Patrik Kamencay – Martina Zachariasova– Robert Hudec – Miroslav Benco – Jan Hlubik – Slavomir Matuska *

IMAGE SEGMENTATION AND FEATURE EXTRACTION USING SIFT-SAD ALGORITHM FOR DISPARITY MAP GENERATION

In this paper, a stereo matching algorithm based on image segments and disparity measurement using stereo images is presented. We propose the hybrid segmentation algorithm that is based on a combination of the Belief Propagation and Mean Shift algorithms with aim to refine the final disparity map by using a stereo pair of images. Firstly, a color based segmentation method is applied for segmenting the left image of the input stereo pair (reference image) into regions. The aim of the segmentation is to simplify representation of the image into the form that is easier to analyze and is able to locate objects in images. Secondly, results of the segmentation are used as an input of the SIFT-SAD matching method to determine the disparity estimate of each image pixel. This matching algorithm is proposed by combining Scale Invariant Feature Transform (SIFT) with the Sum of Absolute Difference (SAD). Finally, the comparisons between the three robust feature detection methods SIFT, Affine SIFT (ASIFT) and Speeded Up Robust Features (SURF) are presented. The obtained experimental results demonstrate that the proposed method has a positive effect on overall estimation of disparity map and outperforms other examined methods.

Keywords: *Belief Propagation, Mean Shift, SIFT, ASIFT, SURF, disparity map.*

1. Introduction

This paper describes a set of algorithms for structure, motion automatic recovery and visualization of a 3D image from a sequence of 2D images. The important step to perform this goal is matching of corresponding pixels in the different views to estimate the depth map. The depth of an image pixel is the distance of the corresponding world point from the camera center. Detecting objects, estimating their pose, geometric properties and recovering 3D shape information are a critical problem in many vision and stereo computer vision application domains such as robotics applications, high level visual scene understanding, activity recognition, and object modeling [1]. The structure and motion recovery system follows a natural progression, comprising the following phases:

- feature matching using SIFT descriptor,
- image segmentation,
- feature detection using SIFT-SAD algorithm,
- disparity and depth map generation.

A classical problem of stereo computer vision is the extraction of 3D information from stereo views of a scene. To solve this problem, knowledge of view properties and feature point between views is needed. However, finding these points is notoriously hard to do for natural scenes. The fundamental idea behind stereo computer vision is the difference in position of a unique 3D point in two different images. As the object moves closer to the cameras, the relative position of object will change, and the positions in each image will move away from each other. In this way, it is possible to calculate the distance of an object, by calculating its relative positioning in the two images. This distance between the same objects in two images is known as disparity [1]. Disparity map computation is one of the key problems in 3D computer vision.

This paper employed a new feature projection approach based on SIFT-SAD method using hybrid segmentation algorithm. A comparison between these two different approaches for the image segmentation (Mean Shift and Belief Propagation) is described in [2]–[4].

The outline of the paper is as follows. The section 2 gives brief overview of the state-of-the-art in stereo matching and stereo correspondence. The proposed method of disparity map estimation from corresponding points using SIFT-SAD algorithm is described in section 3. Finally the experiment results and architecture of reconstruction algorithm are introduced in Section 4 and brief summary is discussed in Section 5.

2. Related work

In this section, we review related stereo. We refer the reader to a detailed and updated taxonomy of dense, two-frame stereo correspondence algorithms by Scharstein and Szeliski [5]. It also provides a tested for quantitative evaluation of stereo algorithms.

* **Patrik Kamencay, Martina Zachariasova, Robert Hudec, Miroslav Benco, Jan Hlubik, Slavomir Matuska**
Department of Telecommunications, University of Zilina, Slovakia, E-mail: patrik.kamencay@fel.uniza.sk

A stereo algorithm is called a global method if there is a global objective function to be optimized. Otherwise it is called a local method. The central problem of local or window-based stereo matching methods is to determine the optimal support window for each pixel. An ideal support region should be bigger in texture less regions and should be suspended at depth discontinuities. The fixed window is obviously invalid at depth discontinuities. Some improved window-based methods, such as adaptive windows [6], shift-able windows [7] and compact windows [7] try to avoid the windows that span depth discontinuities.

In stereo correspondence, two previous comparative papers have focused on the performance of sparse feature matchers [8] and developed new criteria for evaluating the performance of dense stereo matchers for image-based rendering application [9]. Our work is a continuation of the investigations begun by Szeliski and Zabih [10], which compared the performance of several popular algorithms.

3. Proposed method

In this section we present our method for fusing two approaches to disparity map estimation from input stereo images: hybrid segmentation algorithm and SIFT-SAD representation. This proposed algorithm based on the combination of the hybrid segmentation algorithm with SIFT descriptor and SAD stereo matching algorithm is faster, since a small portion of whole left and right images pixels are used for matching. The proposed method shown in Fig. 1 is implemented in MATLAB environment and improves the performance of disparity map calculation.

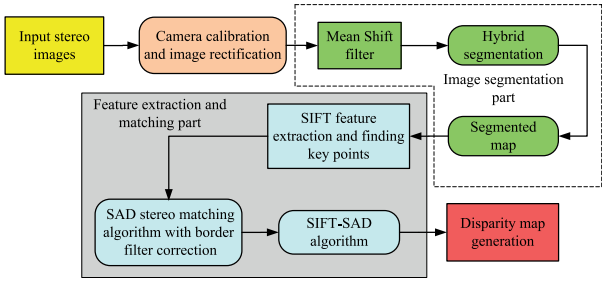

Fig. 1 Architecture for disparity map computation.

First, step is image rectification. It is transformation which makes pairs of conjugate epipolar lines become collinear and parallel to the horizontal axis (baseline). For the epipolar rectified images pair, each point in the left image lies on the same horizontal scan line as in the right image. This approach is used to reduce a search space for disparity map estimation algorithm. Next, we apply image filtering by Mean Shift filter. This step is very useful for noise removing, smoothing and image segmentation [11]. After filtration, the filtered image is split into segments using hybrid segmentation algorithm. Image segmentation (automatically partitioning an image into regions) is an important stage of our proposed algorithm for disparity map estimation. The combination of Mean Shift and Belief Propagation segmentation algorithms are deployed in order to improve precision of the key points search using SIFT and overall complexity. Finally, matching is performed using SAD algorithm, where a disparity map is obtained. The accuracy of SIFT-SAD algorithm depends on the correctness and quality of hybrid image segmentation.

3.1 Hybrid segmentation algorithm

This hybrid approach delivers accurately localized and closed object contours and brings together the advantages of both segmentation algorithms. Mean Shift is quick and Belief Propagation is very accurate segmentation. Initially, the noise corrupting the image is reduced by a noise reduction technique that preserves edges remarkably well, while reducing the noise quite effectively. At the second stage, this noise suppression allows a more accurate calculation and reduction of the number of the detected false edges.

First, we apply image filtering by Mean Shift algorithm. This step is very useful for noise removing, smoothing and image segmentation. For each pixel of an image, the set of neighboring pixels is determined. For each pixel of an image, the set of neighboring pixels is determined. Let X_i be the input and Y_i filtered image, where $i = 1, 2, ..., n$. The filtering algorithm comprises of the following steps

- Compute through the Mean Shift the mode where the pixel converges.
- Store the component of the gray level of the calculated value $Z_i = (x_{i,s}, y_{i,c})$, where $x_{i,s}$ is the spatial component and $y_{i,c}$ is the range component.

Secondly, the image is split into segments using Mean Shift algorithm. In the third step, means of segments are retrieved by applying mean shift theory. Fourth, the small segments are merged together to the most similar adjacent segments by the Belief Propagation method. Finally, we have integrated our proposed hybrid segmentation algorithm with the SIFT descriptor and Sum-of-Absolute-Differences (SAD) stereo matching algorithm. This proposed combination is able to produce highly accurate disparity map [12].

Parameters used in hybrid algorithm Tab.1

Parameter	Set value
p	5
S	50
Min_sh	1
Max_sh	40

The set up parameters of the used hybrid segmentation algorithm are shown in Table 1. The spatial resolution parameter p

affects smoothing and connectivity of segments. Moreover, parameter *S* is a size of the smallest segment, *Min_sh* is minimum shift and *Max_sh* is maximum shift of the pixels.

Advantages of the proposed hybrid algorithm:
- Efficient edge preserving smoothing guided by Mean Shift.
- Ability to change the image topology by using a simple merging mechanism, thus reducing over-segmentation.
- Relatively low sensitiveness to noise.
- Execution time directly proportional to the image size.

3.2 SIFT-SAD algorithm

The performance of stereo matching algorithms depends on the choice of matching cost. In our experiment we proposed SIFT-SAD matching method as matching cost. Scale Invariant Feature Transform (SIFT) is a local descriptor of image features insensitive to illuminant and other variants that is usually used as sparse feature representation. SIFT features are features extracted from images to help in reliable matching between different views of the same object [13]. Basically, in SIFT descriptors the neighborhood of the interest point is described as a set of orientation histograms computed from the gradient image. SIFT descriptors are invariant to scale, rotation, lighting and viewpoint change (in a narrow range). The most common implementation uses 16 histograms of 8 bins (8 orientations), which gives a 128 dimensional descriptor [13]. The SAD algorithm is based on accumulating absolute differences of the left image and right image pixels within a given window. It works by taking the absolute value of the difference between each pixel in the original block and the corresponding pixel in the block being used for comparison. The more similar the pixels are the less the SAD value becomes. These differences are summed over the block to create a simple metric of block similarity, the $L1$ norm of the difference image [14]. SIFT descriptor delivers most of local gradient information and SAD provides local intensity information. SIFT-SAD consists of two parts. Firstly, we get the $L1$ distance of SIFT between pixel p in the left image and $p + d_p$ in the right image.

$$D_{SIFT}(d_p) = \left\| x_L(p) - x_R(p + d_p) \right\|, \tag{1}$$

where d_p is the disparity of pixel p, $\|x_L(\mathrm{p}) - x_R(p + d_p)\|$ is the $L1$ distance. Next, we define SAD matching cost as:

$$D_{SAD}(d_p) = \exp\left(-SAD(p, p + d_p)\right), \tag{2}$$

where SAD $(p, p + d_p)$ is the SAD score in a square neighborhood searching window. Our algorithm computes the disparity for all pixels with a window size dimension at a square of 9×9 pixels. The minimum difference value over the frame indicates the best matching pixel, and position of the minimum defines the disparity of the actual pixel [15]. Then, a linear combination of SIFT-SAD algorithm is proposed as

$$D(d_p) = D_{SIFT}(d_p) + \lambda D_{SAD}(d_p), \tag{3}$$

where λ is a weighting factor that controls the contribution of SIFT part and SAD part. We set $\lambda = 1$ in all the experiments. Finally, we use one dimensional Gaussian weight with a scale factor s to get the matching cost. The underlying assumption is that if a minimum corresponds to the true surface, the neighboring pixels should have near values at a similar depth [15].

Quality of final 3D disparity map depends on square window size, because a bigger window size corresponds to a greater probability of correct pixel disparity calculated from matched points, although the calculation gets slower [15].

4 . Experimental results

In this section, some of the obtained experimental results will be presented. All the experiments were implemented in Matlab. We conducted experiments on Middlebury image database [16] using an Intel(R) Core2 Quad CPU with 2.40 GHz processor. The input to the proposed algorithm is a stereoscopic pair images. A SIFT-SAD matching algorithm lies in the heart of the 3D reconstruction procedure.

First, the proposed hybrid algorithm with three segmentation algorithms were compared using automatic algorithm evaluating the precision of segmentation, as is shown in Table 2. This plays important role for two reasons:
- it can be placed into a feedback loop to enforce another run of segmentation algorithm that may include more sophisticated steps for high precision segmentation,
- outcome of this evaluation can be treated as a quality factor and thus can be used to design a quality driven adaptive recognition system.

The definition of precision (P), recall (R) and $F1$ is given by

$$P = \frac{C}{C + F} * 100, \tag{4}$$

$$R = \frac{C}{C + M} * 100, \tag{5}$$

where C is the number of correct detected pixels that belonging to the boundary, F is the number of false detected pixels and M is the number of not detected pixels. Parameter $F1$ is a combined measure from precision and recall [15]. The definition of $F1$ is given by

$$F1 = \frac{2PR}{P + R} * 100, \tag{6}$$

Best results of image segmentation algorithms Tab. 2

Segmentation algorithm	P [%]	R [%]	F1 [%]	Computing time [s]
Belief Propagation	55.34	19.47	21.03	34.19
K-Means	43.27	15.13	17.56	34.21
Mean Shift	53.09	21.35	23.12	37.02
Hybrid segmentation	61.49	25.09	27.52	54.33

Next, combination of SIFT-SAD algorithm was tested (see Table 3 and Table 4). The result of this stereo matching process is a disparity map that indicates the disparity for every pixel with corresponding intensity. Quality of disparity map is represented as percentage of pixels with disparity errors (bad matching pixels (see Table 3) [5]:

$$P = \frac{1}{X * Y}\sum_{i=1}^{X}\sum_{j=1}^{Y}\left(\left|d_C(i,j) - d_T(i,j)\right|\right), \quad (7)$$

where $X * Y$ represent the size of the image, d_C is the computed disparity map of the test image and d_T is the truth disparity map.

$$d_T = \frac{fBI_{RES}}{D_T h}, \quad (8)$$

where D_T is ground truth depth map, h is height from the ground plane, $D_T * h$ is ground truth distance, B is baseline between the cameras, I_{RES} is image resolution and f is focal length.

Quantitative evaluation of the proposed method in terms percentage of error rates. Tab.3

	Aloe	Dolls	Reindeer
Graph Cut	4.27	3.59	5.03
Graph Cut + Occlusion	3.83	4.15	4.72
Dynamic Programming	4.35	3.78	4.89
Proposed method (SIFT-SAD)	2.12	1.97	2.75

Figure 2 indicates the computed disparity maps for the three scenes together with the used ground-truth maps. For quantitative evaluation, we examine the performance of several algorithms by their error rates. Table 3 shows the error rate for three test stereo image pairs ("Aloe", "Dolls" and "Reindeer").We see that the proposed disparity estimation procedure has boosted up the perfor-

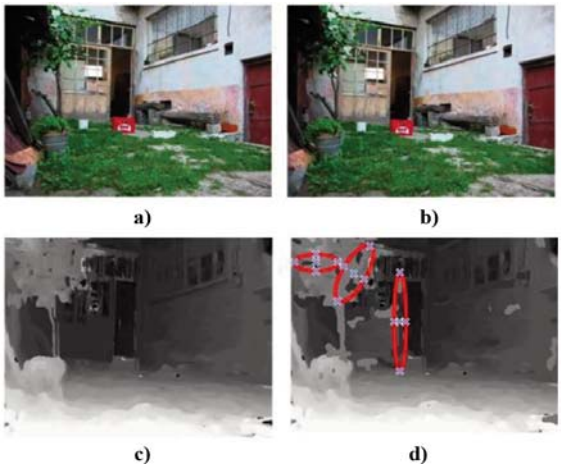

Fig. 3 Results for test "House" stereo image pair: a) left image, b) right image, c) the disparity results of the proposed method, d) the disparity result of [19]

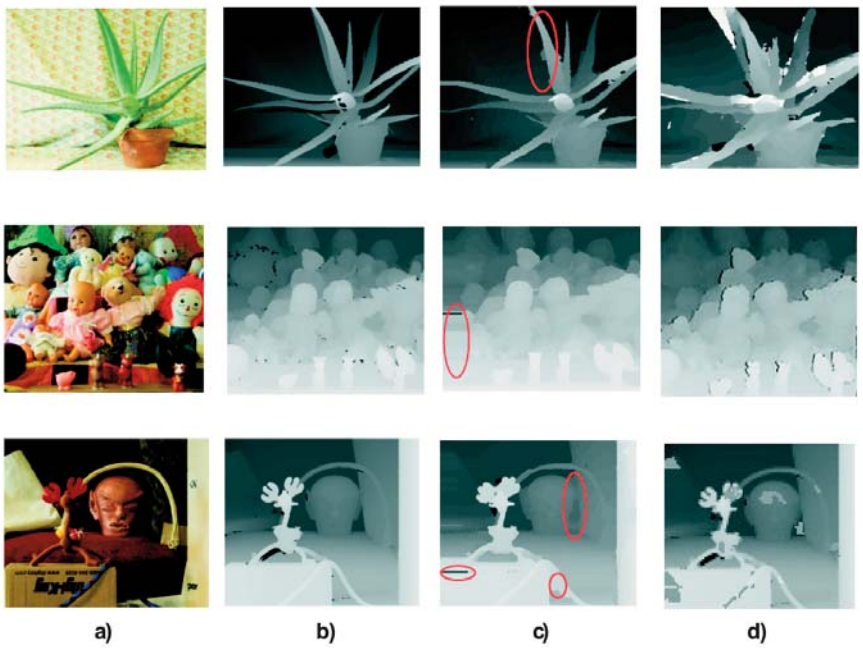

Fig.2 Aloe, Dolls and Reindeer, a) original images, left view; b) ground-truth referring to the left view with black labeled occlusions; c) computed disparity maps using hybrid segmentation algorithm and SIFT-SAD, computed disparity maps using only SAD method

mance of our algorithm. The ground truth disparity map is the inverse of the ground truth distance scale by the image resolution and the focal length. Equation (8) shows how to calculate the ground truth disparity map from the depth map [18]. The depth map is a 16 bit map with values ranging from 0 to 1 where the ground plane was at $D = 1$ and the cameras were at $D = 0$. D is distance of object from the camera. Time is counted for the complete processing which includes feature detecting and matching. Table 3 shows that SURF is the fastest one and SIFT-SAD is the slowest, but it finds most matches.

Although the proposed approach (see Fig. 1) improves the quality of the final disparity map and handles the occlusion, they cannot estimate correct disparity values when the background area behind an object is textureless. This wrong estimation as false matching are called. To give an example, the "Aloe" stereo image pair has many inaccurate disparity regions as illustrated by red circles in Fig. 2, which is obtained by the proposed algorithm. We can see the false matching clearly inside red circles.

In addition, we show another result in Fig. 3 which has the biggest search range of our test images. The size of each left and right images at Fig. 3 (a) and (b) is 800 by 600. The similarity measure was computed with a minimum quadratic 3×3 correlation window because of the high amount of texture in every scene.

The percentage of disparity found correctly, disparity error and the detected occlusion that are correct. Tab. 4

	SIFT	ASIFT	SURF	SIFT-SAD
Disparity correct [%]	86.69	82.07	89.78	92.35
Disparity error [%]	13.31	17.93	10.22	7.65
Occlusion correct [%]	67.45	65.32	72.76	72.03
Total matches	125	135	89	312
Total time [s]	5.52	5.07	2.78	4.95

Table 4 shows summary of overall performance. We compared performances obtained by the proposed method SIFT-SAD with those obtained by three common algorithms (SIFT, ASIFT and SURF). Approximately 92 percent of the disparity values were found correctly for our proposed algorithm. The final disparity map is labeled as correct if it is within one pixel of the correct disparity. Our result in Fig. 3 (c) successfully detects false matching areas and assigns more accurate disparity vector in occlusion regions than the result generated by [19] at Fig. 3 (d) (red circles). For example, our algorithm removes the false matching area around the door and tree which are most complexes regions of the stereo images.

5. Conclusion

The method for reconstructing a 3D scene and proposed algorithm for disparity map measurement from two input images was presented. This algorithm uses image segmentation and SIFT-SAD feature point detection method which extracts more key-points than other feature extraction methods such as SIFT, SURF or ASIFT. The proposed system is based on 3D reconstruction solution using stereo images. This system works with common cameras. The applications of these methods of 3D picture processing are very useful in sphere of medicine, for example detection and identification of tumor in brain and also in other branches as physics, biology or astronomy. In the future we could speed up computation time, improve precision of the hybrid algorithm and apply these methods in real situations.

Acknowledgements

This work was supported by the Slovak Science Project Grant Agency, Project No. 1/0705/13 "Image elements classification for semantic image description" and by project "Competence Center for research and development in the field of diagnostics and therapy of oncological diseases", ITMS: 26220220153, co-funded from EU sources and European Regional Development Fund.

„We support research activities in Slovakia/This project is being co-financed by the European Union"

References

[1] SUN, J., ZHANG, N. N., SHUM, H. Y.: *Stereo Matching Using Belief Propagation,* IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(7):787–800, 2003.

[2] YEDIDA, J. S., FREEMAN, W. T., WEISS, Y.: *Understanding Belief Propagation and it is Generalizations,* Exploring Artificial Intelligence in the New Millennium, Chap. 8, pp. 239–236, January 2003.

[3] DONG, L., OGUNBONA, P., LI, W., YU, G., FAN, L., ZHENG, G: *A Fast Algorithm for Color Image Segmentation,* First Intern. Conference on Innovative Computing, Information and Control, 2006.

[4] YIN, Z., COLLINS, R.: *Belief Propagation in a 3D Spatio - Temporal MRF for Moving Object Detection,* Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA, 2007.
[5] SCHARSTEIN, D., SZELISKI, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *Intern. J. of Computer Vision,* 47(1), pp. 7–42, 2002.
[6] KANADE, T., OKUTOMI, M.: *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment,* PAMI, 16(9), pp. 920–932, 1995.
[7] VEKSLER, O.: *Stereo Matching by Compact Windows via Minimum Ratio Cycle,* ICCV, 2002.
[8] HSIEH, Y. C., McKEOWN, D., PERLANT, F. P.: *Performance evaluation of scene registration and stereo matching for cartographic feature extraction,* IEEE TPAMI, 14(2), pp. 214–238, 1995.
[9] MULLIGAN, J., ISLER, V., DANIILIDIS, K.: Performance Evaluation of Stereo for Tele-Presence, *In ICCV*, vol. 2, pp. 558–565, 2002.
[10] SZELISKI, R., ZABIH, R.: An Experimental Comparison of Stereo Algorithms, *In International Workshop on Vision Algorithms,* Springer, pp. 1–19, 2000.
[11] KUHL, A.: *A Comparison of Stereo Matching Algorithm for Mobile Robots,* Centre for Intelligent Information Processing System, Western Australia, pp. 4–24, 2005.
[12] KAMENCAY, P., BREZNAN, M., JARINA, R., LUKAC, P., ZACHARIASOVA, M.: Improved Depth Map Estimation From Stereo Images Based On Hybrid Method, *The Radioengineering J.,* vol. 21, No. 1, April 2012, ISSN 1210-2512.
[13] LOWE, D. G.: Distinctive Image Feature from Scale Invariant Key Points, *Intern. J. of Computer Vision (IJCV),* vol. 2, No. 60, 2004, pp. 91–110.
[14] JUAN, L., GWUN, O.: A Comparison of SIFT, PCA-SIFT and SURF, *Intern. J. of Image Processing (IJIP),* vol. 3, No. 4, pp. 143–152.
[15] KE, Y., SUKTHANKAR, R.: *PCA-SIFT: A More Distinctive Representation for Local Image Descriptors,* Proc. Conf. Computer Vision and Pattern Recognition, pp. 511–517, 2003.
[16] Middlebury Stereo image pairs dataset. [online 10.10.2012] Available on the internet. http://vision.middlebury.edu/stereo/data/.
[17] ZACHARIASOVA, M., HUDEC, R., BENCO, M., KAMENCAY, P., LUKAC, P., MATUSKA, S.: The Effect of Metric Space on The Results of Graph Based Colour Image Segmentation, *Communications - Scientific Letters of the University of Zilina,* vol. 14, No. 3, 2012, ISSN 1335-4205.
[18] BOLECEK, L., RICNY, V., SLANINA, M.: *Fast Method for Reconstruction of 3D Coordinates,* 35th Intern. Conference on Telecommunications and Signal Processing (TSP 2012), July 2012, ISBN 978-1-4673-1115-1.
[19] KOLMOGOROV, V., ZABIH, R.: *Computing Visual Correspondence with Occlusions using Graph Cuts,* Proc. IEEE Int. Conference Computer Vision 2, pp. 508–515, 2002.