REGION-BASED DENSE DEPTH EXTRACTION FROM MULTI-VIEW VIDEO

Cevahir Çığla[1], Xenophon Zabulis[2] and A. Aydın Alatan[1]

[1]Department of Electrical and Electronics Engineering, M.E.T.U, Turkey
[2]Informatics and Telematics Institute, Thessaloniki, Greece
e-mail: cevahir@eee.metu.edu.tr, xenophon@iti.gr, alatan@eee.metu.edu.tr

ABSTRACT

A novel multi-view region-based dense depth map estimation problem is presented, based on a modified plane-sweeping strategy. In this approach, the whole scene is assumed to be region-wise planar. These planar regions are defined by back-projections of the over-segmented homogenous color regions on the images and the plane parameters are determined by angle-sweeping at different depth levels. The position and rotation of the plane patches are estimated robustly by minimizing a segment-based cost function, which considers occlusions, as well. The quality of depth map estimates is measured via reconstruction quality of the conjugate views, after warping segments into these views by the resulting homographies. Finally, a greedy-search algorithm is applied to refine the reconstruction quality and update the plane equations with visibility constraint. Based on the simulation results, it is observed that the proposed algorithm handles large un-textured regions, depth discontinuities at object boundaries, slanted surfaces, as well as occlusions.

Index Terms— multi-view stereo, segmentation, plane sweeping, angle sweeping

1. INTRODUCTION

Dense depth-map estimation is one of the classic problems in computer vision with its wide application areas, such as 3D object modeling, robot navigation and image-based rendering. There are two main approaches in obtaining dense depth maps, stereo and multiple-view matching. In literature, there are many algorithms making use of stereo images [1]. Most of the algorithms focus on the fact that the estimated dense depth map of a scene should smoothly vary on the surfaces of the objects and change sharply at the object boundaries. Due to the limitations of the observed data, stereo matching algorithms might fail for wide-baseline image pairs, in which the scene contains occlusions frequent and repeated patterns.

In order to increase the precision of the depth map, while handling wide-baselines, repeated structures and occlusions, multiple images are usually utilized [2,3,4]. Multi-view approach to stereo matching increases the amount of observed data, hence should result in more robust depth maps. There are different approaches that combine multiple views, such as pixel-based PDE methods [3], graph cuts [5] and volumetric methods [6]. Taxonomy of multiple-view stereo reconstruction algorithms is given in [7], based on only the volumetric methods, which mostly require a large amount of images.

In last decade, segment-based stereo matching algorithms have been developed with an increasing performance in their resulting depth fields [8, 9, 10]. The main assumption for these algorithms is that the scene is composed of small non-overlapping planes, all of which correspond to distinct segments obtained via grouping pixels of homogenous color. Segment-based stereo matching algorithms are generally studied in the context of small baseline stereo and for almost frontoparallel planes. Hence, they are also expected to suffer from wide-baselines and occlusions, as most other stereo algorithms do. Considering the robustness of the segment-based stereo matching algorithms, which preserve depth discontinuities at object boundaries and support smoothness at similarly colored regions; the same idea can be upgraded for multi-view content. Multi-view extension will result in segment-based stereo matching to handle wide-baselines, repeated patterns and occlusions, while still preserving object boundaries.

This paper proposes a novel multi-view region-based dense matching algorithm, based on a modified plane-sweep method [11]. First, the best plane representation is determined for each segment. Next, 3D plane normals are estimated by angle-sweeping [12] for different planes and the normals are updated via a greedy search among neighboring segments with visibility constraint. The details of the proposed algorithm are explained in Section 2, while the simulation results are presented in Section 3. The last section is devoted to the concluding remarks and the future directions.

2. ALGORITM

The proposed algorithm tries to solve the dense depth map estimation problem by modeling the scene via non-

overlapping planar segments. In this aspect, the projections of these planar segments onto different views, as regions, could be treated similar to the feature points to be matched in different views. Hence, segment matching is performed in order to estimate 3D position of each segment, which involves determining the normal direction of this plane, as well as its distance from the reference camera. For region matching, the observed scene is divided into parallel planes to determine visual consistency between regions in different views. Next, this plane search is refined by varying the orientation of the particular segment to find the best angle position. During the estimation of parameters of the segment planes, color matching and smoothness among neighboring regions are utilized. The algorithm takes N calibrated images as input and a dense depth map is computed for the selected reference view. There are 4 steps in the proposed algorithm; segmentation of the reference image, plane sweeping of segments, angle sweeping and iterative update of the plane parameters. In the following subsections, each of these steps is explained in details.

2.1. Segmentation

All segment-based algorithms assume smooth depth variation within regions of homogenous color, while they allow sharp depth changes at the boundaries of the regions that have high intensity differences. In the proposed algorithm, the same assumption is also valid; in addition, the segments are modeled as 3D planes. In order to meet the assumption of piecewise planar modeling and extract candidate plane patches, the reference image is over-segmented. For segmentation of the images, *mean-shift image segmentation* algorithm [13] is utilized. A typical segmentation mask for the reference image from *Break-dancer* multi-view sequence [14] is presented in Fig.1.

2.2. Initial Depth Estimation via Plane-Sweep

In most of the segment-based stereo algorithms, initial depth estimation step from two views is performed via local matching in the pixel domain, which is followed by plane fitting within segments. In this method, a novel depth estimation approach is proposed by treating all segments, as if they are feature points and matching these segments between the views. Moreover, additional information from multi-view is utilized, which increases the robustness of the algorithm against noise and repeated regions.

In order to assign a plane equation to a segment, the space is assumed to be initially divided into parallel planes [11] that are located at different depths. The separation between the planes is assumed to be located perpendicular to the principal axis of the reference camera, as *z-axis* (see Fig. 2). Hence all the planes are initially selected to be parallel to the reference view. For a particular depth plane, the relation between the locations of the pixels on all multi-view images and the reference image can be defined by a set of homography, which is explicitly given in (1) and (2)

Figure 1 : *Breakdancers*: (left) Reference view, (right) random colored over-segmentation result of the reference view

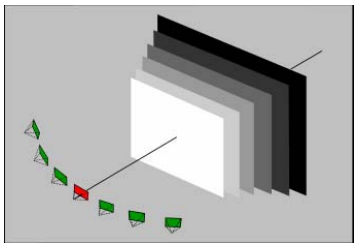


Figure 2 : 3-D space is divided into parallel planes, perpendicular to the principal axis of the reference (red) camera

$$H_i = [P_1^i \; P_2^i \; P_4^i] - [n_1/n_3 \; n_2/n_3 \; n_4/n_3].P_3^i \quad (1)$$

$$H_{kin} = H_k^{-1}.H_i \quad (2)$$

where P_j^i corresponds to the j^{th} column of the projection matrix of the i^{th} camera, n_i indicates the plane normal, H_i is the homography from the plane to the i^{th} camera and H_{kin} indicates the homography from k^{th} image to the i^{th} image.

At the plane-sweep step, each segment is assigned to a depth plane with constant depth by searching the depth space for minimizing a cost function. The cost value of a segment is defined for every depth plane, unifying all intensity differences of a segment between various views by the help of homographic mappings in (2). Since the effects of occlusions might be significant, especially for the segments having covered and uncovered points (see Fig. 3), the cost function should also take such cases into account and ignore the occluded pixels from the corresponding cost.

At this point, since the real structure of a segment could be different from parallel plane, the strategy for the calculation of the cost function is slightly updated, in such a way that the cost is calculated via only the best matching pixels for a segment. In other words, cost term is calculated by the summation of only the pixel intensity differences smaller than a threshold. Hence, the final cost value for the corresponding depth plane is determined by the summation of all cost values for each camera, normalized by the total number of contributing pixels. In this manner, the undesired high intensity differences due to covered or uncovered regions are eliminated.

After the evaluation of the cost function for all planes in the depth range, the planes, resulting with the best-n cost

values are further utilized for each segment individually in the refined search, during angle-sweep step.

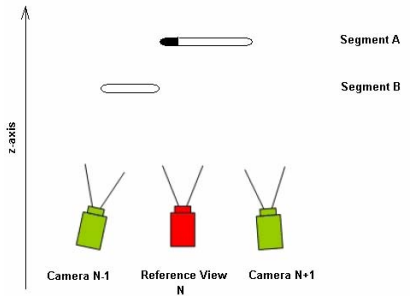


Figure 3 : The dark pixels of segment-A is occluded in $(N-1)^{th}$ camera, whereas non-occluded in $(N+1)^{th}$ camera

2.3. Refinement of Depth Estimation by Angle-Sweep

In the last step, the positions of the segments have been estimated coarsely, assuming all planar patches to have the same normal direction, parallel to the principal axis. This property causes segments to have constant depth with respect to the reference camera, which is valid only for very small segments. As the segment size increases, the constant depth assumption should fail, if there are slanted or curved surfaces in the scene. The curved surface cases are beyond the scope of this work and could only be modeled by concatenation of small planar segments. However, in order to handle the slanted surfaces, such large segments are rotated in *x*- and *y*-directions, around their centeroids (see Fig. 4). Rotation of a plane changes the initial homography, which can still be obtained via (1) and (2), after accordingly modifying the plane normal.

After the determination of the initial depth candidates in the plane-sweep step, the best rotation and depth combination (among *n* candidate planes from the last step) is searched for by sweeping the normal of the segments at different angles in the angle-space. Combination that gives the minimum cost is selected as the location of the planar segment in 3-D space. In this part, the same cost function in the plane-sweep step is utilized.

2.4. Iterative Plane Refinement

So far, the 3D locations of segments have been searched by only considering intensity matching cost without considering the relation between neighboring segments. In this step, the plane equations of the segments are also updated by considering the smoothness and visibility constraints in order to refine the depth map. Considering the general properties of a depth map, smoothness between similar colored regions is a realistic assumption, since depth field changes smoothly in a real environment, except for the object boundaries. On the other hand, the visibility of the segments should also increase the reliability of the depth map, since occlusions are implicitly considered in the scene model. In the plane-sweep step, occlusions are already modeled indirectly via the cost function by utilizing a threshold. After modifications in the scene model, the reliability of the depth map is measured in terms of the reconstruction quality of the other views from the reference image.

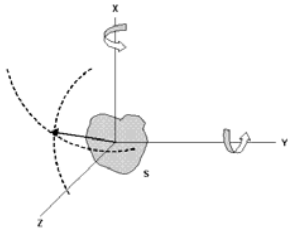


Figure 4: Segments are rotated around their centeroids in *X* and *Y* directions

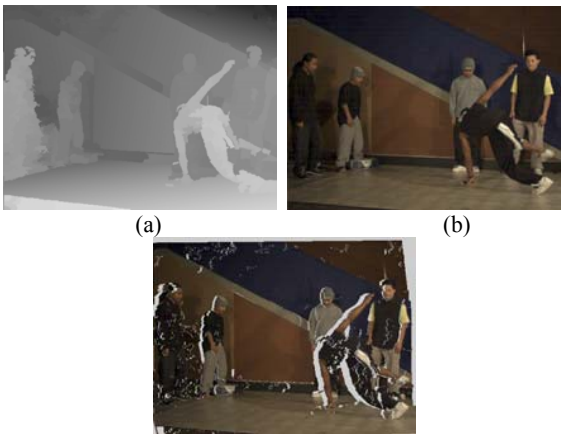
In order to fulfill the above requirements, a new cost function is defined, which consists of a smoothness term that tries to minimize the depth differences of the neighboring pixels on the segment boundaries. The other term in the new cost function takes the visibility constraint into account. Since the reference image is warped onto the other cameras by utilizing the plane equations and the resulting homographies, some pixels might have more than one correspondence from the pixels on the reference image during such a reconstruction. In such a case, the pixel, which is closer to the camera, should be rendered. For realizing the visibility, the reconstructed images are stored in different *Z-buffers* and the pixels on top of Z-buffers are utilized during the intensity filling [8]. The intensity similarity is measured by only the visible pixels, without thresholding, since occlusions can be modeled via visibility constraint. Hence, the new cost function can be written as:

$$C = \frac{1}{N}\Big(\sum_{k=0}^{imageNO}\sum_{x\in V_i}\left|I_{rk}(x) - I_k(x)\right|\Big) + \lambda \sum_{\substack{X\in B_i \\ X'\in N_i}}\left|D(x) - D(x')\right|$$

where D is the depth map, I_{rk} is the reconstructed image of the k^{th} camera, V_i is the set of visible pixels of i^{th} segment, N is the total number of pixels utilized, B_i is the boundary and N_i is the neighboring pixels of the segment.

The depth map for each segment is obtained via the iterative optimization of this new cost function. The problem is a labeling problem, and there are different types of solutions, such as graph-cut [9], belief propagation [15]. In the proposed algorithm, a method, which is similar to the greedy-search algorithm, given in [16] is applied. In this method, for each segment, a search is conducted in the depth-space, bounded only with the depth range of its neighboring segments. For considerably large segments, angle sweeping is also applied within the bounded region. The segment is warped to all other images, and the cost function is evaluated for each angle and depth combination.

The combination, which gives the best improvement in the cost function, is assigned to the segment; however, the models are updated after all of the segments are visited. This operation is performed iteratively, until there is sufficiently small number of updates among all segments.



(a) (b)

(c)

Figure 5: a) Estimated depth map, **b)** original neighboring view, **c)** the reconstructed view from the depth map via warping

3. RESULTS

We have tested the proposed algorithm on a number multi-view sequences, including Microsoft *Break-dancer* [14]. In Fig. 5, the estimated depth map of the reference image, as well as one of the reconstructed images (closest to the reference view), and the original image are illustrated. In the resultant depth map, the depth discontinuities at the object boundaries can be observed, clearly, and the depth field of the un-textured regions is also estimated with a quite satisfactory performance. In addition, the correct estimation of 3D slanted surface locations could be observed. Since there is no hole-filling in the reconstruction stage, the occlusions could be observed. In Fig. 6, the depth map given in [14] is also illustrated which is obtained by the algorithm given in [17]. As can be observed, the proposed algorithm results in a reliable depth map which is comparable, or better, against that of competing methods.

4. CONCLUSION AND FUTURE WORKS

A novel multi-view segment-based depth estimation algorithm is proposed via plane and angle sweeping in 3D space. Considering the high performance of segment-based stereo matching algorithms, the proposed algorithm formulates a multi-view extension which handles occlusions, repeated structures and wide-baselines, as well. In addition, arbitrary-shaped segment matching is performed, instead of pixel matching, which reverses the pixel-to-segment strategy.

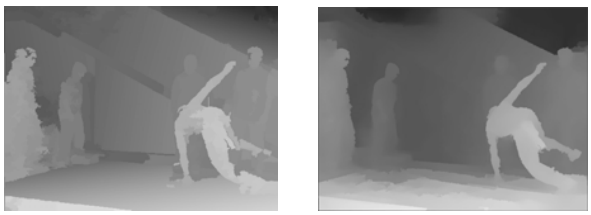


Figure 6: Depth map of (left) the proposed algorithm and (right) the method in [17]

The algorithm depends on the initial segmentation of the scene, therefore an erroneous segmentation of homogenous color regions might result in a inferior depth map. In order to overcome such a case, a region splitting algorithm could be applied, as proposed in [10], which is thought as a future work of the study. In addition, the algorithm will be extended for depth estimation from the whole multi-view sequence.

6. ACKNOWLEDGEMENT

This work is funded by EC IST 6th Framework 3DTV NoE and partially funded by TÜBİTAK under Career Project 104E022.

11. REFERENCES

[1] D. Sharstein and R. Szeliski, *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. In IJCV, Volume 47, pg 7-42, April 2002

[2] M.Okutomi and T.Kanade, *A Multiple-Baseline Stereo*, IEEE Transactions on Pattern Analysis and Machine Intelligence, April 1993

[3] S.Christopp *et al, Dense Matching of Multiple Wide-baseline Views,* in ICCV 2003

[4] S.B.Kang *et al, Handling Occlusions in Multi-view Dense Stereo,* in CVPR 2001

[5] V.Kolmogorov and R.Zabih, *Multi-camera Scene Reconstruction via Graph Cuts*. In ICCV, volume II, pages 508-515, 2001.

[6] K.Kutulakos and S.Seitz, *A Theory of Shape by Space Carving,* IJCV, 38(3):197-216, 2000

[7] S.M. Seitz et al, *A Comparison and Evaluation of Multi-view Stereo Reconstruction Algorithms*, In CVPR 2006.

[8] M. Bleyer and M. Gelautz. *A Layered stereo algorithm using image segmentation and global visibility constraints*, In ICIP 2004.

[9] L. Hong and G. Chen, *Segment-based stereo matching using graph cuts*, in CVPR, 1, pp. 74-81, 2004.

[10] Y.Wei and L.Quan, *Region-Based Progressive Stereo Matching,* in CVPR 2004

[11] Robert T.Collins, *A Space-Sweep Approach to True Multi-Image Matching*. In Proc. CVPR96, pages 358-363, 1996

[12] X. Zabulis ans K. Daniilidis, *Multi-camera reconstruction based on surface normal estimation and best viewpoint selection*. In Proceedings of the 2nd International Symposium on 3DPVT, 2004.

[13] D. Comaniciu and P. Meer. *Mean shift: A robust approach toward feature space analysis*. *IEEE:PAMI*, 24(5):603–619,May 2002.

[14] http://research.microsoft.com/IVM/3DVideoDownload/

[15] A. Klaus et all. *Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*, ICPR 2006.

[16] H. Tao and H. Sawhney, .*Global matching criterion and color segmentation based stereo*,. in *WACV*, 2000, pp. 246.253.

[17] C. Lawrence Zitnick et al, *High-quality view Interpolation using a Layered Representation,* in Proceedings of the 2004 SIGGRAPH Conference Volume 23 , Issue 3