Accomplishments and Challenges of Computer Stereo Vision

Miran Gosta [1], Mislav Grgic [2]
[1] Croatian Post and Electronic Communications Agency (HAKOM)
Broadcast and Licensing Department, Jurisiceva 13, HR-10000 Zagreb, Croatia
[2] University of Zagreb, Faculty of Electrical Engineering and Computing
Department of Wireless Communications, Unska 3, HR-10000 Zagreb, Croatia
miran.gosta@gmail.com

***Abstract* - Segmentation and grouping of image elements is required to proceed with image recognition. Due to the fact that the images are two dimensional (2D) representations of the real three dimensional (3D) scenes, the information of the third dimension, like geometrical relations between the objects that are important for reasonable segmentation and grouping, are lost in 2D image representations. Computer stereo vision implies on understanding information stored in 3D-scene. Techniques for stereo computation are observed in this paper. The methods for solving the correspondence problem in stereo image matching are presented. The process of 3D-scene reconstruction from stereo image pairs and extraction of parameters important for image understanding are described. Occluded and surrounding areas in stereo image pairs are stressed out as important for image understanding.**

***Keywords* - Computer Stereo Vision; Epipolar Rectification; Correspondence Problem; Disparity; Depth Calculation; Occlusion; Depth Discontinuities**

I. Introduction

The computer vision is defined as technology concerned with computational understanding and use of the information present in visual images [1]. Visual images are commonly considered as two dimensional representations of the real (3D) world. Computer stereo vision implicates to acquisition of images with two or more cameras horizontally displaced from each other. In such a way different views of a scene are recorded and could be computed for the needs of computer vision applications like reconstruction of original 3D scene. Stereo image pairs definitely contain more information about the real world than regular 2D images, for example scene depth, object contours, surface orientation and creases.

Artificial intelligence systems like computer vision try to imitate the mechanisms performed in human (visual) system and in human brain so the accomplishments in neuroscience and psychology should be used when researching computer vision.

A scene pictured with two horizontally and exactly displaced cameras will obtain two slightly different projections of a scene. If comparing these two images, additional information, like the depth of a scene, could be reached. This process of extraction of three dimensional structure of a scene from stereo image pairs is called computational stereo [2].

Computational stereo includes the following calculations on images:

1. image rectification,
2. image matching,
3. calculation of disparity map,
4. handling occlusion.

The outline of this paper is as follows. Section 2 describes the epipolar rectification of the stereo image pair. In Section 3 principle of depth calculation from stereo images is described, while in Section 4 principles of solving correspondence problem are considered. A survey of stereo algorithms with emphasis on their characteristics is given in Section 5. Section 6 describes the challenges of using stereo vision in image and video segmentation. Conclusions are given in Section 7.

II. The Geometry of Stereo Vision

Problem of matching the same points or areas in stereo image pairs is called the correspondence problem. Image pair matching could be performed by the algorithms that include search and comparison of the parts of two images.

If the camera geometry is known, the two dimensional search for corresponding points could be simplified to one dimensional search. This is done by rectification of images which is based on epipolar geometry (epipolar rectification) [3].

Epipolar rectification is shown in Fig. 1 [4]. If there are two pinhole cameras with the optical centers C_l and C_r, the scene point P will be projected onto the left (L) and right (R) image plane as p_l and p_r.

Any other possible scene point on a ray P-C_l will be projected on a line p_r-e_r which is called epipolar line of p_l, and where e_r is called epipole which geometrically represents picture of left optical center C_l in the right camera. If the same is done on left camera, rectification could be obtained by transforming the planes L and R in a way that p_l-e_l and p_r-e_r form single line in the same plane.

If all the scene points are transformed in the same way, the result will be that the same elements are situated on the same

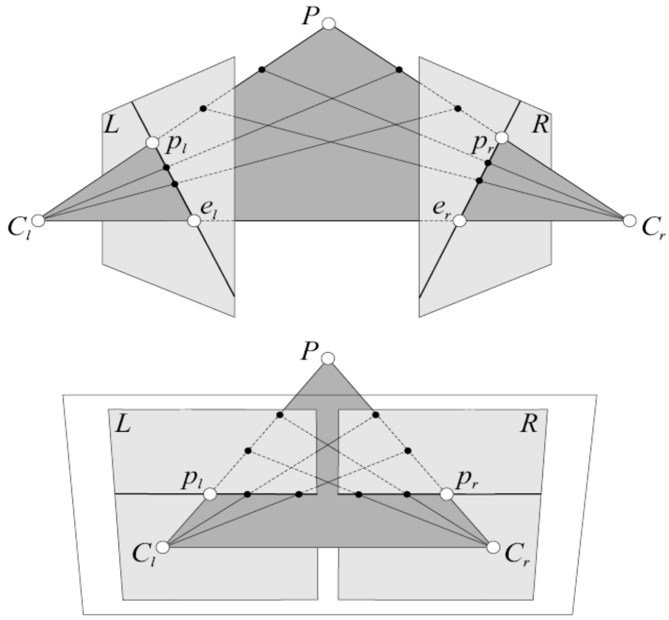


Figure 1. Epipolar rectification

horizontal line in left and right image, Fig. 2, [5] and the correspondence problem should be solved only in one (axis) direction:

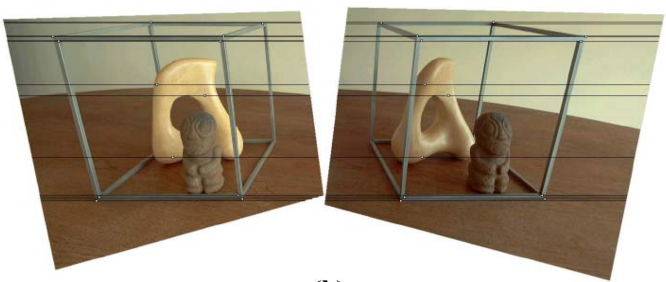
$$x_r = x_l + d\,, \quad \text{while} \qquad y_r = y_l \tag{1}$$

Actually, the value d is horizontal offset (displacement) between the two corresponding pixels and is called disparity. Disparities could be calculated for all image points so that disparity map is constructed.

III. Depth Calculation

The epipolar rectification assumes like the cameras are parallel to each other and that they have the identical focal lengths (f). If the geometry of the two camera system is known,

(a)



(b)

Figure 2. (a) Original image pair; (b) Rectified image pair

the baseline length C_l-C_r is a constant value (B). Comparing the triangles (triangulation), distance to an object point (X,Y,Z) or depth (Z) could be determined [6]:

$$\frac{X}{Z} = \frac{x_l}{f} \qquad \text{and} \qquad \frac{X-B}{Z} = \frac{x_r}{f} \tag{2}$$

which can be derived into:

$$Z = \frac{B \cdot f}{x_l - x_r} = \frac{B \cdot f}{d} \tag{3}$$

IV. Stereo Matching Problems

Although finding the corresponding points of the left image on the right image is simplified, there are still some specific matching problems in stereo vision.

Missing of photo consistency between the images is common due to the facts that intensity and colors vary depending on the viewpoint. Intensity and colors in images in a pair could also vary due to the different camera sensors characteristics. Additionally, camera electronics produces noise that affects image acquisition. Usually all these differences are quite small so can be neglected and the photo consistency is assumed as constraint in stereo matching algorithms.

There is also a problem in uniquely matching two points due to the fact that large regions with constant luminance exist (e.g. untextured regions or repetitive patterns), and in such regions more than one corresponding point could be identified. If there are repetitive patterns the more than one unique corresponding point will be identified.

The biggest problem in stereo matching arrives due to the fact that for some pixels in the left image the corresponding pixel in the right image does not even exist. This is the consequence of the fact that some scene parts are visible to one camera but occluded to other camera due to the obstacles, Fig. 3, [2]. If there is no corresponding pixel, the depth calculation and 3D reconstruction is impossible for that pixel.

Stereo matching algorithms commonly use constraints that provide acceptable results in order to solve stereo correspondence problem. These constraints assume the following:

1. Epipolar constraint,
2. Photo consistency,
3. Smoothness,
4. Uniqueness,
5. Ordering.

Smoothness means that disparity of neighboring points varies smoothly, except at depth boundaries. The smoothness assumption is a consequence of observing natural objects and cognition that their surfaces are smooth. The smoothness assumption in algorithms is very effective on scenes that contain compact objects, but is ineffective when computing scenes with thin fine structured shapes (e.g. grass or hair).

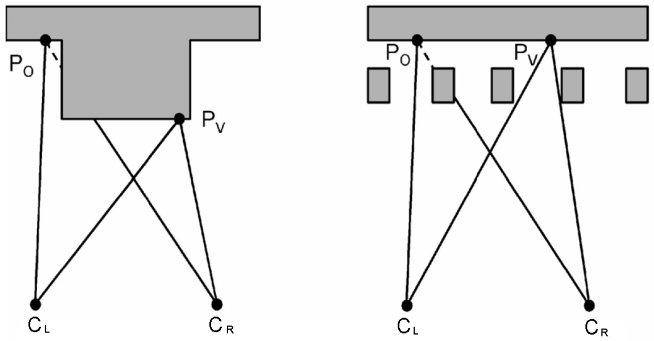


Figure 3. Two examples of occlusions: P_V visible points to both cameras; P_O - occluded points to right camera

Smoothness is applied in different algorithms in implicit (local correspondence methods) or explicit way (global correspondence methods).

Uniqueness of correspondence is defined in a way that a pixel of one view corresponds to only one pixel of other view. If a pixel of one view corresponds to no pixel of other view it is interpreted as occluded in other view [7]. The uniqueness constraint is effective with opaque surfaces. Uniqueness can not be provided when computing transparent surfaces due to the fact that the depth could be provided for two surfaces - the foreground transparent surface and background surface which is visible because the front surface is transparent. Uniqueness of correspondence is also lost if slanted surfaces are observed, because the projections of the same 3D line results with different line lengths, Fig. 3, [8]. If there are different lines lengths in image pair it is obvious that one pixel will correspond to more pixels in the image with wider line length.

The ordering assumption can be explained on an example: if there are scene points P_1 and P_2, and in the left image projection p_{1L} appears left to point p_{2L}, then ordering assumption claims that in the right image projection, p_{1R} should appear left to p_{2R}. The ordering assumption exists in most of the real scene cases with the exceptions of the scenes containing thin foreground objects that could reverse order with certain background pixels.

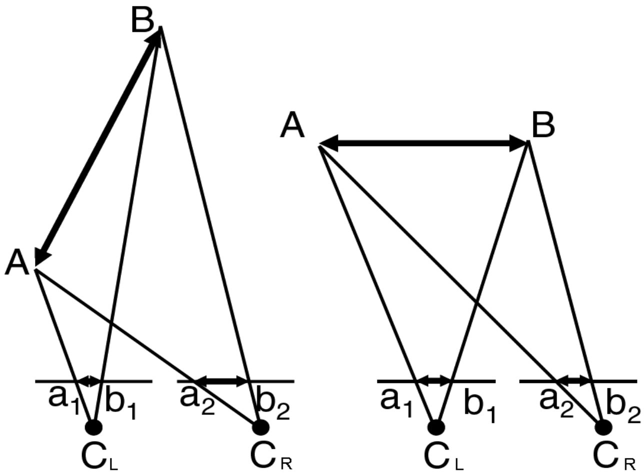


Figure 4. (a) Two different-length projections of horizontally slanted line; (b) Two similar-length projections of straight line

V. Stereo Algorithms

A lot of research in stereo vision is done in solving the correspondence problem. A large number of algorithms for stereo correspondence have been developed, and it is possible to evaluate a new algorithm and compare its characteristics [9] using the software available on the Middlebury College web-site [10].

When analyzing the scene objects, the surface interiors and the boundaries of the objects become the important issue. At the object surface the depth is very smooth, but at the object boundary the depth is non-smooth so the depth discontinuity appears. It is suggested in [11] that depths, surface orientation, occluding contours and creases should be estimated simultaneously when the stereo algorithm is designed. Such an approach enables the recovery of the surface interiors by estimating their depth and orientation, and recovery of the surface boundaries by estimating occluding contours and creases.

In [12] a three-axis categorization of binocular stereo algorithms according to their interpretation of continuity and uniqueness is suggested. The axes are:

1. continuity - over disparity values within smooth surface patches,
2. discontinuity - at the boundaries of smooth surface patches,
3. uniqueness - to the occlusions that accompany depth discontinuities.

The disparity values within smooth surface patch can be constant, can be discrete but numerically as close as possible to the neighboring pixels, and can vary over the real numbers.

The discontinuity at the boundaries of smooth surface patches can be differently penalized. Discontinuity might not be specifically penalized (free), or might be penalized infinitely (disallowed). If the discontinuities are allowed, they could be penalized as a finite, positive, convex function of the size of the jump of the discontinuity (convex), or could be penalized as a non-convex function of the size of the jump of the discontinuity (non-convex). Common convex functions are the square or the absolute values of the size of the jump of the discontinuity. The typical non-convex functions of the size of the jump of the discontinuity are statistical transformation functions.

The uniqueness constraint can not be applied on the transparent and occluded regions. As the occlusion region is defined as an image region where there is no disparity defined, algorithms differ in a way how they treat occlusion regions. For transparent image regions uniqueness should not be assumed. The uniqueness to the occlusions that accompany depth discontinuities can be one-way assumed, asymmetric two-way or symmetric two-way assumed. If uniqueness is one-way assumed, then, to every pixel in the reference image, single disparity is assigned, but each disparity can be pointed by multiple pixels from the second image. If uniqueness is two-way assumed, then one same disparity is pointed by pixels from both images of a stereo pair. Uniqueness is asymmetric if it is encouraged by both images, but only one image is used as reference, while uniqueness is symmetric if it is enforced by both equally treated images.

A. Window-based correlation

The idea of matching corresponding pixels in image stereo pair could be simply implemented in a way that each pixel from left image is matched with the pixel from the right image that has the same color or the most similar color. This point-based correspondence method does not consider any smoothness and continuity and in practice results with a lot of false matched pixels because a large number of candidate pixels that satisfy the color similarity condition exist.

The simplest implementation of smoothness assumption implies that neighboring pixels have similar disparities. If the windows are formed (e.g. 8×8 pixel windows) and slid through the images, Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), Normalized Cross-correlation (NCC) or other measure model could be used for matching windows. Instead of calculating disparity for every pixel, the disparity is calculated for all pixels (x, y) inside the window and a three-dimensional structure (x, y, d) called disparity space image (DSI) could be formed. The smoothness is implicitly assumed to be locally constant inside the window.

The main challenge with windowed correspondence is selection of the window size. If the windows are too small, then wrong matches are more possible like for pixel-to-pixel matching. On the other hand, with the larger window sizes there are less wrong matches, but the local treatment of smoothness (constant disparity inside the window) is not correct. The better results could be reached with adaptive windows where smaller windows are used near discontinuities, and larger windows are used away from discontinuities.

B. Cooperative methods

The improvement of window-based methods could be reached with additional analysis of window-based results. Cooperative methods are based on iterative procedure of updating the window-based matching results or the DSI values.

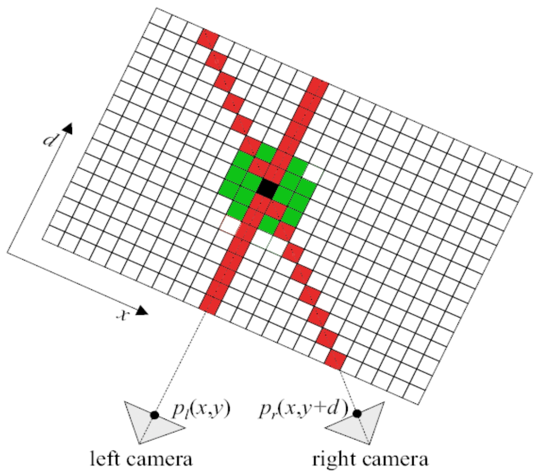


Figure 5. Cooperative approach: current match (black), support region (green), inhibited area (red)

TABLE I. OVERVIEW OF THE STEREO MATCHING ALGORITHMS

Stereo matching method	Reference	Characteristics
Window-based correlation	Hirschmuller, Innocent, Garibaldi [13] Kanade, Okutomi [14] Scharstein, Szeliski [15] Fua [16] Fusiello, Roberto, Trucco [17] Veksler [18]	• Smooth and textured regions are well treated • One-way uniqueness is applied • Discontinuties are not specifically penalized • Blur results across discontinuities • Poor results with untextured regions • Efficient in computation
Cooperative methods	Scharstein, Szeliski [9] Zhang, Kambhamettu [19] Zitnick, Kanade [20] Marr, Poggio [21] Mayer [22]	• Support discontinuities with non-convex penalties • Asymmetrically two-way uniqueness is encouraged • Results are dependable on initialization • Blurred discontinuties • High computational effort
Dynamic programming	Baker, Binford [23] Belhumeur [24, 11] Belhumeur, Mumford [25] Cox, Hingorani, Rao, Maggs [26] Dhond, Aggarwal [27] Geiger, Ladendorf, Yuille [28] Intille, Bobick [29] Bobick, Intille [30] Ohta, Kanade [31] Birchfield, Tomasi [32]	• One dimensional discrete approach to the smoothness (inside the scanlines) • Dependent upon the ordering constraint • Errors with low texture regions (horizontal streaks) and thin foreground objects (wrong ordering) • Efficient in computation
Graph-based methods	Boykov, Veksler, Zabih [33, 34, 35] Ishikawa, Geiger [36] Kolmogorov, Zabih [7] Roy, Cox [37] Veksler [38]	• Powerfull and efficient stereo algorithms • Discrete approach to the smoothness

This is done by cooperatively assuming the smoothness by creating the support region and the uniqueness by creating the inhibition area.

Firstly, the support region around the current DSI value is created. The smoothness is extended and aggregated to the entire support region, e.g. by computing the average value. In this way, the DSI values with higher weight (matching score) are propagated.

In the next step, the inhibition area is registered. It is supposed that if the current match is correct, than according to uniqueness assumption all other matches at the lines of sight of left and right camera can not be considered any more so they are inhibited for further matching, Fig. 5, [4].

Cooperatively, the weight of the strongest match is increased to neighboring support region and potential matches on the line of sight are reduced.

The whole process is done to update all scores in the DSI and then iterated until the convergence. After reaching convergence new real-valued DSI weights are compared with one another and threshold to determine final correspondences is set. If the weight is bellow the threshold, the area is classified as occluded. Although cooperative algorithms give good results, the depth boundaries are blurred due to the rectangular support regions.

C. Dynamic programming

Matching algorithms based on dynamic programming do not treat smoothness locally as window-based and cooperative methods. Instead, smoothness is globally treated, which in dynamic programming concretely means that continuity is considered inside the horizontal scanlines. The new constraint introduced in dynamic programming algorithms is the ordering constraint.

After the DSI is calculated, two horizontal scanlines are analyzed as shown in Fig. 6. Each cell in Fig. 6 shows the possibility of matching the two pixels. The path that connects opposite corners C_s and C_e is built in a way that matching of the two neighboring pixels is presented with diagonal movement, and occlusion is presented with horizontal (occluded pixel in right scanline) or vertical movement (occluded pixel in left scanline). Some of the pixels are prohibited due to the ordering constraint (left pixels) or limitation of maximum disparity. The optimal path between C_s and C_e will be achieved if every path between the starting cell C_s and some intermediate cell C_i is also minimal or the path of lowest cost. Recursively, optimal paths to every cell can be calculated.

In stereo matching algorithms, the matching cost function is defined to perform actual matching. The matching cost function should be minimal for a correct match, and large for an unlikely match.

The idea with dynamic programming is to decompose complex optimization into simpler sub-problems. When building the optimal path, the costs of including a diagonal move are the matching costs, while the costs of including vertical and horizontal moves are the cost for occlusions and can be used for penalizing occlusions. The optimal predecessor of a cell can be calculated as the one whose costs plus the cost of the move joining two cells are lowest. In such a way every cell can record the costs and the link to its optimal predecessor.

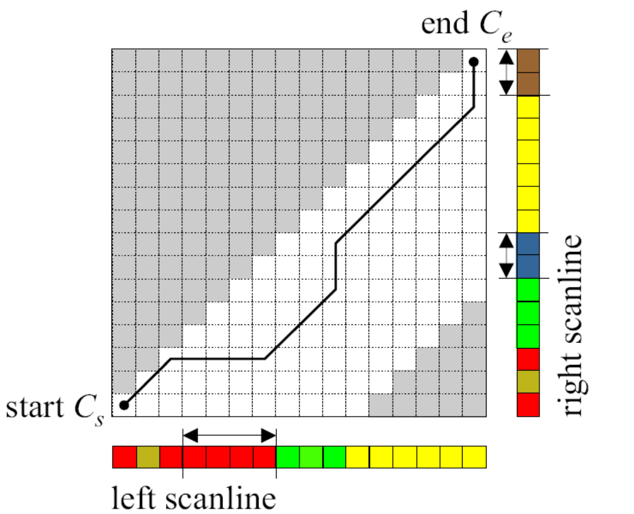


Figure 6. Dynamic programming - finding the minimum cost path through a disparity space image (DSI) (grey colored cells are prohibited)

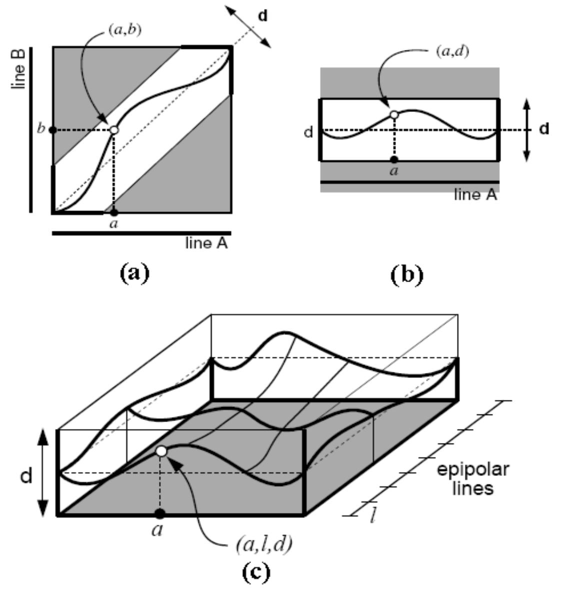


Figure 7. Matching whole images

Dynamic programming algorithms are characterized as computationally efficient and able to explicitly identify occlusions in left and right image. Their main weakness is the consequence of one-dimensional treatment of smoothness.

D. Graph-based algorithms

Similar like in dynamic programming the graph-based algorithms treat smoothness global, but instead of optimizing continuity only over the scanlines (one-dimensional), continuity is optimized over the entire image (two-dimensional).

As shown in Fig. 7(a) [37], the same principle like in dynamic programming, but without ordering constraint, is used to create the path that connects opposite corners of the epipolar

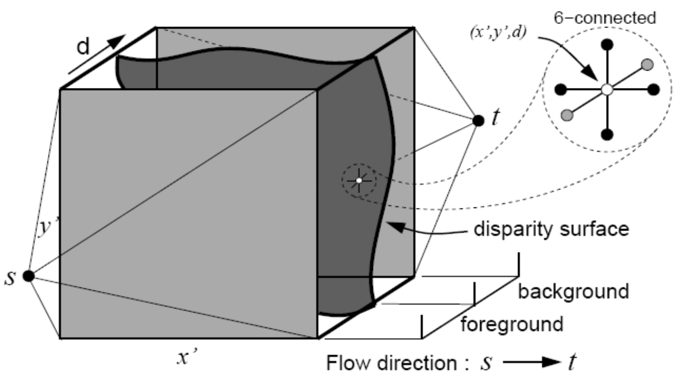


Figure 8. Graph cut - finding the minimum cost surface through the network nodes

lines (left scanline is presented with line A and right scanline is presented with line B). In Fig. 7(b) the same is presented but using only the reference line A. In Fig. 7(c) all epipolar lines are stacked together like in reference image. The problem of finding the minimum-cost paths which define the matching of every single epipolar line is now assembled into finding single minimum-cost surface. This surface will contain the disparity information of entire reference image so is called disparity surface. Due to the smoothness, the local coherence is considered as constraint in a way that disparities are locally very similar. Dynamic programming is unusable in the new surface-environment, but the graph theory can be used to calculate minimum-cost surface [37]. With the graph theory, the minimum cut problem is turned to the maximum flow problem.

The directed graph *G* is created like presented in Fig. 8 [37]. The source node *s* and the sink node *t* are connected through the network nodes, where every node is 6-connected with the other network nodes. Four of six connections are made with other image pixels (x',y'), while two connections are made with the corresponding pixel disparity weights. In Fig. 8 the disparity surface cuts the graph into two parts. The task of graph based algorithms is to find the minimum cut of the graph *G* and so to solve the correspondence problem. If a single image pixel $p_1(x',y')$ is observed like in Fig. 9(a) and disparity is limited in range from zero to three, then disparity of minimum cost of a pixel p_1 is derived by computing the minimum cut.

If we imagine the graph *G* as the water pipe network, than we can define the capacity of pipes as the maximum amount of water that can flow through them. If we calculate the maximum flow through the edge defined by graph cut, then the edge capacity will be known and it is equal to the cost of the minimum cut.

In Fig. 9(b) minimum cut, which is simultaneously calculated for six pixels p_1-p_6, is presented with red-dotted line, while the blue-solid line presents the corrected cut after establishing connections with neighboring pixels. When connections with neighboring pixels are established, the smoothness could be assumed to improve the results. Actually, the edge can be either in $x'y'$-plane or on *d*-axis. If the edge is

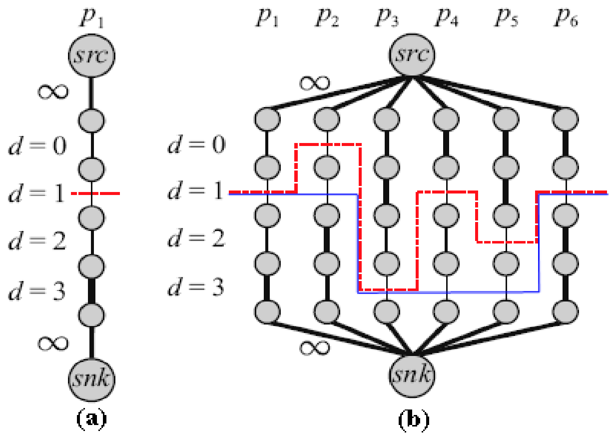


Figure 9. (a) Finding disparities for single pixel using graph cut; (b) Finding disparities for six pixels of a line using graph cut

oriented along the *d*-axis, then it is called disparity edge, while the edge oriented along $x'y'$-plane is called occlusion edge. The smoothness is directly connected with the costs of occlusion edges and is described by constant user-defined smoothness parameter λ. If two neighboring points' disparities d_1 and d_2 differ by $d_1 - d_2$ pixels, the minimum cut must include $|d_1 - d_2|$ smoothness edge. The smoothness could be described with the non-convex function (smoothness parameter is λ when neighboring points have different disparities or 0 (zero) when they have same disparities).

VI. Recognition and Stereo Image Pairs

Stereo parameters that are important for object recognition process are also researched. The half-occluded and surrounding areas are very important for the image recognition process.

Psychological, physiological and biomedical researches on binocularity [39, 40] developed the theory of the neural mechanisms for processing binocularity in humans and animals. The neural theory of binocular rivalry [41] identifies special neurons in cortex that are responsible for binocular processing and the other neurons for monocular processing of visual information. The eye movement also helps in fine tuning the correspondence and has an important role in stereo perception [42].

The 2½ D sketch model is suggested in [43], as viewer centered representation of the depth, orientation and discontinuities of the visible surfaces. The 2½ D sketch model is simplified representation of a 3D scene which contains the fundamentals for the recovery of three-dimensional, object centered description of object shapes and their spatial organization in images. The model is criticized in [44] due to the viewer centered limitations and so caused individuality which is not appropriate for solving general vision problem. Instead, layered representation of visible curves and surfaces based on tensors is suggested.

With a layered representation, image is segmented into smaller parts - layers. Each layer can then be excluded from the rest of image and used for different applications such as motion analysis, object identification, coding algorithms, or background replacement in video applications. When looking the natural scene, different objects contained in a scene are differently situated inside the scene and depth information is useful if we wish to segment the image into layers presenting reasonable objects.

The algorithms that are capable to perform depth-segmentation could perform layer extraction as a part of stereo algorithm or as independent post stereo algorithm [45]. Layer-model could be used to improve the stereo algorithm in a way

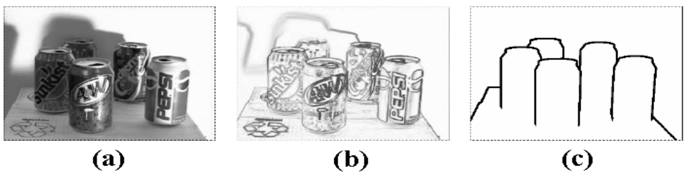


Figure 10. (a) Original picture; (b) Detected edges - based on intensity; (c) Depth discontinuities

that left and right images are divided into regions and instead of matching corresponding pixels, matching is done with similar regions [46-50]. The result of this is the disparity map that is piecewise-smooth, like the natural objects are really. These layered based methods are quite effective, because they support occlusion regions, two way uniqueness and real-value disparities together with the property to extract layers from the resulting disparity map.

Automatic separation of layers from stereo alone is typically error prone, so the segmentation is done in cooperation of stereo vision with:

- motion extraction [51, 52],
- color extraction [53-58],
- motion and color extraction [59].

The importance of half-occluded and surrounding areas for the processes which preserve object boundaries is studied in [30] and the depth discontinuities are identified as more closely tied to the geometry of the scene than intensity edges, Fig. 5, [60].

VII. CONCLUSION

Using stereo image pairs as representations of a scene, stereo information of a visible scene can be extracted. The epipolar image rectification simplifies the matching of corresponding image parts to one dimensional search and one dimensional matching. Finding corresponding pixels in two images, displacements or disparities could be calculated and depth can be extracted from disparities. The process is possible for all image pixels that are correctly matched in image pairs, but depth is lost for unmatched pixels. Matching of some pixels is impossible due to obstacles that are seen only to one camera, so some image regions are occluded in the one image of a stereo image pair. Occlusions are identified to be in a connection to discontinuities of a depth. The algorithms that can handle the occlusions even during the matching process are developed.

The biggest challenge of stereo image segmentation is to define exact and fine layer contour, which is quite difficult and depends on the treatment of object boundaries and occlusions by the stereo algorithm. Besides the layered methods, the best results are achieved with algorithms that globally treat smoothness and exactly detect occlusions like dynamic programming and graph-cut algorithms. Further developments and improvements of developed algorithms can be researched. Research should be based on state of the art Kolmogorov et al. algorithm [60] and extensions to multi-layer segmentation.

ACKNOWLEDGMENT

The work in this paper was conducted under the research project "Intelligent Image Features Extraction in Knowledge Discovery Systems" (036-0982560-1643), supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

REFERENCES

[1] "Encyclopedia of Science and Technology", McGraw-Hill, 2009, pp. 594-596

[2] M. Z. Brown, D. Burschka, G. D. Hager, "Advances in Computational Stereo", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, 2003, pp. 993-1008

[3] Z. Zhang, T. Kanade, "Determining the Epipolar Geometry and its Uncertainty: A Review", International Journal of Computer Vision, vol. 27, 1998, pp. 161-195

[4] M. Bleyer, "Segmentation-based Stereo and Motion with Occlusions", Vienna University of Technology, Ph.D. dissertation, 2006

[5] C. Loop, Z. Zhang, "Computing Rectifying Homographies for Stereo Vision", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 1999

[6] R. Szeliski, "Computer Vision: Algorithms and Applications", Springer, 2010 (in press)

[7] V. Kolmogorov, R. Zabih, "Computing visual correspondence with occlusions using graph cuts", Proceedings. Eighth IEEE International Conference on Computer Vision, vol. 2, 2001, pp. 508-515

[8] A. S. Ogale, Y. Aloimonos, "Stereo correspondence with slanted surfaces: critical implications of horizontal slant", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2004. pp. 568-573

[9] D. Scharstein, R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", International Journal of Computer Vision, Springer Netherlands, vol. 47, 2002, pp. 7-42

[10] D. Scharstein, R. Szeliski, Middlebury Stereo Vision Page, available: http://vision.middlebury.edu/stereo

[11] P. N. Belhumeur, "A Bayesian approach to binocular stereopsis", International Journal of Computer Vision, vol. 19, 1996, pp. 237-260

[12] M. Lin, "Surfaces with occlusions from layered stereo", Stanford University, Ph.D. dissertation, 2002

[13] H. Hirschmuller, P. R. Innocent, J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors", International Journal of Computer Vision, vol. 47, 2002, pp. 229-246

[14] T. Kanade, M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 16(9), 1994, pp. 920-932

[15] D. Scharstein, R. Szeliski, "Stereo matching with nonlinear diffusion", International Journal of Computer Vision, vol. 28(2), 1998, pp. 155-174

[16] P. V. Fua, "Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities", International Joint Conference on Artificial Intelligence, 1991, pp. 1292-1298

[17] A. Fusiello, V. Roberto, E. Trucco, "Efficient stereo with multiple windowing", Conference on Computer Vision and Pattern Recognition, 1997, pp. 858-863

[18] O. Veksler, "Stereo correspondence with compact windows via minimum ratio cycle", IEEE Transactions on Pattern Analysis and Machine In-telligence, vol 24(12), 2002, pp. 1654-1660

[19] Y. Zhang, C. Kambhamettu, "Stereo matching with segmentation-based cooperation", Proceedings of the European Conference on Computer Vision, vol. 2, 2002, pp. 556-571

[20] C. L. Zitnick, T. Kanade "A cooperative algorithm for stereo matching and occlusion detection" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 22(7), 2000, pp. 675-684

[21] D. Marr, T. Poggio, "Cooperative computation of stereo disparity", Science 194, 1976, pp. 283-287

[22] H. Mayer, "Analysis of means to improve cooperative disparity estima-tion", International Archives ofthe Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34(Part 3/W8), 2003, pp. 25-31

[23] H. H. Baker, T. O. Binford, "Depth from edge and intensity-based stereo", Proceedings of the International Joint Conferences on Artificial Intelligence, 1981, pp. 631-636

[24] P. N. Belhumeur, "A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion"

Proceedings of the IEEE International Conference on Computer Vision, 1993, pp. 431-438

[25] P. N. Belhumeur, D. A. Mumford, "A Bayesian treatment of the stereo correspondence problem using half-occluded regions", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1992, pp. 506-512

[26] I. J. Cox, S. L. Hingorani, S. B. Rao, B. Maggs, "A maximum likelihood stereo algorithm", Computer Vision and Image Under-standing, vol. 63(3), 1996, pp. 542-567

[27] U. R. Dhond, J. K. Aggarwal, "Stereo matching in the presence of narrow occluding objects using dynamic disparity search", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 17(7), 1995, pp. 719-724

[28] D. Geiger, B. Ladendorf, A. Yuille, "Occlusions and binocular stereo", International Journal of Computer Vision, vol. 14(3), 1995, pp. 211-226

[29] S. S. Intille, A. F. Bobick, "Disparity-space images and large occlusion stereo", Proceedings of the European Conference on Computer Vision, vol. 2, 1994, pp. 179-186

[30] A. F. Bobick, S. S. Intille, "Large Occlusion Stereo", International Journal of Computer Vision, Springer Netherlands, vol. 33, 1999, pp. 181-200

[31] Y. Ohta, T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7(2), 1985, pp. 139-154

[32] S. Birchfield, C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo", International Journal of Computer Vision, vol. 35(3), 1999, pp. 269-293

[33] Y. Boykov, O. Veksler, R. Zabih, "Markov random fields with efficient approximations", Proceedings ofthe IEEE Conference on Computer Vision and Pattern Recognition, 1998, pp. 648-655

[34] Y. Boykov, O. Veksler, R. Zabih, "Approximate energy minimization with discontinuities", Proceedings of the IEEE International Workshop on Energy Minimization Methods in Computer Vision, 1999, pp. 205-220

[35] Y. Boykov, O. Veksler, R. Zabih, "Fast approximate energy minimization via graph cuts", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23(11), 2001, pp. 1222-1239

[36] H. Ishikawa, D. Geiger, "Occlusions, discontinuities, and epipolar lines in stereo", Proceedings of the European Conference on Computer Vision, vol. 1, 1998, pp. 232-249

[37] S. Roy, I. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem, Proceedings of the IEEE International Conference on Computer Vision, 1998, pp. 492-499

[38] O. Veksler, "Efficient Graph-Based Energy Minimization Methods in Computer Vision". Cornell University, Ph.D. dissertation, 1999.

[39] A. Anzai, I. Ohzawa, R. D. Freeman, "Neural Mechanisms for Processing Binocular Information I. Simple Cells", The Journal of Neurophysiology vol. 82, 1999, pp. 891-908

[40] A. Anzai, I. Ohzawa, R. D. Freeman, "Neural Mechanisms for Encoding Binocular Disparity: Receptive Field Position Versus Phase", Journal of Neurophysiology vol. 82, 1999, pp. 874-890

[41] R. Blake, "A Neural Theory of Binocular Rivalry", American Psychological Association, Inc., Psychological Review, Vol. 96, 1989, pp. 145-167

[42] D. Marr, T. Poggio, "A Computational Theory of Human Stereo Vision", Proceedings of the Royal Society of London. Series B, Biological Sciences, vol. 204, 1979, pp. 301-328

[43] D. Marr, T. Poggio, "A Theory of Human Stereo Vision", Massachusetts institute of technology, Artificial intelligence laboratory, Memo No. 451, 1977.

[44] G. Medioni, M.-S. Lee, C.-K. Tang, "A Computational Framework for Segmentation and Grouping", Elsevier, 2000

[45] D. Han, B. Lee, J. Cho, D. Hwang, "Real-time object segmentation using disparity map of stereo matching", Applied Mathematics and Computation, Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation, vol. 205, 2008, pp. 770-777

[46] S. Baker, R. Szeliski, P. Anandan, "A layered approach to stereo reconstruction", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1998, pp. 434-441

[47] S. Birchfield, C. Tomasi, "Multiway cut for stereo and motion with slanted surfaces", Proceedings of the IEEE International Conference on Computer Vision, 1999, pp. 489-495

[48] M. Lin, C. Tomasi, "Surfaces with occlusions from layered stereo. In Conference on Computer Vision and Pattern Recognition", 2003, pp. 710-717

[49] M. Bleyer, M. Gelautz, "Graph-cut-based Stereo Matching using Image Segmentation with Symmetrical Treatment of Occlusions", Signal Processing: Image Communication, Special issue on three-dimensional video and television, vol. 22(2), 2007, pp. 127-143

[50] M. Bleyer, C. Rother, P. Kohli, "Surface Stereo with Soft Segmentation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, unpublished

[51] S. Wang, X. Wang, H. Chen, "A stereo video segmentation algorithm combining disparity map and frame difference", 3rd International Conference on Intelligent System and Knowledge Engineering, vol. 1, 2008, pp. 1121-1124

[52] Y. Altunbasak, A. M. Tekalp, G. Bozdagi, "Simultaneous motion-disparity estimation and segmentation from stereo," IEEE Int. Conf. Image Processing, 1994, vol. 3, pp. 73-77

[53] S. T. Birchfield, B. Natarajan, C. Tomasi, "Correspondence as energy-based segmentation", Image and Vision Computing, 2007, vol. 25, pp. 1329-1340

[54] R. Ma, M. Thonnat, "Object detection in outdoor scenes by disparity map segmentation", Proceeding of 11th IAPR International Conference on Pattern Recognition, Conference A: Computer Vision and Applications, vol. 1, 1992, pp. 546-549

[55] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, C.Rother, "Bi-layer segmentation of binocular stereo video", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005., vol. 2, 2005, pp. 407-414

[56] N.D.F. Campbell, G. Vogiatzis, C. Hernandez, R. Cipolla, "Automatic 3D object segmentation in multiple views using volumetric graph-cuts", Image and Vision Computing, vol. 28, 2010, pp. 14-25

[57] J. F. Vigueras, M. Rivera, "Registration and interactive planar segmentation for stereo images of polyhedral scenes", Pattern Recognition, Interactive Imaging and Vision, vol. 43, 2010, pp. 494-505

[58] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, C. Rother, "Probabilistic fusion of stereo with color and contrast for bi-layer segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2006, pp. 1480-1492

[59] Q. Zhang, K. N. Ngan, "Multi-view video based multiple objects segmentation using graph cut and spatiotemporal projections", Journal of Visual Communication and Image Representation, Special issue on Multi-camera Imaging, Coding and Innovative Display, vol. 21, 2010, pp. 453-461

[60] S. T. Birchfield, "Depth and motion discontinuities", Stanford University, Ph.D. dissertation, 1999.