

Adaptive algorithm of conforming image matching

V A Fursov^{1,2}, Ye V Goshin^{1,2} and K G Pugachev¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: pugachev_k.g@mail.ru

Abstract. This paper presents an adaptive algorithm of conforming image matching based on the principle of conformity. The algorithm consists of several main stages. At the first stage, we find the corresponding points using a minimum value of conformity as the measure of points' similarity. We define a conformity function as the sum of all possible combinations of squared differences of pixel intensity values on the fragments that are matched. Then, we perform an adaptive procedure of errors correction considering an intensity gradient distribution. An important feature of the algorithm is the finding of error points using a criterion of maximum value of samples' conformity for every fragment of the disparity map. The results of experiments on the "Teddy" test images are shown.

1. Introduction

Image matching is the most important and difficult step in computer vision applications, such as image stitching, the creation of three-dimensional models of objects or scenes by a sequence of frames etc. At least three approaches to solve image matching problem are known: area-based (1), feature-based (2), and symbolic (3) approach. In the framework of the area-based approach [1], which is used in this paper, for the fragments of the first image, the most similar fragment is need to be found.

The choice of fragments similarity measure is very important, when the area-based approach is used. There are many similarity measures. The absolute differences (AD) and squared differences (SD) and the sum of absolute differences (SAD) are the most commonly used to compute matching cost. Also, the sum of squared differences (SSD), normalized cross correlation (NCC), rank transformation (RT), and census transformation (CT) are the popular methods to aggregate matching cost. In detail, these methods were described in paper [2].

In paper [3] Wang et al. applied the AD algorithm to real-time image matching using GPU. The results show that AD algorithm is not capable of producing a smooth disparity map in highly textured images. To overcome it, Min et al. [4] and Pham and Jeon [5] implemented truncated absolute differences (TAD) algorithm, which uses the colour and gradient values in pixels matching to improve its robustness against variations in illumination.

SD algorithm was applied to subpixel image matching by Yang et al. [6]. To improve the flattening of edges and to smooth areas, they used a bilateral filter. In paper [7] Miron et al. researched various matching cost functions. They concluded that the SD algorithm produces the largest error.

Lee and Sharma [8] implemented the real time image matching algorithm using the SAD algorithm to calculate matching costs via GPU. In paper [9] Gupta and Cho proposed a technique, which uses two different sizes of correlation windows in the SAD algorithm. As denoted in [2], the SAD algorithm is fast, but the quality of the initial disparity map is low because of the noise at object boundaries and in homogeneous regions.

In paper [10] Fusiello et al. used the SSD algorithm with multiple fixed window blocks to reduce the incidence of occlusion errors. In paper [11] the SSD algorithm with multiple windows was implemented by Yang and Pollefeys using GPU.

Low-dimensional image features matching method using NCC was proposed in [12] by Satoh. This method achieves a high accuracy, but requires a lot of computational resources. Cheng et al. [13] proposed to calculate matching cost using a zero mean normalized cross correlation (ZNCC).

An image matching algorithm using the RT algorithm was proposed by Gac et al. [14]. The method provides a reliable initial disparity map by using a careful selection of the window sizes. A new RT algorithm was developed in paper [15] to reduce matching ambiguity using a Bayesian model. This model considers similarity between the first and the second image pixels and the level of ambiguity within each image independently.

To improve image matching accuracy, some researchers tried using combinations of various cost functions. For instance, Mei et al. [16] implemented a combination of the AD and CT algorithms to compensate their limitations. They showed that CT algorithms produce incorrect matches in regions with repetitive local structures, and the AD algorithm does not work well in the large regions without texture.

The combination of the SAD and CT algorithms led to higher performance, but computational complexity has become higher as well [17]. The SAD and CT cost measures were obtained individually. The final cost function was calculated as a linear combination of SAD and CT cost measures based on a weighting factor.

Zhang et al. [18] has improved the matching accuracy by using a combination of SAD and the arm length differences (ALD) methods. Lee et al. [19] used combination of CT and gradient difference methods. However, due to similar or repetitive texture patterns, matching ambiguity can appear in some regions.

In paper [20] Perez-Patricio et al. has proposed a fuzzy logic method using the various cost functions simultaneously and decision-making based on membership functions. The method uses both local and global information to reduce errors' count in homogeneous areas and to preserve the edges near discontinuities.

When the area-based approach is used, many points with wrong disparity value can appear in disparity map. In our previous papers we have tried to solve this problem by using a new criterion of similarity. In particular, in papers [21, 22] a new fragments' similarity criterion was presented. Although we have obtained the results comparable and in some cases are better than results of semi-global block matching, however, this problem was not fully solved.

The traditional way of noise removal is smoothing filtering of image representing disparity map. This method can give visually better image, however this smoothing procedure unavoidably distorts into disparity map points with true value of disparity. Consequently, the three-dimensional coordinates of the scene, which are calculated using these points, can also be distorted. So, the problem of filtering procedure creating, which performs error correction without affecting the other points with high reliability, is actual task.

2. Problem statement

In this paper we develop an approach based on the principle of fragments conformity. The method was proposed in [21, 22]. Let us state that F_1 and F_2 are two images, which are obtained by multi-view shooting. For every point of these images we can specify a rectangular fragment with size of $K \times L$, K and L are odd. In general case, $K \neq L$. Every fragment of the first image F_1 can be presented in the form of vector \mathbf{f}_1 , obtained by scanning by rows or columns.

Let's specify rectangular search area with size of $N \times M$ on the second image F_2 for every point of the first image F_1 . For every point in search area we can form vector \mathbf{f}_2 with the same size in a similar way.

The task is to find corresponding point (central point of \mathbf{f}_2 , the most similar for some measure) in the search area of the second image F_2 for every point (central point of \mathbf{f}_1) of first image F_1 . For certainty, we assume that fragments sizes are odd and the search area is a set of points, which can be used as centres of fragments.

For points of the first image with coordinates (k, l) and $N \times M$ points in search area in second image vectors of differences are calculated:

$$\Delta \mathbf{f}(n, m) = \mathbf{f}_2(n, m) - \mathbf{f}_1(k, l), \quad n = \overline{1, N}, \quad m = \overline{1, M}, \quad (1)$$

where N and M are count of rows and columns of the search area.

Then conformity function is calculated as:

$$W(n, m) = \sum_{\substack{i, j=1, \\ i < j}}^S (\Delta f_i(n, m) - \Delta f_j(n, m))^2, \quad (2)$$

where Δf_i is i^{th} element of vector (1), $S = K \times L$ is elements count of \mathbf{f}_1 and \mathbf{f}_2 vectors, $n = \overline{1, N}, m = \overline{1, M}$. Point (n^*, m^*) in search area of the second image with minimum value of conformity function is chosen as corresponding point for point (k, l) at the first image:

$$W(n^*, m^*) = \min_{\forall n, m} W(n, m) \quad (3)$$

We have used this criterion in papers [21, 22], where we have shown that this criterion allows improving an accuracy of image matching. The important feature is that all possible combinations of fragments samples differences are used. So, this criterion gives much more information about pixel intensity distribution of image fragment and provides a high reliability of image matching. However, as we mentioned above, the points with wrong calculated disparity can occur in the areas with low values of gradient function.

Proposed filter, which allows overcoming it, is also based on the principle of samples conformity. The main idea is to use the conformity function like (2) to find error point and replace its value by values of neighbouring points. But the points without errors are not will be affected. This method allows avoiding errors in image areas with low values of pixel intensity gradient.

3. Method of error correction

We develop an error correction algorithm for areas of disparity map image with low values of pixel intensity gradient. Let us state that F_d is image representing obtained disparity map. Let's choose fragment of this image with size of $P \times Q$, P and Q are odd. In general case $P \neq Q$.

For every point (p, q) we calculate value of local conformity with other points of this fragment:

$$W(p, q) = \sum_{\substack{i=1, P, \\ j=1, Q \\ i \neq p, j \neq q}}^{S-1} (f(p, q) - f(i, j))^2, \quad (4)$$

where $f(p, q)$, $f(i, j)$ are values of fragments samples in points (p, q) and (i, j) , and $S = P * Q$.

Now S functions of local conformity are calculated. We search point (p^*, q^*) with the maximum value of local conformity and assume it as error point:

$$W(p^*, q^*) = \max_{\forall p, q} W(p, q) \quad (5)$$

This statement is based on an assumption that all points of this fragment are in areas with low value of pixel intensity gradient. So, the three-dimensional coordinates of elements of the same objects can't be too different from neighbouring points, because the objects don't contain any tears of surface.

To correct the value of error point (p^*, q^*) , an error value is replaced by mean value of variational range of fragment samples. In case, when count of errors in disparity map is low, it's useful to compare an error value with mean value of variational range. If the difference is low and can be compared with possible differences in coordinates of surface elements, this replacement will not change anything. It can be seen that all points will remain the same. So, the coordinates of surface can be calculated accurately using the values of obtained disparity map.

On the fragments with objects edges this procedure will not work because the values of conformity function (4) will be high due to the difference between fragments samples of different objects. So, these image areas should be excluded from consideration. Moreover, this filter works efficiently, when count of error points on the fragment is low.

In view of the above, the next section proposes a multistage technology with adaptation to the features of the various parts of the disparity map, ensuring the efficient correction of error points.

4. Overall technology of disparity map forming

The technology consists of several main stages:

1. Disparity map forming using a method of conforming estimates;
2. Preliminary disparity image processing, single errors' correction;
3. Object edges detection and binary template construction;
4. Error points correction on the areas with the low value of gradient function (on the objects surface).

Let us describe briefly these stages of technology.

At the first stage we form disparity map using method and algorithm which are described in [21, 22]. The main idea of this method was described in Section 2.

Preliminary disparity image processing is performed to correct values of single error points. The algorithm is implemented by several iterations (by rows and columns) on the image. For each point of the image difference from neighbouring points (by vertical or horizontal) and the difference between neighbouring points are calculated. If the first difference is higher than threshold value and the second value is low, this point is chosen as an error point. Then, the value of the central pixel is replaced by half-sum of neighbouring pixels values. Similar procedure can be used to correct double, triple and etc. errors.

To construct a binary template with detected edges, various methods can be used to find the differences, similarities [23] and image segmentation [24]. In this paper we use the simplest and widely used approach. We use a median filter to blur images and Sobel operator to form the field of gradients. Then we perform the contrasting and binarization procedures.

At the last stage, we perform iterative error correction in the areas of disparity map with low value of gradient. We divide disparity image into disjoint fragments with size of $P \times Q$ and iteratively perform an error correction using method, which is described in Section 3. If the fragment contains objects edges, an error correction is not performed. When the fourth stage is performed, some clusters of failed points can be "transformed" into the single error points. Therefore, the second stage correction can be repeated without loss of accuracy.

5. Results of experiment

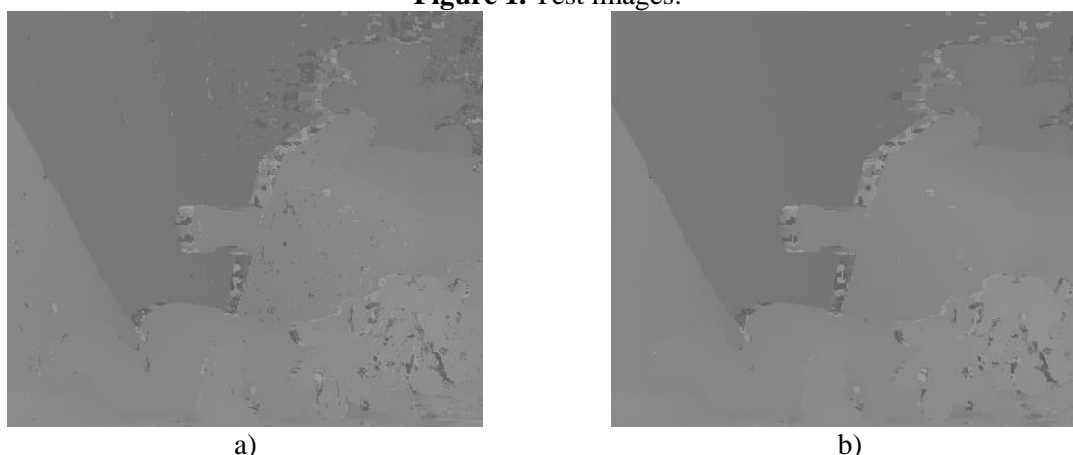
The described technology was used to form a disparity map for a set of multi-view test images «Cones» [25]. Test images are presented at Figure 1. For these test images, the disparity map shown at Figure 2a was obtained using method and algorithm described in [21, 22]. Search window with the size of 5×7 and search area with the size of 1×65 were used. Then median intensity of pixels was set as 128. It can be seen that disparity map contains many points with the wrong disparity. The most of them, as expected, are in the area without any local features (textures, symbols etc.).

Figure 2b shows a disparity map obtained after several iterations of single, double and triple error's removal (Stage 2). Also, correction of big areas with errors was performed. We can see, that this disparity map is still consists of many errors. Moreover, most of these points are contained in the areas of the test images with a small value of gradient.



a) left image; b) right image

Figure 1. Test images.



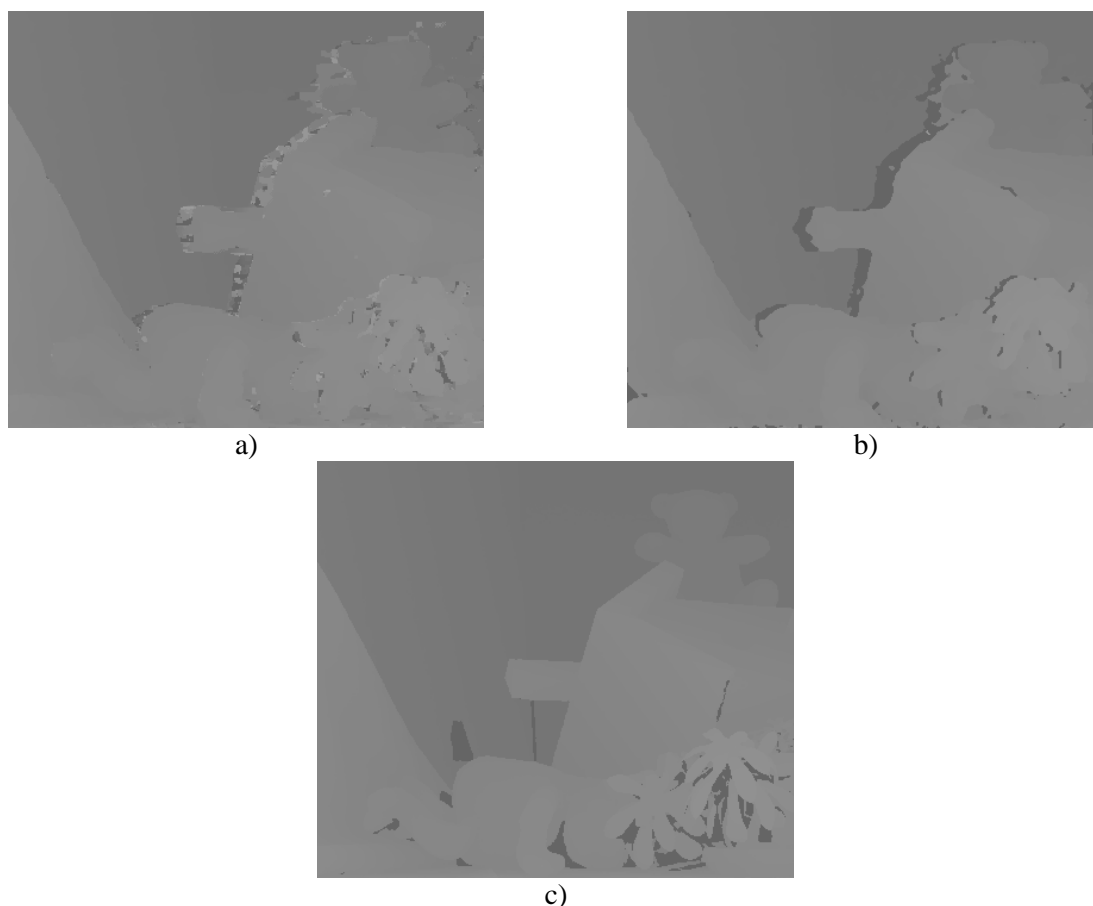
a) Obtained disparity map (first stage); b) Processed disparity map (second stage)

Figure 2. Disparity maps.

Figure 3a shows disparity map obtained using our error correction algorithm, which is described in Section 3. To compare the obtained results with the other methods and algorithms, we have constructed a disparity map using semi-global block matching method [26]. We have used the implementation of this method from OpenCV library [27] with parameters:

- 1) minDisparity=0;
- 2) numberOfDisparities=80;
- 3) SADWindowSize=5;
- 4) disp12MaxDiff=1;
- 5) preFilterCap=31;
- 6) uniquenessRatio=5;
- 7) speckleWindowSize=5;
- 8) speckleRange=5.

Disparity map obtained by semi-global block matching is shown at Figure 3b. Figure 3c shows the ground-truth disparity map. It can be seen that semi-global matching smooths errors. However, this method also smooths edges of objects.



a) Filtered disparity map (fourth stage); b) Disparity map, obtained using Semi-Global Block Matching; c) Ground-truth disparity map

Figure 3. Disparity maps.

Accuracy comparison for semi-global matching method and our performed technology was performed. For both methods an error of image matching was calculated using the formula:

$$\sigma = \sqrt{\frac{1}{NM} \sum_{i,j=1}^{N,M} (I_{i,j}^1 - I_{i,j}^2)^2}$$

Calculation was performed for points, where disparity value is known. So, for our proposed method error value is 5,22 before processing. After processing error value became 3,98. For semi-global matching error value is equal to 4,00. We can see that proposed method of error correction can significantly reduce the number of errors in the disparity map.

6. Conclusion

A new method and algorithm of error correction on the areas with low value of pixel intensity gradient are proposed. The important feature is that an algorithm allows removing only point with the wrong disparity value. The other points of fragment are not affected. This feature is very useful for disparity maps formation, which affects the three-dimensional models calculation of coordinates. The constructed technology, including this algorithm, has a high accuracy.

7. References

- [1] Guk A P and Altsyntsev M A 2017 Automatic identification of corresponding points for aerial images of forest areas *Vestnik of SSUGT* **22** 68-77
- [2] Rostam A H and Haidi I 2016 Literature Survey on Stereo Vision Disparity Map Algorithms *J. Sens.* 8742920

- [3] Wang L, Gong M, Gong M and Yang R 2006 How far can we go with local optimization in real-time stereo matching *Proc. 3rd Int. Symp. 3D Data Proc. Vis. Trans.* (Chapel Hill, NC, USA/ IEEE) 129-136
- [4] Min D, Lu J and Do M N 2011 A revisit to cost aggregation in stereo matching: how far can we reduce its computational redundancy? *Proc. IEEE Int. Conf. Comp. Vis. (ICCV '11)* (Barcelona, Spain/IEEE) 1567-1574
- [5] Pham C C and Jeon J W 2013 Domain transformation-based efficient cost aggregation for local stereo matching *IEEE Trans. Circ. Sys. Vid. Tech.* **23** 1119-1130
- [6] Yang Q, Yang R, Davis J and Nistrer D 2007 Spatial-depth super resolution for range images *Proc. IEEE Conf. Comp. Vis. Pat. Rec. (CVPR)* (Minneapolis, Minn, USA/IEEE) 1-8
- [7] Miron A, Ainouz S, Rogozan A and Bensrhair A 2014 A robust cost function for stereomatching of road scenes *Pat. Rec. Let.* **38** 70-77
- [8] Lee S H and Sharma S 2011 Real-time disparity estimation algorithm for stereo camera systems *IEEE Trans. Cons. Electr.* **57** 1018-1026
- [9] Gupta R K and Cho S Y 2013 Window-based approach for fast stereo correspondence *IET Comp. Vis.* **7** 123-134
- [10] Fusiello A, Castellani U and Murino V 2001 Relaxing symmetric multiple windows stereo using Markov Random Fields *Computer Vision and Pattern Recognition of Lecture Notes in Computer Science* **2134** 91-105
- [11] Yang R and Pollefeys M 2003 Multi-resolution real-time stereo on commodity graphics hardware *Proc. IEEE Comp. Soc. Conf. Comp. Vis. Patt. Rec.* **1** 211-217
- [12] Satoh S 2011 Simple low-dimensional features approximating NCC-based image matching *Pat. Rec. Let.* **32** 1902-1911
- [13] Cheng F, Zhang H, Yuan D and Sun M 2014 Stereo matching by using the global edge constraint *Neurocomp.* **131** 217-226
- [14] Gac N, Mancini S, Desvignes M and Houzet D 2008 High speed 3D tomography on CPU, GPU, and FPGA *EURASIP J. Emb. Sys.* 930250
- [15] Zhao G, Du Y K and Tang Y D 2011 A new extension of the rank transform for stereo matching *Adv. Eng. Forum* **2-3** 182-187
- [16] Mei X, Sun X, Zhou M, Jiao S, H. Wang and Zhang X 2011 On building an accurate stereo matching system on graphics hardware *Proc. IEEE Int. Conf. Comp. Vis. Work.* (Barcelona, Spain/IEEE) 467-474
- [17] Cigla C and Alatan A A 2013 Information permeability for stereo matching *Sig. Proc.: Im. Com.* **28** 1072-1088
- [18] Zhang N, Wang H and Cr J 2013 A near real-time color stereo matching method for GPU *Proc. 3rd Int. Conf. on Adv. Comm. Comp.* (Lisbon, Portugal/ IARIA) 27-32
- [19] Lee S, Lee J H, Lim J and Suh I H 2015 Robust stereo matching using adaptive random walk with restart algorithm *Im. Vis. Comp.* **37** 1-11
- [20] Pérez-Patricio M, Aguilar-González A, Camas-Anzueto J-L and Arias-Estrada M 2015 A Fuzzy Logic Approach for Stereo Matching Suited for Real-Time *Proc. Int. J. Comp. App.* **113** 9
- [21] Fursov V A, Gavrilov A V, Goshin Ye V and Pugachev K G 2018 The technology of image matching by the criterion of conformity of image fragments samples *Inf. Tech. and Nanotech.* (Samara, Russia/ New Engineering Ltd.) 2299-2305
- [22] Fursov V A, Gavrilov A V, Goshin Ye V and Pugachev K G 2018 The technology of image matching by the criterion of conformity of image fragments samples *J. Phys.: Conf. Ser.* **1096** 012084
- [23] Plotnikov D E, Kolbudaev P A and Bartalev S A 2018 Identification of dynamically homogeneous areas with time series segmentation of remote sensing data *Computer Optics* **42(3)** 447-456 DOI: 10.18287/2412-6179-2018-42-3-447-456
- [24] Lebedev M A, Rubis A Yu, Vizilter Yu V and Vigolov O V 2018 Detecting image differences based on reference EMD-filters *Computer Optics* **42(2)** 291-296 DOI: 10.18287/2412-6179-2018-42-2-291-296

- [25] University of Tsukuba Computer Vision Laboratory URL: <http://www.cvlab.cs.tsukuba.ac.jp/dataset/tsukubastereo.php>
- [26] Hirschmüller H 2008 Accurate and efficient stereo processing by semi-global matching and mutual information *CVPR, PAMI* **30(2)** 328-341
- [27] OpenCV Library URL: <http://opencv.org>

Acknowledgments

The work was funded by the Russian Federation Ministry of Education and Science (project # 2.891.2017) and RFBR (projects # 17-29-03112).