This is a post-peer-review, pre-copyedit version of an article published in 2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON). The final authenticated version is available online at: https://doi.org/10.1109/3DTV.2017.8280416

STEREO CAMERA UPGRADED TO EQUAL BASELINE MULTIPLE CAMERA SET (EBMCS)

Adam L. Kaczmarek

Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology,
ul. G. Narutowicza 11/12, 80-233 Gdansk, Poland

ABSTRACT

The paper presents the results of using a set of five cameras called Equal Baseline Multiple Camera Set (EBMCS) for making 3D images, disparity maps and depth maps. Cameras in the set are located in the vicinity of each other and therefore the set can be used for the purpose of stereoscopy similarly as a stereo camera. EBMCS provides disparity maps and depth maps which have a better quality than these maps obtained with the use of a stereo camera. Moreover, EBMCS has many advantages over other kinds of equipment for making 3D images such as a Time-of-flight camera (TOF), Light Detection and Ranging (LIDAR), a structured-light 3D scanner, a camera array and a camera matrix. These advantages are described in the paper. The paper also compares the performance of EBMCS to the performance of stereo cameras.

Index Terms — Stereo camera, camera array, camera matrix, depth maps, stereo matching algorithm

1. INTRODUCTION

There are a lot of devices for making 3D images, 3D videos and 3D models. 3D imaging can be performed with the use of stereo cameras [1, 2], laser scanners [3, 4], structured-light 3D scanners [5] and structure from motion scanning [6]. Each kind of an imaging technique has its advantages and disadvantages.

If a stereo camera made a sufficient quality of 3D images, then developing other kinds of 3D imaging devices would not be necessary. However, in many cases results of using stereo cameras are unsatisfactory. This paper proposes using Equal Baseline Multiple Camera Set (EBMCS) which can be perceived as a set of four, integrated stereo cameras. The considered set preserves all advantages of a stereo camera, however it makes it possible to obtain 3D images with significantly better quality than a single stereo camera. Moreover, using EBMCS instead of other kinds of 3D imaging devices have also many other benefits.

Laser scanners used for making 3D images include Light Detection and Ranging (LIDAR) devices [3] and Time of Flight (TOF) cameras [4]. This equipment provides a very precise measurement of a distance between an imaging device and a single point of an object in the field of view. A 3D image obtained by a laser scanner consists of a series of such measurements. However, there are significant limitations to the number of considered points. Therefore, the point cloud detected by laser scanners is sparse in comparison to the results of other 3D imaging devices.

This problem does not occur when a structured-light 3D scanner is used [5]. These scanners provide dense and accurate depth maps consisting of distances between a scanner and analysed objects. The structured light scanner emits a light pattern such as vertical stripes and it traces distortions of the light on objects. The light pattern emitted by the scanner needs to be intensive enough to perform the measurement. As a consequence, structured 3D light scanner are only suitable for making 3D images of relatively small objects. I case of large object, e.g. buildings, it is complicated to generate light pattern covering the entire object. Moreover, scanning is even more problematic under conditions of intensive natural light occurrence. Natural light interferes which the light emitted by the scanner deteriorating the quality of scanning.

Another technique which can be used for making 3D models suitable for 3D videos is recovering the structure from motion [6]. The method bases on making images of the object from different points of view located around the object. A 3D model is acquired by matching the same parts of the object visible in different images. The method leads to reasonable results however its disadvantage is such that it requires making a great number of images from different places. It is less time consuming to make a 3D image from a single point of view.

Such an image can be acquired by a stereo camera [1, 2]. Stereo cameras provide dense depth maps which can be obtained in intensive natural light. It is also not complicated to acquire images of large objects. Moreover, making a 3D image does not require relocating the imaging equipment. Stereo cameras have many advantages, but they have a major drawback which is a problem with the quality of 3D images. In real applications there are many cases in which distances from a camera to objects are estimated improperly, despite significant advances in designing stereo matching algorithms and camera calibration techniques.

This paper does not focus on developing new algorithms for stereo cameras, but it proposes improving the quality of distance estimation by taking advantage of a larger number of cameras. The paper analyses the usage of a set of cameras called Equal Baseline Multiple Camera Set (EBMCS) [7, 8]. The set consists of a central camera and four side cameras. The set has a function of four stereo cameras with the common central camera present in every stereo pair. There are many benefits of using this set instead of other kinds of camera based 3D vision systems such as stereo cameras, camera arrays consisting of cameras located along a straight line [9] and camera matrix [10] with cameras placed on a grid.

Main contributions of this paper includes the analysis of the influence of an aggregating window size on the quality of disparity maps obtained with the use of the algorithm proposed by Park and Inoue [7]. The algorithm was designed for EBMCS. The paper also compares results of this algorithm to the results of obtaining disparity maps with the use of a single stereo camera. Moreover, the paper describes types of algorithms which can be used with EBMCS for obtaining disparity maps, depth maps and 3D images.

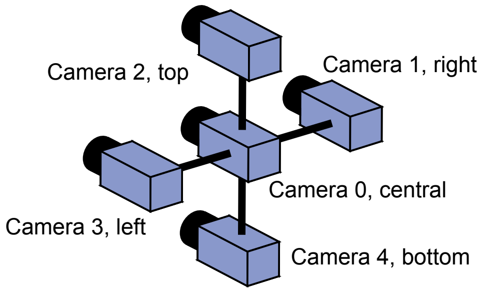

Figure 1. The arrangement of five cameras in Equal Baseline Multiple Camera Set (EBMCS)[8]

2. EQUAL BASELINE MULTIPLE CAMERA SET (EBMCS)

Equal Baseline Multiple Camera Set considered in this paper consists of five cameras having the following locations: central, left, right, up and down (Figure 1). This kind of a camera arrangement was first proposed by Park and Inoue [7]. They also developed an algorithm for obtaining depth maps with the use of the set. Other kinds for algorithms designed for EBMCS were also introduced by the author of this paper [11, 8, 12].

In EBMCS each pair of cameras consisting of a central camera and one of side cameras comprises on a single stereo camera. Therefore, the set can be used as four stereo cameras with the same reference camera which is a central camera in the set. The fact the same camera is included in every considered stereo camera is very significant, because it simplifies merging data acquired from all five cameras in order to obtain a 3D image with improved quality. Stereo cameras with different reference cameras provide 3D images taken from different points of view. Combining such 3D images into a single 3D image requires identifying in these images points corresponding to the same objects. In EBMCS this problem does not occur because all considered stereo cameras have the same point of view.

EBMCS is in fact an improved form of a camera array. Camera array consists of cameras located along a line. In a camera array any two cameras form a stereo camera. However, only these stereo cameras are used which share the same camera [9]. The reason for that is the same as the one described above. The problem with stereo cameras which share the same camera selected from the array is such that they have different baselines. A baseline in a stereo camera is the distance between its constituent cameras. The baseline influences differences between locations of points in the reference image and locations of points referring to the same object in the side image. The longer is the baseline, the greater are these disparities. Different baselines are not advantageous for processing data from stereo cameras [9]. However, this problem also does not apply to EBMCS because in this set baselines are the same in every considered pair of cameras.

Images from EBMCS similarly as images from a stereo camera needs to be processes by a stereo matching algorithm in order to retrieve information about distances from an image making device to objects. This distances are saved in a from of a depth map which combined with the 2D image creates a 3D image. In case of EBMCS there are two types of stereo matching algorithms which can be used.

The first type is the one in which depth maps are made by every stereo camera considered in EBMCS. Then, these depth maps are merged together in order to obtain a single depth map with higher quality than depth maps acquired from different pairs. This method can be interpreted as making four measurements of the same distance and on this basis drawing conclusions about its length. In this method images made by every considered pair of cameras are processes independently from each other. This method will be called in this paper merging methods.

In another type of a stereo matching algorithm used for EBMCS a depth map is obtained by simultaneously using data from every image obtained from EBMCS. In this method depth maps are not merged because the result of a matching process is a single depth map calculated on the basis of all input images. This type of methods will be called the simultaneous methods.

3. MERGING METHODS FOR EBMCS

One of merging methods designed for EBMCS was proposed by the author of this paper [8]. The method was called Exceptions Excluding Merging Method (EEMM). The method is based on excluding distance estimations which have low credibility.

The result of the method is a disparity map obtained from four disparity maps created with the use of stereo cameras included in EBMCS. Four maps processed by the method are calculated by stereo matching algorithms developed for a single stereo camera. These algorithms can be used with EEMM without any modifications.

Stereo matching algorithms does not always provide data about disparities for all points of disparity maps. These are the points for which the disparity could not be determined. Therefore, in case of merging four disparity maps the number of merged disparities in a single point varies from zero to four. This fact was reflected in the design of the EEMM method. The method calculated disparities differently with regard to the number of available values of disparities. The description of the method is presented in [8].

The method was applied to eight different matching algorithms. Experiments included algorithms implemented in the OpenCV library [13], algorithms based on Markov Random Field available in Middlebury Stereo Vision Page (http://vision.middlebury.edu/MRF) [14] and Eficient Large-scale Stereo Matching (ELAS) algorithm [15]. The results showed that, on average, EBMCS used with EEMM creates disparity maps containing over 26% less errors in disparity estimation than disparity maps obtained with the use of a single stereo camera. These results were obtained when the same stereo matching algorithm was used with both a single stereo camera and four stereo cameras included in EBMCS. The best results were obtained for algorithm based on a graph cut which was used with EBMCS containing five cameras.

4. SIMULTANEOUS METHODS FOR EBMCS

Simultaneous methods of obtaining depth maps with the use of EBMCS were also developed by the author of this paper [11]. The author proposed an algorithm called Multiple Similar Areas (MSA). The algorithm was based on analysing areas in images in which adjacent points have a similar color. For points of a central image the MSA algorithm searched for the longest sequences of points in side images having a the same color as the one located in the central image. The search was performed simultaneously in all side images. A detailed description of MSA is available in [11].

A simultaneous method for the EBMCS set was also introduced by Park and Inoue who proposed this kind of a camera system [7]. Their algorithm was based on Sum of Square Differences measure commonly used for obtaining disparity maps on the basis of images from stereo cameras [9]. The algorithm proposed by Park and Inoue calculates the matching function presented in Eq.

1. The results of this function are obtained for points of the central image.

$$\hat{d}(\mathbf{x}) = \underset{d}{argmin}(min[SSD_{left}(\mathbf{x}, d), SSD_{right}(\mathbf{x}, d)] + min[SSD_{top}(\mathbf{x}, d), SSD_{bottom}(\mathbf{x}, d)]) \quad (1)$$

where $\mathbf{x}$ are coordinates of a point in a cental image, d is a disparity for which the function is calculated, $\hat{d}$ is the disparity which is the result of a function and SSD is the SSD function presented in Eq. 2.

$$SSD_i(\mathbf{x}, d) = \sum_{\mathbf{j} \in \mathbf{W}} (I_0(\mathbf{x} + \mathbf{j}) - I_i(\mathbf{x} + \mathbf{j} + \mathbf{d_i}))^2 \quad (2)$$

where $\mathbf{d_i}$ is the disparity in a stereo camera containing a side camera i, I_i is the intensity of the point at coordinates $\mathbf{x}$ in the image from the camera i and $\mathbf{W}$ is an aggregating window identifying points located in the vicinity of the point $\mathbf{x}$.

The matching function proposed by Park and Inoue interprets cameras included in EBMCS as two intersecting camera arrays. The first array consists of three cameras located along a horizontal line and a second array are cameras located along a vertical line. The algorithm selects from each array a side camera for which the matching cost is the lowest. The matching cost corresponds to the similarity level between an area in a central image and areas in side images. The algorithm finds two side cameras from different arrays for which the sum of matching costs is the lowest. There are cameras which does not form with a central camera a straight line.

The algorithm takes advantage of the SSD measure which uses aggregating windows. The aggregating window defines the size of areas which are matched by the algorithm. The size of an aggregating window needs to be selected properly because it influences the performance of stereo matching algorithms. The greater is the window, the more number of calculations needs to be performed by the algorithm. However, if the size of the window is too low, the results of matching can have a low quality. This next section of this paper analysis the influence of the aggregating window size on the quality of results obtained with the use of the algorithm proposed by Park and Inoue.

5. EXPERIMENTS

This paper presents experiments which compared the results of using EBMCS to the results of using a single stereo camera. Two methods of obtaining disparity maps were analysed. The first method was the algorithm proposed by Park and Inoue. In the second method disparity maps were obtained from a single stereo camera with the use of a stereo matching algorithm based on the SSD measure which is also used in the Park and Inoue algorithm. The stereo camera used in the experiment is the one which consists of the central camera and the left one included in EBMCS. The results of these algorithms were compared with regard to the size of aggregating window used for obtaining disparity maps.

Experiments were based on a testbed described in [11]. The testbed consisted of six set of images. Each set contained five images made with the use of five cameras included in EBMCS. Camera included in EBMCS were calibrated with the use of the OpenCV library [8, 13]. Every test set contained also ground truth which was a disparity map with real values of disparities. The image of a real EBMCS used for making these images is presented in Figure 2.

Figure 2. Real EBMCS used in the experiments [8]

The quality of the results obtained in the experiments were measured with the use of the bad matching pixel (BMP) measure. The formula for the measure is presented in Eq. 3.

$$BMP = \frac{1}{N} \sum_{\mathbf{x}} (|D_M(\mathbf{x}) - D_T(\mathbf{x})| > Z) \quad (3)$$

where $D_M(\mathbf{x})$ is the disparity of the point $\mathbf{x}$ in the disparity map, $D_T(\mathbf{x})$ is the disparity in ground truth, N is the total number of points and Z is the threshold value.

The measure calculates differences between ground truth and points in a disparity map which is the result of a matching algorithm. The point is accepted as matched correctly if the difference if below the given threshold Z. In other cases it is considered to be matched incorrectly.

Figure 3 shows the results of the experiments. The line with triangles show the values of BMP obtained with the use of a stereo camera and the SSD measure. The line with crosses shows the results for EBMCS used with the algorithm proposed by Park and Inoue. The vertical axis show the size of a side of an aggregating window. Results show that for each aggregating window size the results obtained with EBMCS were significantly better than the results obtained for a single stereo camera. The improvement varies from 22% to 50% of the value acquired for a stereo camera. The greatest improvement was obtained for a window size equal to 5.

Results also show that another benefit of using EBMCS is such that the size of an aggregating window can be reduced. When stereo camera was used, the best results were obtained for the window size equal to 13. In this case the window contained 169 points. The best results for EBMCS were acquired with the window of a size 7. The window has 49 points. Therefore, the window is over 3 times smaller. It is very significant because not only results are better with the use of EBMCS but also smaller aggregating windows can be used when EBMCS is used instead of a stereo camera.

There results were obtained for the similar kind of matching algorithms. Both the Park and Inoue algorithm and the algorithm based on SSD measure are a local type of matching methods. These methods in order to acquire the value of a disparity in a single point analyze only a part of input images. Both of these algorithm do not use global optimization which significantly improves the quality of disparity maps.

There results presented in this paper are consistent with previous work performed by the author of this paper [8, 12]. Th previous research the author analysed the performance of the Sum of

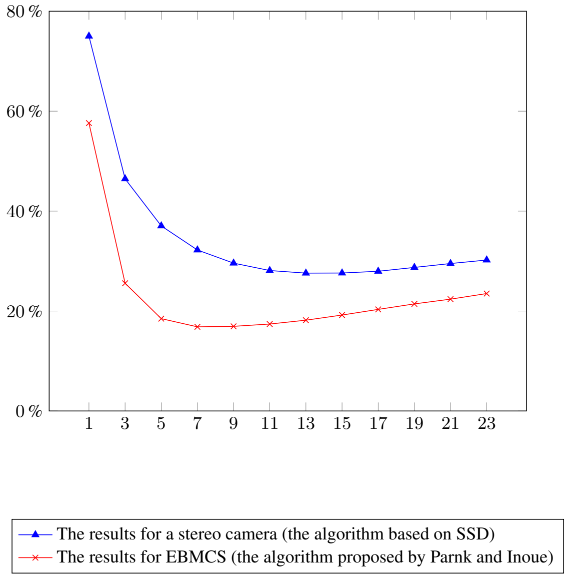

Figure 3. The quality of disparity maps with regard to aggregating window size

sum of Square Differences (SSSD) measure modified for the purpose of using this measure with EBMCS [12]. SSSD is a kind of a simultaneous method of a local type. Experiments with SSSD were performed on a smaller data set than the experiments with Park and Inoue algorithm presented in this paper. The author also showed that using a EBMCS with a merging method using stereo matching algorithm with global optimization improves results by over 26% percent [8]. A potentially highly important area of research is designing simultaneous methods for EBMCS which take advantage of global optimization. It is possible that such a method would produce better results than existing methods.

6. SUMMARY

Advantages of using EBMCS are evident. Better quality of depth maps can be acquired with the use of a device similar to a stereo camera. It is surprising that this technology is not commonly used for making 3D videos. In further research the author plans to develop other kinds of algorithms for EBMCS. The author is looking for cooperation with other researchers of stereo matching algorithms and multi-camera systems to design an algorithm for EBMCS which would lead to even better results than currently known methods.

7. REFERENCES

[1] S. A. Fezza, A. Chetouani, and M. C. Larabi, "Universal blind image quality assessment for stereoscopic images," in *2016 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2016, pp. 1–4.

[2] D. Grogna, G. Guerri, and J. G. Verly, "Stereo matching method using multiple angular signatures matched by multidimensional dynamic timewarping (MD-DTW)," in *2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2014, pp. 1–4.

[3] J. Gao and R. Yang, "Online building segmentation from ground-based LiDAR data in urban scenes," in *2013 International Conference on 3D Vision - 3DV 2013*, 2013, pp. 49–55.

[4] J.-D. Yoo, J. H. Cho, I.-Y. Jang, and K. H. Lee, "3D content generation using a TOF camera and HDR light information," in *2011 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2011, pp. 1–4.

[5] W. Jang, C. Je, Y. Seo, and S. W. Lee, "Structured-light stereo: Comparative analysis and integration of structured-light and active stereo for measuring dynamic shape," *Optics and Lasers in Engineering*, vol. 51, no. 11, pp. 1255–1264, 2013.

[6] Y. Zhang, P. Teng, Y. Shimizu, F. Hosoi, and K. Omasa, "Estimating 3D leaf and stem shape of nursery paprika plants by a novel multi-camera photography system," *Sensors*, vol. 16, no. 6, 2016.

[7] J.-I. Park and S. Inoue, "Acquisition of sharp depth map from multiple cameras," *Signal Processing: Image Communication*, vol. 14, no. 1-2, pp. 7–19, 1998.

[8] A. L. Kaczmarek, "Stereo vision with equal baseline multiple camera set (EBMCS) for obtaining depth maps of plants," *Computers and Electronics in Agriculture*, vol. 135, pp. 23–37, 2017.

[9] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 4, pp. 353–363, 1993.

[10] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," in *ACM SIGGRAPH 2005 Papers*, New York, NY, USA, 2005, SIGGRAPH '05, pp. 765–776, ACM.

[11] A. L. Kaczmarek, "Improving depth maps of plants by using a set of five cameras," *Journal of Electronic Imaging*, vol. 24, no. 2, pp. 023018, 2015.

[12] A. L. Kaczmarek, "Influence of aggregating window size on disparity maps obtained from equal baseline multiple camera set (EBMCS)," in *Image Processing and Communications Challenges 8: 8th International Conference, IP&C 2016 Bydgoszcz, Poland, September 2016 Proceedings*. 2017, pp. 187–194, Springer International Publishing.

[13] G. R. Bradski and A. Kaehler, *Learning Opencv, 1st Edition*, O'Reilly Media, Inc., first edition, 2008.

[14] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.

[15] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Computer Vision ACCV 2010*, vol. 6492 of *Lecture Notes in Computer Science*, pp. 25–38. Springer Berlin Heidelberg, 2011.