing the resulting images of SGM and GC in Figure 2 and Figure 3.

Operating efficiency comparison. The algorithms' operating times are shown in Table 3. All the test images have a resolution 600 by 400 pixels.

Table 3. Algorithm operating times (ms).

	Image							
Algorithm	**Aloe**	**Rocks**	**Midd**	**Plastic**	**Scene1**	**Scene2**	**Scene3**	**Scene4**
BM	66.7482	73.9296	89.5153	55.7921	68.5243	79.3685	98.265	86.4503
SGM	530.706	520.884	641.854	511.029	629.254	812.362	926.304	858.638
GC	66529.4	51232.4	251386	203799	98654.6	121759	146292	130258

Note: 1 second equals 1000 milliseconds.

From Table 3, the average time of whole eight images test shows that the efficiencies of the three matching algorithms are BM > SGM > GC. The table also shows the operating time of BM and SGM is less than one second, which means that these two matching algorithms have the great ability to achieve a real-time or a near real time processing. This processing time is vital for the surveillance monitoring system with real-time requirement.

Discussion. The BM algorithm performs best in terms of efficiency, but the quality of the matching results is not as good as the other two algorithms. The graph cuts based matching method can produce the most precise disparity images, whereas the operating time normally is over one hundred seconds which limits its use in real-time applications. The semi-global matching algorithm in this experiment is proved to be the matching method with both high operating efficiency and good matching result quality. These superb characterizes enable this method with the potential to be applied in projects that require both high accuracy and real-time operation, such as temporary structure monitoring.

CONCLUSION

This paper evaluated three typical image matching algorithms (block matching, semi-global matching and graph cuts based matching) for 3-D digital image correlation method in order to reconstruct the 3-D models and calculate the deformation and displacement of construction site products, such as temporary structures. The SGM method exhibits a much more comprehensive potential to achieve a matching process for 3-D DIC, which is a critical component in achieving real time or near real time structure safety monitoring. Nevertheless, there are still several limitations of this method. For instance, the object boundaries are often not well matched and the miss-matching in those non-textured, discontinuity of disparity regions and occlusion areas are still quite difficult issues. Thus, further investigation is still needed.

ACKNOWLEDGEMENTS

The authors acknowledge NASA WV EPSCoR for its partial support.

REFERENCES

Chang, X., and Qunge, H. (2009). “Dynamic structural monitoring using the kinematic positioning of a robotic total station.” *9th International Conference on Electronic Measurement & Instruments*, IEEE, pp. 2-69 - 2-73.

Chen, X., Udupa, J. K., Bagci, U., Zhuge, Y., and Yao, J. (2012). “Medical image segmentation by combining graph cuts and oriented active appearance models.” *IEEE Transactions on Image Processing*, 21(4), 2035-2046.

Choi, S. W., Kim, Y., Kim, J. M., and Park, H. S. (2013). “Field monitoring of column shortenings in a high-rise building during construction.” *Sensors*, 13(11), 14321-14338.

Feng, Y. P., Zhou, H., Zhai, P. F., Zhang, Y. N., and Sun, L. L. (2010). “Development of 3-D reconstruction system of the craniofacial hard tissue based on 2-D digital radiograph.” *Shanghai journal of stomatology*, 19(5), 480-485.

Hirschmuller, H. (2008). “Stereo processing by semiglobal matching and mutual information.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328-341.

Hirschmuller, H., and Scharstein, D. (2009). “Evaluation of stereo matching costs on images with radiometric differences.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9), 1582-1599.

Väänänen, S. P., Amin Yavari, S., Weinans, H., Zadpoor, A. A., Jurvelin, J. S., and Isaksson, H. (2013). “Repeatability of digital image correlation for measurement of surface strains in composite long bones.” *Journal of Biomechanics*, 46(11), 1928-1932.

Wang, J., and Zheng, N. (2013). “A novel fractal image compression scheme with block classification and sorting based on Pearson's correlation coefficient.” *IEEE Transactions on Image Processing*, 22(9), 3690-3702.