

# Multi-view Dense Depth Map Estimation

Cevahir Çiğla

Department of Electrical and Electronics Engineering  
M.E.T.U, Turkey  
+90 312 2104508

cevahir@eee.metu.edu.tr

A. Aydın Alatan

Department of Electrical and Electronics Engineering  
M.E.T.U, Turkey  
+90 312 210 2351

alatan@eee.metu.edu.tr

## ABSTRACT

A novel dense depth map estimation algorithm is proposed in order to meet the requirements of *N-view plus N-depth* representation, which is one of the standardization efforts for the upcoming 3D display technologies. Hence, extraction of multiple depth maps is achieved from multi-view video. Starting from the piecewise planarity assumption of the scene, estimation of 3D structure of the patches, obtained through color-based over-segmentation, is achieved by plane- and angle-sweeping for every view independently. Markov Random Field (MRF) modeling is utilized for each view in pixel-wise manner in order to relax and refine the estimated planar models while incorporating visibility and consistency constraints. In this algorithm, the fusion of multiple depth maps is performed by updating belief values on the observed nodes based on depth and color consistency during the refinement step. The proposed method handles untextured surfaces, as well as depth discontinuities at object boundaries, due to its initial modeling of the scene as piecewise planar regions. The experimental results illustrate reliability and the robustness of the proposed algorithm for different type of scenes.

## Keywords

N-view plus N-depth, Color Segmentation, Belief Propagation, Dense Depth Map

## 1. INTRODUCTION

The recent advances in 3-D extraction [2] and representation [2] technologies, as well as the introduction of auto-stereoscopic (or multiscopic) 3D displays in the market, have pioneered new alternatives to traditional television setups which involve free-view TV and 3DTV. In free-view TV, the TV-viewers freely select their viewing position by generating virtual views of the scene, whereas in a typical 3DTV, the perception of the third dimension, depth, is provided to the viewer. For both of these approaches, any mono-view video data should be converted into a 3-D scene representation, so that 3D structure of the captured scene is available for these displays. Hence, the extraction of the 3D content from multiple images becomes crucial for the perception of the third dimension.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IMMERSCOM 2009, May 27-29, 2009, Berkeley, California, USA.  
Copyright C 2009 ICST ISBN #978-963-9799-39-4

Dense depth map representation [2] is one of the most popular and endeavored way of 3D representation of a scene. In this representation, a depth value is assigned for each pixel and a point-wise modeling is achieved where depth assignments could be utilized to map 3D points onto images to render arbitrary view points. Dense depth map estimation is an inverse problem of 3D extraction from 2D projections; thus, some assumptions, such as smoothness or color consistency, are required in order to obtain high quality models [3]. In that manner, many algorithms have been proposed in order to estimate depth maps that utilize stereo [3] or multiple views [4]-[7]. Most of these methods assign depth maps for only one image which is chosen as the reference view of the data set. However, video rendered from different viewpoints through only the reference view and its depth map, will be relatively poor in quality due to occlusions, especially when the view point moves away from the reference camera. Such occlusions result with disturbing visualization at the output of any 3DTV system. In order to provide more satisfactory and realistic visualization, contribution from multiple views should be considered. Such an approach requires depth maps for all of the cameras in multi-view setups, which is denoted as *N-view-plus-N-depth* representation [8]. This extension utilizes the maximum information belonging to multi-view video and is one of the standardization efforts to be utilized in the upcoming 3D display technologies.

The extraction of several depth maps from multiple images have not received much attention so far; however, as video capture technology developed, multi-view content become available and the requirement for multiple depth maps has increased. There are mainly two approaches for the solution of this problem; the first one is utilizing stereo algorithms for pairs of images among multiple views and then merging the estimated depth maps for consistency and obtaining a final representation [8]. This approach is sub-optimal, since the extra information, especially for the occluded regions, from multiple views is discarded and depth maps are initially obtained with the similar problems as most stereo algorithms face. However, utilizing multiple images for the estimation of depth map of each camera as proposed in [9] and [10] reduces the ambiguities and handles occlusions by constraining multiple match probabilities. In both [9] and [10], MRF models are constructed and depth assignment is achieved by loopy belief propagation (BP). In addition, the visibility, which is a powerful tool for depth estimation is also considered in these methods. The utilization of visibility constraint provides robustness against large occluded regions and consistent, reliable depth maps for all of the images. 3D extraction presented in [11], utilizes a similar representation as *N-view-plus-N-depth*, however the images are captured by a video camera and consecutive frames

are utilized as the multiple views. Thus, the problem is considered in the context of small base line and fusion of multiple depth maps via visibility is achieved for refining the initially estimated depth maps [11].

Although multiple depth map estimation algorithms provide satisfactory results, they still suffer from the untextured regions that might lead ambiguities. In this work, a novel segment-based approach is proposed followed by a refinement through belief propagation in pixel-domain in order to overcome these problems. The geometrical verification of untextured regions are difficult to achieve, since color variation at those regions is quite small; hence, the pixels may be mismatched no matter how many images are utilized during depth estimation. In the proposed method, however, those regions are treated as *super-pixels*, consisting of a group of pixels and matching is performed for all of the pixels belonging to the same segment. Such an approach enforces pixels to have depth values strictly (or parametrically) related with the neighboring pixels, namely a *planarity constraint* [12]. The relaxation of the depth map from the planarity assumption is achieved by applying belief propagation among pixels considering the depth consistencies, visibility and color match as well.

This paper is organized as follows, Section 2 gives an overview of the algorithm by explaining the fundamental assumptions; in Section 3, the details of the region based plane- and angle-sweeping dense depth map estimation methods are given, which is followed by the refinement algorithm based on belief propagation in Section 4. Section 5 is devoted to the experimental results and finally, the paper concludes in Section 6.

## 2. OVERVIEW OF THE ALGORITHM

The proposed algorithm estimates the depth maps for all images given in a multi-view sequence whose internal and external calibration parameters are provided at a certain time instant. The estimation is achieved in two steps as shown in Figure 1; in the first step, the scene is modeled via non-overlapping planar segments, which are obtained through over-segmenting each image by any color segmentation algorithm. The initial 3D models for these segments are estimated based on plane- and angle-sweeping, which are valuable tools for relating multiple images of the same scene with each other. During sweeping operations, the segments are swept among different depth candidates and rotated around their centroids in order to obtain the best possible match based on their intensity similarity, while being mapped to multiple images. The sweeping operations are performed for all images of the multi-view data independently and the initial depth maps are obtained for each image. In the second step, in order to improve the depth estimate accuracy, the planarity assumption is relaxed by modeling each image as MRF and utilizing pixel-based belief propagation that takes into account the visibility of scene points between different views. Pixels are considered as the nodes of the MRFs, whose states correspond to different depth candidates, and the final assignment is achieved by iteratively updating the messages coming from the neighboring nodes in loopy belief propagation algorithm. In the proposed MRF, each node has 4 neighbor nodes and the state probabilities are calculated according to the current depth maps and their correspondence reliability. After certain number of iterations, the states probabilities are updated according to the refined depth maps.

This approach provides the reliability and the consistency between the depth maps of different views. In this manner, the refinement and updates are performed iteratively until convergence of the models.

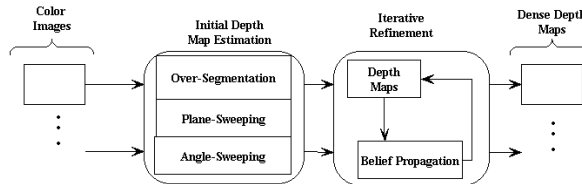


Figure 1: Flowchart of the proposed method

The region-based algorithms [14]-[17] [22], which have been popular in last decade, mostly focus on dense depth estimation from stereo pairs. The proposed method extends the region-based matching approach to multi-view data and deviates from the general direction of those methods by modifying the algorithm “from pixel matching to segment correspondence” into “from segment correspondence to pixel matching”. The novelty of the proposed method is mainly due to treating each segment as *super-pixels* to be matched and applying the iterative belief propagation refinement, which relaxes depth values within super-pixels and provides depth consistency. Assigning 3D models to each segment individually provides these pixels to have constrained depth values with their neighbors. Hence, the proposed algorithm provides reliable models, especially for the untextured regions, where matching could be problematic due to less texture variation. Proposed region-based approach assumes the scene to be piecewise planar; however, this assumption may not be valid as long as the scene involves roundly shaped, natural objects, which is a common case actually. Thus, the planarity models are relaxed via MRF-modeling in order to consider such objects as well.

## 3. INITIAL DEPTH MAP ESTIMATION

In this study, initial dense matching is provided by using a method, in which plane- and angle-sweeping is performed to estimate the planar model of the scene. In addition, special attention is devoted to overcome the difficulties resulting from large untextured regions and the effects of enlarging or shrinking of some regions due to rotation of cameras in a multiple-camera setup. The first main step of the proposed method is plane- and angle-sweeping of the segments in 3D scene during intensity matching process.

### 3.1 Color Segmentation

In order to exploit the piecewise planarity assumption for the scene, segmentation of some planar patches within this scene is required. According the *coherence principle*, it is assumed that the regions having similar color distribution should belong to the same object. This principle can be extended by stating that these regions should also have associated depth values, since they belong to the same object. In that aspect, the input views are over-segmented into a relatively higher number of regions, which do not have any semantic meaning, whereas correctly defining homogeneity in terms of color distribution. For the segmentation of the input images, among various tested methods, *mean-shift image segmentation algorithm* [18] is preferred, since it is insensitive to gradient changes which might lead to false contours. The performance of the initial depth map estimation algorithm

depends on the segmentation quality; an erroneous segmentation (i.e. a region belonging to more than one semantic objects or planes) might result with inferior depth maps. Fortunately, these errors could still be handled in the pixel-based refinement step during the relaxation of these depth maps.

### 3.2 Plane- and Angle- Sweeping

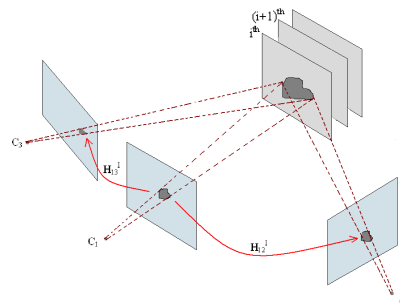
The plane-sweeping approach has been introduced in [20], in order to match the object boundaries between multiple images. The main idea stems from the requirement of relating multiple edge images to each other and extracting depth information. In this manner, a surface in any view can be easily back-projected to 3D space and by using the projection matrices, the corresponding locations of the surface in the other views are determined easily. The relations between images are defined via homographies, which also determine the projective transformations between planes. In order to map 3D surfaces to multiple images, the space is initially sampled with planes, which are parallel to each other along the principal direction of the reference view. The mapping between the cameras is achieved by homographic relations from 3D planes to camera planes as illustrated in Figure 2 for an arbitrary camera.

Sweeping through the depth planes provides the best depth values for each segment in the context of visual similarity [20]. The visual similarity is measured by summing the intensity differences between the pixels of the warped segment and the pixels onto which the segment is mapped. In the relation below, the formulation for a segment  $S_i$  in  $j^{\text{th}}$  image is given explicitly, where  $x'$  denotes the target pixels on which  $x$  in  $S_i$  is mapped according to the homographies between camera  $j$  and  $k$  at corresponding to depth  $d$ ,  $M$  denotes total number of different camera views,  $N_{total}$  corresponds to the total number of pixels mapped for  $S_i$ .

$$C_{S_i}^k(d) = \frac{1}{N_{total}} \sum_{\substack{k=1 \\ k \neq j}}^M \sum_{x \in S_i} |I_j(x) - I_k(x')| \quad (1)$$

In (4), it is observed that all of the pixels within a segment have contribution to the cost function; however, this may not be the case for partially occluded sub-regions within segments. In order to handle these cases, the visibility of the pixels should be considered which cannot be determined at the initial steps of the algorithm.

However, in this method, before the estimation of the depth maps, such cases are handled by thresholding the intensity similarities of the pixels. It is assumed that if the intensity difference for a pixel is above a threshold, then it should belong to an occluded region and the contribution of that pixel to the cost function should be ignored. After the initial depth estimates are obtained, the visibility is modeled precisely, which will be explained in the upcoming sections. Plane-sweeping provides the segments to have a constant depth value (i.e. distance from the camera) for all of their pixels within its boundaries. This assumption is valid as long as the size of segments is considerably small. As the segment size increases, the constant depth assumption may be violated, especially for the slanted and complex surfaces in the scene.



**Figure 2:** Each segment is mapped onto multiple images via homographies.

In the proposed method, such surfaces are modeled by angle-sweeping [17][21], based on the planarity assumption of the scene. In angle-sweeping, the segments are rotated around their centroids and the best similarity match is searched.

### 3.3 Iterative Update of Planar Structures

Up to this point, for the proposed algorithm, smoothness assumption is only enforced for the pixels belonging to the same segment via the planarity assumption, whereas smoothness between neighboring segments that belong to same object is not considered. In addition, the partially occluded segments are pre-modeled by intensity thresholding without any visibility constraint. In this step, the initial depth estimates are refined via smoothness constraint between neighboring segments and the visibility during the warping of the pixels to the other views. The previous cost function in (1), which is based on intensity similarity, is improved by a smoothness term that takes depth differences of the pixels on the boundaries of the segments and the visibility of the pixels into account, as follows:

$$C_{S_i}^k(d) = C_{S_i}^k(d) + \lambda \sum_{x \in S_j} |DI(X) - DI(X')| \quad (2)$$

In (5),  $X'$  is related to  $X$  by a homography depending on the angles at which the segments are rotated around their centroids and  $d$ , the corresponding depth location.  $DI$  denotes the estimated depth image;  $V_j$  is the set of visible pixel in  $I_k^j$  which is the reconstructed image of the  $k^{\text{th}}$  camera from the  $j^{\text{th}}$  camera. The reconstruction is achieved by warping all the segments in the corresponding image to the target image via the homographies between these two images. During such a reconstruction, some pixels in  $I_k^j$  might have more than one correspondence from the pixels on the source image. In this case, the pixel, closer to the camera, is rendered, which is actually the *visibility constraint*. The reconstructed image is stored in a *Z-buffer* [15] and the pixels on the top of this *Z-buffer* are utilized for the intensity filling in order to implement this constraint.

The minimization of the cost function in (2) is an *np-complete* problem and there are different solution approaches, such as *graph cuts* [14] and *belief propagation* [16]. In the proposed method, a greedy search algorithm introduced in [15] is utilized. In this method, for each segment a search is performed in the depth space, considering the neighboring smoothness and the reconstruction quality of the warped images. Hence, the segments are enforced to have similar depth values with the neighboring

segments that have similar color distribution which provides smoother depth maps.

#### 4. REFINEMENT VIA BELIEF PROPAGATION

Until this step, estimation of the depth maps has been performed independently for all the images in a multi-view camera setup. In addition, the scene is assumed to consist of non-overlapping planar structures in order to model untextured regions and object boundaries precisely. *N-view-plus-N-depth* representation requires the estimated depth maps to be consistent, that is, the points that are visible in multiple images should have the same depth values assigned in the corresponding depth maps. Thus, a refinement step is required to relax the planarity assumption and relate depth maps of different images with each other. In this study, a pixel-based belief propagation method is proposed for providing relaxation and consistency. Each view is modeled as a MRF whose nodes correspond to the pixels and the states of the nodes are the candidate depth values of the sampled space. In this model, each pixel is labeled ( $f_x$ ) to a depth value such that a predefined energy function is minimized. It is assumed that the neighboring pixels' depth assignments are smooth almost everywhere, unless the nodes are at the object boundaries or at sharp discontinuities.

For stereo problems, the energy functions in MRF models generally consist of two terms [23]; a *compatibility* term based on the observation cost  $\Phi(f_x)$ , and a *smoothness* term between the states of the neighboring hidden nodes  $\Psi(f_p)$ . A similar formulation is also utilized in the proposed work with the formula given as below:

$$E(f) = -\sum_{x \in P} \Phi(f_x) - \sum_{x \in P} \Psi(f_x) \quad (3)$$

The compatibility term originating from the observations, is a function that is based on the color similarity and the depth consistency of the labeling. The color similarity,  $E_{color}$ , is calculated as below by the absolute difference of the intensities at the corresponding pixel locations for visible depth candidates.

$$E_{color}(f_x) = \frac{1}{M-1} \sum_{\substack{i=0 \\ i \neq k}}^M |I^k(x) - I^i(x_f^i)| \quad (4)$$

In (4),  $x_f^i$  is the pixel in  $i^{th}$  camera onto which the pixel in the  $k^{th}$  image is mapped via labeling ( $f_x$ ),  $I$  is the intensity image of the  $i^{th}$  camera. The visibility can be easily modeled for this case without intensity thresholding by help of initial depth map estimates. The intensity match function is calculated for the depth labels which provide the current pixel to be visible. The visibility constraint is illustrated in Figure 3, for the corresponding depth hypothesis of point A in center camera ( $C_3$ ); the intensity function utilizes observations from 4<sup>th</sup> and 5<sup>th</sup> cameras, where the point is visible according to the estimated depth surface. Other cameras are discarded, since for that depth value point A is occluded in 1<sup>st</sup> and 2<sup>nd</sup> cameras.

The depth consistency,  $E_{depth}$ , for each pixel is measured by the absolute differences of the candidate depth value ( $f_x$ ) and the depth values at which the corresponding pixel is mapped onto, as follows:

$$E_{depth}(f_x) = \frac{1}{M-1} \sum_{\substack{i=0 \\ i \neq center}}^M |f_x - D^i(x_f^i)| \quad (5)$$

where  $D^i$  corresponds to the current depth map estimate of the  $i^{th}$  camera. Depth consistency provides the depth estimates to be similar for the pixels in different cameras that represent the same surface in the scene. Depth consistency and intensity similarity terms are combined with weighting factors to obtain the compatibility term, as follows:

$$\Phi(f_x) = \lambda_c \cdot E_{color}(f_x) + \lambda_d \cdot E_{depth}(f_x) \quad (6)$$

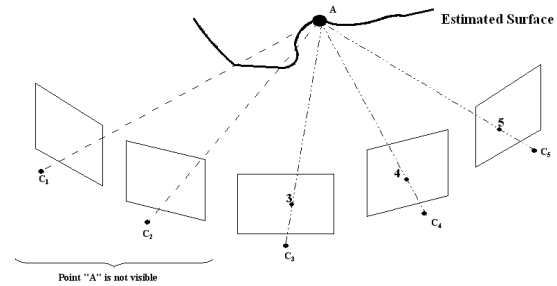
In (6),  $\lambda_c$  and  $\lambda_d$  correspond to the weighting factors of the color and depth information. In the proposed algorithm, these factors change in time as the depth maps are refined. The updating strategy can be summarized as follows: initially, the color information is more reliable compared to the depth estimates; hence, the weighting factor for color term is higher. However, as the depth maps are refined, the reliability of depth information increases; hence, the weighting factor for the depth term is also increased. The smoothness term, which controls the neighboring state relations of the hidden nodes, is a linear function of the state differences between neighboring nodes, where  $z$  denotes the neighboring nodes of  $x$ :

$$\Psi(f_x) = \sum_{z \in N(x)} |f_x - f_z| \quad (7)$$

The depth estimates are refined by the maximization of the total energy function. In this study, min-sum belief propagation algorithm [23] is utilized in order to propagate messages to the neighboring nodes and update the states of the nodes. Belief propagation is an iterative method that passes messages in a local manner, and widely utilized in computer vision problems due to its convergence property.

Each node sends messages to their neighboring nodes indicating the energy values at different state cases, iteratively. In the equation below, the belief update relation is given, where  $m_{z \rightarrow x}^T(f_x)$  denotes the message passed to node- $x$  for the labeling of  $f_x$  from node- $z$  at iteration  $T$ . Finally, the assignment of the states is achieved by the minimization of the beliefs among  $f_x$  at each node individually.

$$b_x(f_x) = \Phi(f_x) + \sum_{z \in N(x)} m_{z \rightarrow x}^T(f_x) \quad (8)$$



**Figure 3:** Color similarity and depth consistency are measured for visible states of the pixels.

Belief update is considered for MRF of each view independently; however, MRFs of different views should be related to each other in order to obtain consistent depth maps. This relation is provided with a similar method as presented in [9]. The procedure can be summarized as follows:

1. Calculate the energy function (3) for each node at each state by the current depth map estimates.
2. Perform iterative belief propagation for each MRF model (each view) independently in order to refine the depth maps.
3. Go to step 1 with the refined depth maps.
4. Iterate the first three steps until the convergence of the refinement of the depth maps.

Since the updates are performed in pixel domain for belief propagation, the relaxation of the planar structures which are estimated segment-wise is achieved. In addition, especially for the untextured regions, there is a possibility to trap in local minima which correspond to false depth estimations. However, the proposed method first constrains the pixels in these regions and enforces them to be strictly planar; later, relaxes them without violating the smoothness assumption.

## 5. EXPERIMENTAL RESULTS

The proposed algorithm has been tested on various multi-view data sets, such as *Break-Dancing* [19], *Ballet* [19] and *Uli* [24], which are widely utilized in multi-view coding, rendering and depth estimation research. Sample frames of those sequences are illustrated in Figure 4. During these experiments, for each data set, 5 different views are utilized and the corresponding 5 depth maps are estimated. The initial images are independently segmented into approximately 500-1000 sub-regions by mean-shift image segmentation algorithm [18].

The estimated initial depth maps for the reference camera of different data sets and the refined depth maps as a result of applying pixel-based belief propagation are presented in Figure 5, Figure 6 and Figure 7, respectively, where the darker intensities indicate the regions closer to the cameras. It can be easily observed that the erroneous regions in the initial estimates are corrected after the refinement step; in addition, the consistency of the depth maps is also achieved. When the refined depth estimates are analyzed, it is observed that the proposed method also handles large occlusions and untextured regions quite well. In *Break-Dancing* and *Ballet* frames, the depth distributions at floor and wall do not involve spikes generally originating from texture similarities; hence, smooth depth variation is obtained. Thus, the advantage of the segment-based approach is obviously utilized to obtain robust and realistic depth maps.

The quality of the resulting depth maps is also measured objectively by calculating the PSNR values of the reconstructed frames that are obtained from the depth maps and different viewpoints. It should also be indicated that higher PSNR values can be obtained via better frame rendering algorithms, whereas in this work, the algorithm presented in [25] is utilized during the rendering stage. PSNR values are calculated by discarding the occluded parts of the reconstructed images; thus, the comparisons based on PSNR might not yield the results with subjective visual quality. The proposed method is also compared to the work

presented in [11] (*median-fusion*) for *Break-Dancing* sequence. The comparison is performed by PSNR calculation of the consecutive reconstructions,  $n^{\text{th}}$  camera from  $(n+1)^{\text{th}}$  camera, in order to minimize the ratio of the occluded regions. PSNR values are given in Table 1. According to these results, the proposed algorithm outperforms [11] and provides the best reconstruction for different camera positions.

In Figure 8 and Figure 9, reconstruction results of the center camera views of *Break-Dancer* and *Ballet* sequences from the left and right neighboring cameras are illustrated. In that scenario the occluded regions are handled, thus PSNR values are calculated among the whole image. The quality of the reconstruction for *Break-Dancing* is 32.26 dB while it is 31.98 dB for the *Ballet* sequence. In Figure 9, the reconstructions from initial and refined depth maps are also compared, and the effect of refinement is clearly observed in terms of visual quality and PSNR measure.

## 6. SUMMARY AND CONCLUSIONS

A novel algorithm is proposed for the estimation of multiple depth maps from multi-view utilizing cues from each camera in order to provide consistency and reliability. The approach of “from segment correspondence to pixel matching” provides the untextured regions and object boundaries to be modeled correctly, while the smoothness of estimated depth maps is also satisfied. Moreover, associating depth maps with each other refines the estimated models and it is clearly observed that consistency between neighboring depth maps is crucial for extracting 3D information precisely. The experimental results show that the proposed approach is applicable to different kind of scenes, indoor or outdoor, and it is robust to texture and illumination variations. The future work is devoted to the utilization of temporal consistency in addition to spatial consistency among the estimated depth maps of different cameras in multi-view video that will provide smoother frame transition in rendered video sequences.

**Table 1:** PSNR values for the reconstructions via three different depth map results

PSNR	Initial [17]	Proposed	Median-Fusion [11]
<b>view-3 from view-2</b>	30.68	<b>31.41</b>	30.98
<b>view-4 from view-3</b>	31.86	<b>32.53</b>	31.68
<b>view-5 from view-4</b>	31.20	<b>32.13</b>	31.77
<b>view-6 from view-5</b>	31.13	<b>31.59</b>	31.43



**Figure 4:** Typical frames from (a) *Ballet*, (b) *Break-Dancing* and (c) *Uli* multi-view data sets.



Figure 5: (a) Initial, (b) refined depth map for *Break-Dancer*.



Figure 6: (a) Initial depth estimate, (b) refined depth map for *Ballet*.

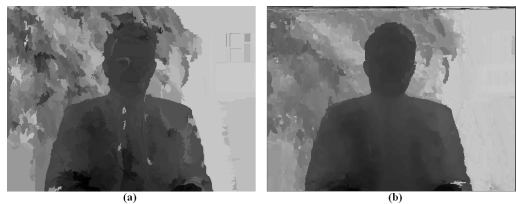


Figure 7: (a) Initial depth estimate, (b) refined depth map for *Uli*.



Figure 8: (a) Original 4<sup>th</sup> camera view, (b) Reconstruction from two neighboring cameras (32.26 dB).

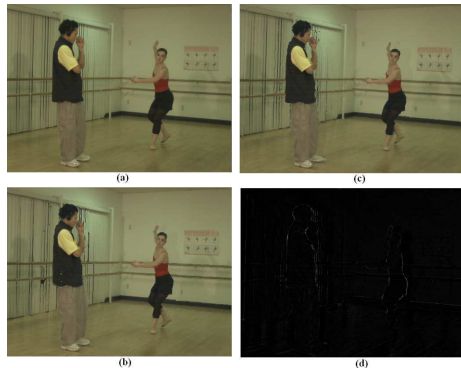


Figure 9: (a) Original 4<sup>th</sup> camera view, (b) reconstruction via initial depth estimate (30.78 dB), (c) reconstruction with refined depth estimate (31.98 dB), (d) difference map between the original frame and best reconstruction.

## 7. ACKNOWLEDGEMENT

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

## 8. REFERENCES

- [1] E. Stoykova, A. A. Alatan *et al.*, *3-D Time-Varying Scene Capture Technologies - A Survey*. IEEE Trans. Circuits Syst. Video Techn. 17(11): 1568-1586 (2007)
- [2] A. A. Alatan, Y. Yemez, *et al.*, *Scene Representation Technologies for 3DTV - A Survey*. IEEE Trans. Circuits Syst. Video Techn. 17(11): 1587-1605 (2007)
- [3] D. Sharstein and R. Szeliski, *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. International Journal on Computer Vision, Volume 47, pg 7-42, April 2002
- [4] M.Okutomi and T.Kanade, *A Multiple-Baseline Stereo*, IEEE TPAMI, April 1993
- [5] S.Christopp *et al.*, *Dense Matching of Multiple Wide-baseline Views*, International Conference on Computer Vision (ICCV) 2003
- [6] S.B.Kang *et al.*, *Handling Occlusions in Multi-view Dense Stereo*, IEEE CVPR 2001
- [7] V.Kolmogorov and R.Zabih, *Multi-camera Scene Reconstruction via Graph Cuts*. ICCV 2001.
- [8] P. Kauff *et al.*, *Depth Map Creation and Image Based Rendering for Advance 3DTV Services Providing Interoperability and scalability*, Image Communication Vol. 22, Issue 2 pages 217-234.
- [9] C. Lawrence Zitnick *et al.*, *High-quality view Interpolation using a Layered Representation*, Proceedings of the 2004 SIGGRAPH Conference Volume 23, Issue 3
- [10] E. Scott Larsen, P.Mordohai, H.Fuchs and M.Pollefeys, *Simplified Belief Propagation for Multiple View Reconstruction*, 3DPVT 06, pages: 342-349, 2006
- [11] Paul Merrell *et al.*, *Real-Time visibility-Based Fusion of Depth Maps*, International Conference on Computer Vision (ICCV) 2007.
- [12] Takahashi H. and Tomita F., *Planarity Constraint in Stereo Matching*, ICPR1988, vol 1, pages: 446-449.
- [13] Robert T.Collins, *A Space-Sweep Approach to True Multi-Image Matching*, IEEE CVPR, pages 358-363, 1996.
- [14] M.Bleyer and M. Gelautz, *Graph-based surface reconstruction from stereo pairs using image segmentation*, Proc. SPIE vol.5665, January 2005.
- [15] H. Tao and H.D.Sawhney, *Global matching criterion and color segmentation based stereo*, applications of Computer Vision, 2000, fifth IEEE Workshop pages:246-253.
- [16] A. Klaus *et al.*, *Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*, International Conference on Pattern Recognition 2006.
- [17] C. Cigla, X.Zabulis and A.A. Alatan, *Region-Based Dense Depth Extraction from Multi-view Video*, ICIIP 2007.
- [18] D. Comaniciu and P. Meer. *Mean shift: A robust approach toward feature space analysis*. IEEE PAMI, 24(5):603-619, May 2002.
- [19] <http://research.microsoft.com/IVM/3DVideoDownload/>
- [20] Robert T.Collins, *A space-sweep approach to true multi-image matching*, Proc. CVPR96, pages 358-363, 1996.
- [21] X.Zabulis and K. Daniilidis, *Multi-camera reconstruction based on surface normal estimation and best viewpoint selection*, In Proceedings of 3DPVT, 2004.
- [22] Sang Yoon Park, Sang Hwa Li, and Nam Ik Cho, *Segmentation based disparity estimation using color and depth information*, International Conference on Computer Vision 2004.
- [23] Pedro F. F. and Daniel P. H., *Efficient Belief Propagation for Early Vision*, International Journal of Computer Vision Vol. 70, No. 1, October 2006.
- [24] [https://www.3dtv-research.org/3dav\\_CfP\\_FhG\\_HHI/](https://www.3dtv-research.org/3dav_CfP_FhG_HHI/)
- [25] William R. Mark, L. McMillan, G. Bishop, *Post-rendering 3D Warping*, Proceedings of 1997 Symposium on Interactive 3D Graphics