Multidim Syst Sign Process
DOI 10.1007/s11045-012-0191-6

Reconstruction of coarse depth estimates using hierarchical image partitions

Camilo Dorea · Ricardo L. de Queiroz

Received: 24 April 2011 / Revised: 13 April 2012 / Accepted: 30 April 2012

Abstract This paper presents a novel depth map enhancement method which takes as inputs a single view and an associated coarse depth estimate. The depth estimate is reconstructed according to the region boundaries proposed within a set of image partitions of the associated view. Image partitions are determined through a color-based region merging procedure and organized within a hierarchy. Novel depth estimates are computed according to plane-fitting procedures using the proposed regions as support and selected coarse depth estimates. The use of partition hierarchies provides an efficient mechanism for evaluating plane-fitting results over various scales and, in this way, reduce the dependency of the results on the quality of a single color-based segmentation. Experimental results on publicly available test sequences illustrate the potential of the algorithm in significantly improving low quality depth estimates.

Keywords Depth map reconstruction · Hierarchical image partitions · Image segmentation

1 Introduction

Renewal of interest in 3D video services has also been accompanied by rising demand for new technologies surrounding acquisition, processing and display of the images necessary for depth perception. In order to simultaneously provide multiple images at varying viewing angles of the same scene and faced with the large increase in raw data size imposed by such requirements, several video representation formats have adopted the use of depth maps (Merkle et al. 2007). A depth map, associated to a particular view of the scene, is used to represent the distance of each point in the scene to a reference camera. The per pixel depth map forms a gray-scale image in which objects that are closer to the camera appear brighter

C. Dorea (✉) · R. L. de Queiroz
Department of Computer Science, University of Brasilia, Brasilia, DF 70919-970, Brazil
e-mail: camilo@cic.unb.br

R. L. de Queiroz
e-mail: querioz@ieee.org

Published online: 12 May 2012

while farther ones are shown darker. Depth map structure is thus generally formed by smooth areas delimited by sharp depth discontinuities.

Processing techniques using depth-based representations rely on rendering algorithms (Kauff et al. 2007) to synthesize other virtual views using a given view of the scene, its corresponding depth map and known camera calibration parameters. Thus, a single depth map can allow the synthesis of virtual views within a range of arbitrary angles while occupying relatively little storage space. In spite of these advantages, widespread usage of depth maps still faces a few obstacles. Depth map accuracy is fundamental to the quality of view rendering, however, its precise computation, either via disparity estimation or direct measuring equipment (e.g., light ranging systems), remains a challenging and costly task. Inaccurate depth map estimation, particularly around the discontinuities along the contours of objects at different depths, can lead to annoying artifacts such as incomplete or disconnected objects in rendered views.

In multiview systems, the impact of depth map inaccuracies and their consequent rendering artifacts may be alleviated by combining complementary information from other available views and depth maps. For example, a blending step in Muller et al. (2008) identifies depth discontinuity strips of fixed width and avoids the usage of such points in view interpolation, preferring the use of points from complementary views which are beyond their discontinuity areas. In Zhao et al. (2010), misaligned depth discontinuities are precisely located and also suppressed from the rendering step. Success of these approaches is contingent on the existence of complementary information from other views. Numerous 3D systems, however, are based on single-view formats such as video-plus-depth. Off-the-shelf systems composed of a single color camera and a range sensor (Kolb et al. 2009) are becoming increasingly common and, often, the result of depth map estimates from low-cost range sensing devices is coarse and/or of low-resolution. In such circumstances, depth map enhancement is the only means of guaranteeing view rendering quality.

Improvement of initial depth map estimates has been the subject of several studies aimed at 3D modeling and 3D television. In computer vision applications, segmentation-based approaches have been used to complete missing depth information (Torres-Mendez and Dudek 2004; Wang et al. 2008). Video images and a limited amount of range data are combined in an MRF framework to estimate unknown depth values (Torres-Mendez and Dudek 2004). In (Wang et al. 2008) an inpainting algorithm allows the user to remove an object from the scene and then fill in missing texture and depth areas with background information from stereo images. The previous approaches rely on multiple views to estimate depth within a user defined area. Segmentation-based approaches have also been used as a means of estimating and/or refining disparity in stereo algorithms (Scharstein and Szeliski 2002). For example, in Zitnick et al. (2004) the authors propose a color segmentation-based stereo algorithm to generate and refine disparity among pairs of multiple cameras for view interpolation. Both disparity estimation and refinement steps are, however, based on the usage of complementary views. Furthermore, refinement stages are generally coupled to the estimation results. Coarse estimates, produced by separate, low-cost range sensing devices, for example, can present significant errors beyond the capacity of dedicated disparity refinement.

A second family of approaches is based on the design of non-linear, edge-preserving filters which incorporate single image information to reinforce depth map structure. The bilateral filter uses color information as weights to precisely define discontinuities in low-resolution depth maps (Yang et al. 2007). A similar reconstruction filter is adopted in Oh et al. (2009) to compensate for depth coding errors and recover object boundaries. These filtering approaches, however, depend on correct registration between the color image and depth map.

Lastly, the technique in Ekmekcioglu et al. (2009) uses an adaptive median filter as well as image and depth information from multiple views to enhance coded depth maps.

This paper presents a novel, segmentation-based method for the reconstruction of coarsely estimated depth maps. The proposal relies on a single color image and its associated depth estimate. Estimates are reconstructed in accordance to the boundaries proposed by the regions within a hierarchy of color-based image partitions. The partitions serve as regions of support in which novel depth estimates are computed by fitting depth planes to selected coarse depth values. This proposal improves upon the color-based depth reconstruction scheme initially presented in Dorea et al. (2011) by extending reconstruction beyond the relevant depth discontinuity areas to the entire depth map. Furthermore, the plane-fitting procedures permit the assignment of non-uniform depth values within their regions of support. The usage of partition hierarchies, formed through efficient region merging procedures, grants the method greater robustness to segmentation errors since depth reconstruction results may be evaluated at various scales within the partition hierarchy. Instead of relying on adjacent images for complementary information, the hierarchy can efficiently organize multiple partitions of the same color image in order to improve significantly distorted depth estimates. The proposal may be applied, for instance, towards the enhancement of low-resolution or low-quality depth estimates arising, for example, when a separate range sensor is used in conjunction with a single color camera. Enhancement is validated through visual inspection of reconstructed depth maps as well as view synthesis results using publicly available data sets.

The remainder of the paper is organized as follows. In Sect. 2 we introduce the mechanisms for building the color-based image partition hierarchies. Using the hierarchy, the reconstruction procedure for coarse depth estimates is discussed in Sect. 3. Experimental results, presented in Sect. 4, illustrate the potential of the algorithm in reconstructing severely distorted depth maps. Section 5 concludes the paper.

2 Hierarchical image partitions

Region-based representations generally involve a partition of the image that groups together similar pixels. Such representations have several advantages. First, they greatly reduce the number of primitives, i.e., the number of regions is much lower than the number of pixels. Second, by grouping similar pixels, the consistency of models or measurements within each region may be improved while retaining accuracy in the representations, i.e., the contours remain well defined. Finally, regions or groups of regions can correctly represent the objects within a scene.

Object definition may be further improved with hierarchical or multi-scale representations. Instead of offering a single partition of the image, these representations offer a family of partitions at various scales or levels. The different levels allow greater flexibility in precisely defining objects. In this work, the regions within the partition hierarchy will serve as regions of support for depth reconstruction. It is assumed the discontinuities along the contours of objects at different depths are a subset of the region borders present within the hierarchy. Regions will be defined through color homogeneity metrics also including area and shape.

Construction of the hierarchical image partition layers is achieved with the region merging algorithm of Garrido and Salembier (1998). Consider the hierarchy of partitions pictured in Fig. 1. In the hierarchy, each region of a partition P_k is a subset of a region of P_l for $k < l$. Partition P_k is said to be finer than P_l and, conversely, partition P_l is coarser than P_k. Each region of P_l can be obtained through a union or merging of regions of P_k. Associated to

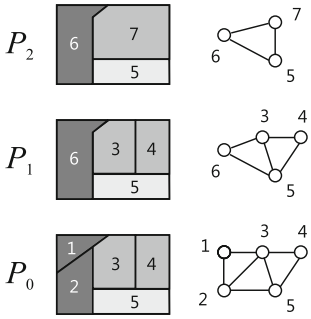

Fig. 1 Example of a partition hierarchy and the associated RAGs. Beginning with the finest partition (*bottom row*) region merging is used in constructing the hierarchy

each partition, a region adjacency graph (RAG) is also presented. The RAG is a set of nodes representing the regions. The links within the RAG connect neighboring regions. Throughout this work, neighboring or adjacent regions are defined under 4-connectivity. The merging algorithm of Garrido and Salembier (1998) proceeds iteratively by: (1) removing a link from the RAG, (2) merging the associated nodes in order to create a new region and (3) updating the links to the new neighborhood configuration.

The merging algorithm is further specified through a merging order which defines the order in which the links are processed. The order is computed as a similarity measure expressing homogeneity between adjacent regions. In this work we have limited our analysis to image features including region color, area and shape complexity in composing a similarity measure. Models based on such features can be efficiently computed and updated for each region as the merging algorithm progresses, greatly reducing processing time. Once a merging order has been specified, beginning from an initial partition, the algorithm may proceed iteratively as described. The results of each merging step define regions of similar features which may be organized within a partition hierarchy for further analysis.

Color is described in the YUV color space. The color model $\mathbf{C}_R$ for a region R is a column vector containing the mean YUV color components over all pixels $p \in R$. A color similarity measure (Garrido and Salembier 1998) between regions R_i and R_j which also takes into account region area $|\cdot|$ is given by:

$$\begin{aligned} S_a(R_i, R_j) = |R_i| \left\| \mathbf{w}^{1/2} \left(\mathbf{C}_{R_i} - \mathbf{C}_{R_i \cup R_j} \right) \right\|^2 \\ + |R_j| \left\| \mathbf{w}^{1/2} \left(\mathbf{C}_{R_j} - \mathbf{C}_{R_i \cup R_j} \right) \right\|^2 \end{aligned} \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm and $\mathbf{w}^{1/2} = [\sqrt{w_Y}\ \sqrt{w_U}\ \sqrt{w_V}]$ is a vector of weights such that $w_Y + w_U + w_V = 1$. The usage of area-based weights avoids the formation of partitions containing a few large regions surrounded by a large number of tiny regions. Employed segmentation parameter values are summarized within Sect. 4.

Shape complexity within the resulting partitions can be reduced by introducing an additional term into the color similarity metric which expresses the increment in perimeter relative to the area when annexing a region. Assume, without loss of generality, that R_i is the region of smallest perimeter, $C(R_i, R_j)$ is the ratio of the increment in perimeter and the increment in area of $R_i \cup R_j$ relative to R_i and $cp(R_i, R_j)$ is the common perimeter between R_i and R_j. The final color-based similarity measure prioritizes the formation of regions of low complexity and is given by

$$S(R_i, R_j) = \alpha S_a(R_i, R_j) + (1 - \alpha) \frac{C(R_i, R_j)}{cp(R_i, R_j)}. \tag{2}$$

(a) (b) (c)

Fig. 2 **a** The *Breakdancers* sequence (view 2, frame 0) and hierarchical partitions: **b** fine partition (2,000 regions) and **c** coarse partition (500 regions)

In this work, the merging algorithm is initiated upon individual pixels where each pixel is assumed as a region in an initial partition as exemplified in Fig. 2. The color-based similarity measure of (2) is used to order the iterative merging algorithm defined over the RAG. Adjacent regions with highest similarity are merged successively until a termination criterion is reached. Such a criterion may be the total number of regions or a maximum color modeling error beyond which merging of regions will likely violate relevant color boundaries (Dorea et al. 2009). Further studies on precise limits for meaningful region merging can be found in Huihai et al. (2007).

In our application, depth discontinuities among objects within the scene are assumed to be mostly contained within a color-based image partition of typically 500 regions. A finer partition of typically 2,000 regions, formed at one of the previous merging iterations, is assumed to contain any depth discontinuities not included among the borders of regions within the coarser partition. Examples of such partitions are depicted in Fig. 2b, c, respectively. Such partitions form the hierarchy of coarse and fine partitions employed by the proposed depth map reconstruction method. The use of a coarser as well as a finer partition to guarantee the existence of all relevant depth discontinuities among their combined region borders allows the method to operate with a general and straightforward termination criterion (total number of regions) in region merging.

3 Depth fitting

Once an adequate partition hierarchy has been defined, the depth reconstruction method establishes the validity of the coarse depth estimates in terms of their fit within the regions defined in the hierarchy. Then, using this sparse set of valid depth estimates, a plane fitting procedure is used to compute a complete set of depth values within each region of support. The results are contrasted among the levels of the hierarchy and a final depth map with reconstructed estimates is produced.

As noted in Sect. 1, depth maps generally present a structure formed of smooth areas separated by discontinuities. Inaccuracies within maps usually manifest themselves as displaced discontinuities. Consider the coarse depth estimate shown in the top row of Fig. 3. Depth discontinuities are clearly misaligned with respect to objects within the scene. Nevertheless, coherent depth values may be found within the smooth areas of the corrupted map. The distortions of the type presented in this depth map can arise, for example, upon up-sampling a lower resolution depth estimate originating from range sensing devices.

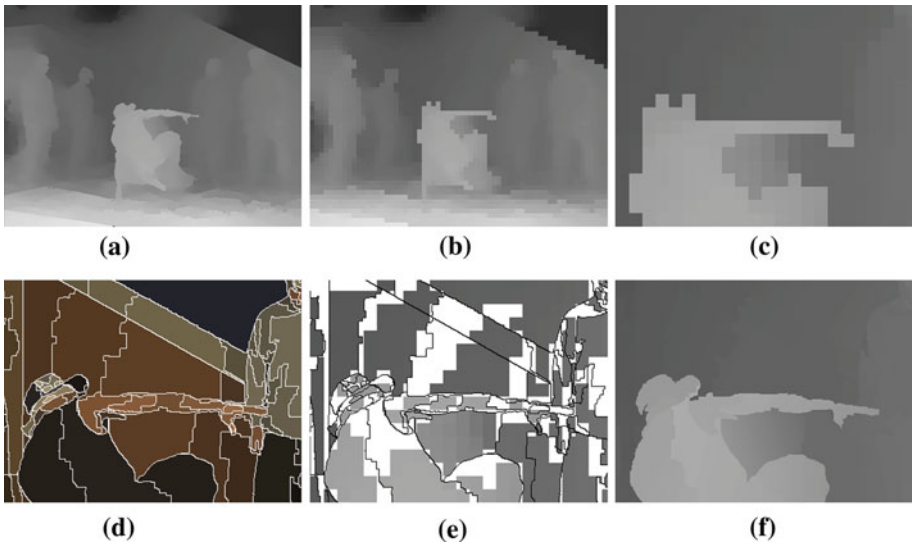

Fig. 3 **a** The original depth map for the *Breakdancers* sequence (view 2, frame 0), **b** coarse (corrupted) depth estimate, detail crops of **c** the coarse estimate, **d** the color-based partition, **e** valid depth estimates within each region (eliminated estimates in *white*) and **f** the reconstructed depth estimate

Depth fitting commences by identifying the smooth areas, or flat zones, within a coarse depth estimate. Flat zones are defined as regions of strictly flat intensity, i.e., constant value. For each region of a color-based image partition, see Fig. 3d as an example, the coarse estimates whose flat zones are contained in at least 50 % within such region are deemed as valid for that region. This fitting procedure, illustrated in Fig. 3e, selects for each region of the image partition, a set of valid coarse estimates which respect the proposed color-based region boundaries. Depth estimates whose flat zones are not majoritarily contained within the region are eliminated. Fitting requirements are relaxed for small regions which contain no depth flat zones by at least 50 %. In this case, the largest of the flat zones is deemed as valid.

Valid estimates selected from the coarse depth map are used to determine the best fit depth plane for each region of support. Assuming the depth value z is functionally dependent on the x and y coordinates, the plane $z = ax + by + c$ which best fits the set of given samples (x_i, y_i, z_i) within a region is determined by minimizing the sum of squares errors between z_i and the plane values $ax_i + by_i + c$. Once the a, b and c plane coefficients have been determined through least squares fitting, a set of best depth fit estimates is computed for the entire region. These results may be viewed in Fig. 3f for the given example and compared to the corrupted estimates displayed in Fig. 3c. Note that the use of planes can correctly capture gradual transitions in depth such as those presented in the floor area of the *Breakdancers* sequence.

The reconstruction algorithm computes a set of best depth fit estimates for each of the partitions within the hierarchy: (i) the coarse partition (typically 500 regions) and (ii) the fine partition (typically 2,000 regions). Due to their generally smaller sizes, regions within the fine partition are mostly fitted with frontal-parallel planes, i.e., constant depth values. The reconstruction algorithm resorts to the fine layer to recover from possible over-segmentation errors present within the coarse partition layer. Possible segmentation errors are detected by comparing average absolute differences between each set of best fit depth estimates

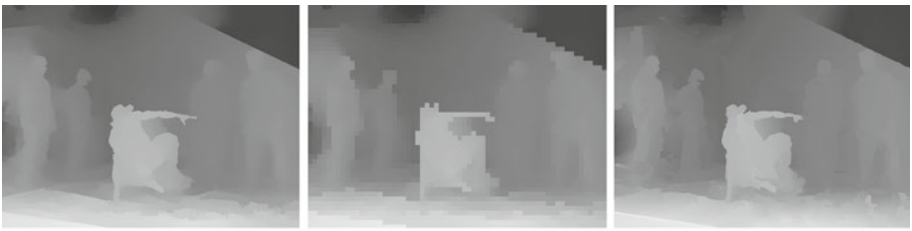

Fig. 4 **a** The original depth map for the *Breakdancers* sequence (view 2, frame 0), **b** coarse (corrupted) depth estimate with 16×16 blocks and **c** the final reconstructed depth estimate

(originating from coarse and from fine partitions) for each region of the fine partition. Large differences ($>\delta$) indicate the fine region is likely to belong to another neighboring coarse region. In this case, the fine region is incorporated by the neighboring coarse region with which it shares the largest common perimeter and assumes its corresponding depth plane model. Examples of this correction procedure are further discussed in Sect. 4. The best fit depth estimates from the corrected coarse partition form the final reconstructed depth map.

4 Experimental results

The proposed depth map reconstruction method was tested on publicly available data sets *Breakdancers* and *Ballet* (Zitnick et al. 2004). Enhancement potential of the proposal for view synthesis was demonstrated on horizontally-aligned data sets *Akko & Kayo*, *Balloons* (Tanimoto Lab at Nagoya University) and *Poznan Street* (Domanski et al. 2009) as specified by Ref. "Call for proposals on 3D video coding technology" (2011) as well. All data sets have been furnished with depth maps, henceforth called the original depth maps. In all tests, coarse and fine partitions with 500 and 2,000 total number of regions, respectively, are employed. Color weights are set to $w_Y = w_U = w_V = 1/3$, shape complexity proportion $\alpha = 0.25$ and the average absolute difference threshold between depth estimates is $\delta = 20$ in all experiments.

Coarse depth estimates are simulated by producing severe blocking effects on the originally provided depth maps. Original depths are substituted by their median values within 16×16 blocks as has been depicted in Fig. 3 for the example involving the *Breakdancers* sequence (view 2, frame 0) and reproduced in Fig. 4 for comparison purposes. The complete reconstructed depth map is presented in Fig. 4c. The proposed method eliminates the severe blocking effects, correctly aligning depth discontinuities with object boundaries. The improvements are appreciable along the stronger depth discontinuities but also noticeable around weaker discontinuities such as the spectators in the background.

A similar setup was used to reconstruct coarse depth estimates for the *Ballet* sequence (view 2, frame 0). The image, coarse partition and various depth estimates are presented in Fig. 5. Results from the best depth fit estimates over the coarse partition of Fig. 5d are depicted in Fig 5e. Notice that segmentation errors within the coarse partition have a strong influence over the depth fitting results. The protrusions around the shoulders of the dancers, which are part of the background, assume depth plane values from the foreground objects. By resorting to the fine partition layer, these regions are identified and correctly incorporated by the background areas. The final reconstructed depth map is shown in Fig. 5f. The plane fitting

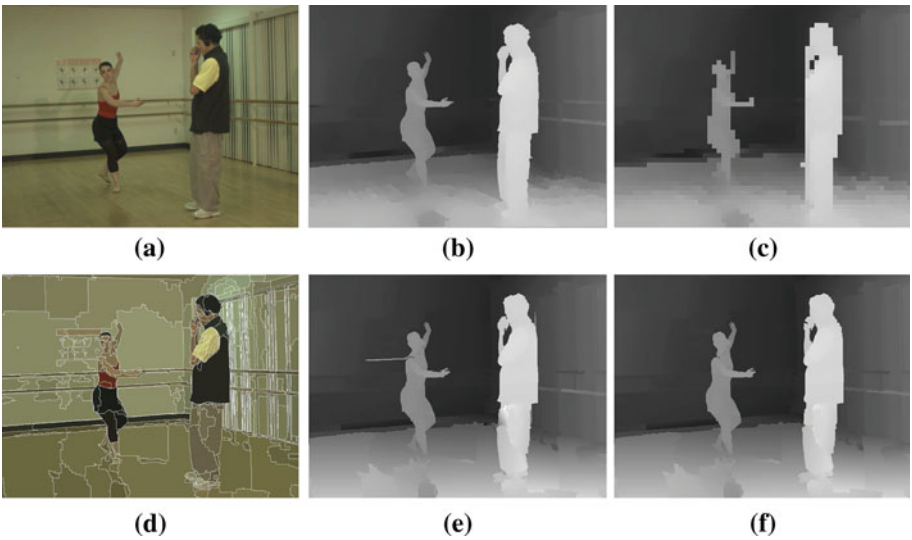

Fig. 5 **a** Image from *Ballet* sequence (view 2, frame 0), **b** original depth map, **c** coarse (corrupted) depth estimate, **d** color-based partition (500 regions), **e** best depth fit from the coarse partition and **f** the final reconstructed depth estimate

Table 1 PSNR and mean SSIM of corrupted depth maps and maps resulting from usage of fine partition, coarse partition and final reconstruction (both partitions) for the *Breakdancers* sequence

Depth map	PSNR (dB)	SSIM
Corrupted	28.81	0.947
Fine partition	29.17	0.961
Coarse partition	29.02	0.963
Final reconstruction	29.17	0.964

procedures allow the reconstructed map to accurately follow the gradual depth transitions seen along the walls of the scene, for example.

The high-quality depth maps originally provided by Zitnick et al. (2004) are used as ground truth for measuring quality gains of the reconstructed depth maps. Tables 1 and 2 present the PSNR and mean SSIM (Wang et al. 2004) of the (i) corrupted depth map, (ii) results from depth fitting over the coarse partition (500 regions), (iii) depth fitting over the fine partition (2,000 regions) and (iv) final reconstructed map using both partition layers, all with respect to the mentioned ground truth depth map. For both sequences there are clear quality gains in PSNR and less significant SSIM gains from using either coarse or fine partitions to improve the corrupted maps. For the *Breakdancers* sequence, the final reconstructed map is of approximately the same PSNR of the map resulting from usage of fine partition only. The gains of using the multiple partition layers are appreciated in Table 2 where the final reconstructed depth map is superior to depth maps resulting from usage of either fine or coarse partitions only.

The second set of experiments explore the effects of depth map reconstruction upon view synthesis results. For this, the MPEG View Synthesis Reference Software (VSRS) (2008), version 3.0, was employed along with the previously discussed *Breakdancers* and *Ballet* data sets and the horizontally-aligned, multiview plus depth sequences: *Akko & Kayo*, *Balloons*

Table 2 PSNR and mean SSIM of corrupted depth maps and maps resulting from usage of fine partition, coarse partition and final reconstruction (both partitions) for the *Ballet* sequence

Depth map	PSNR (dB)	SSIM
Corrupted	23.76	0.930
Fine partition	25.29	0.952
Coarse partition	25.64	0.952
Final reconstruction	25.91	0.955

Fig. 6 Detail crops of the synthesized view for the *Breakdancers* sequence (view 3, frame 0) using **a** original, high-quality depth maps, **b** coarse (corrupted) depth estimates and **c** reconstructed depth estimates

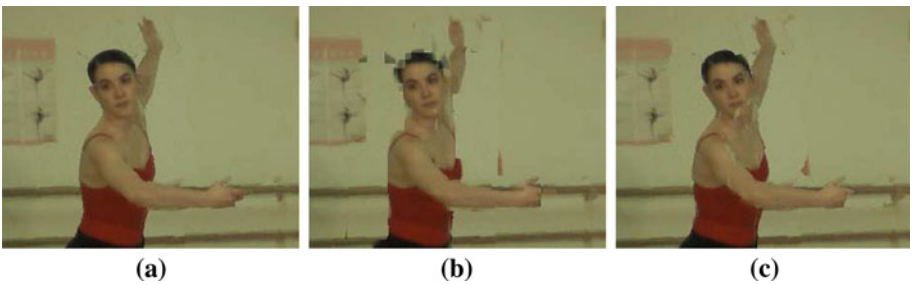

Fig. 7 Detail crops of the synthesized view for the *Ballet* sequence (view 3, frame 0) using **a** original, high-quality depth maps, **b** coarse (corrupted) depth estimates and **c** reconstructed depth estimates

and *Poznan Street*. Note that VSRS takes as inputs two neighboring views and two corresponding depth maps to synthesize an intermediate view. In these experiments, the input depth maps to VSRS are either both coarse (corrupted) or both reconstructed. Although significant quality gains in view synthesis are achieved by VSRS with its usage of adjacent views, this experimental setup is intended to illustrate the improvements of the proposed single-view depth reconstruction algorithm with respect to the coarse estimates.

Synthesis results for view 3 of the *Breakdancers* sequence depicted in Fig. 6 are obtained by using adjacent views 2 and 4 and their corresponding depth maps. The original, coarse and reconstructed depth maps of view 2, presented in Fig. 4, and their counterparts for view 4 are employed. In spite of the high quality of the original maps, corona-like artifacts (Muller et al. 2008) are perceivable along the vicinities of the dancer's contours in Fig. 6a. These, and blocking artifacts due to the imposed depth corruption, are greatly aggravated in Fig. 6b. View synthesis with the reconstructed depth maps is seen to substantially lessen distortions along the depicted contours in Fig. 6c.

Objective quality differences among synthesized views for *Breakdancers* are reported in Table 3. The PSNR and mean SSIM values are measured with respect to the corresponding view captured by a camera at the same location and orientation chosen for synthesis.

Table 3 PSNR and mean SSIM of synthesized view for the *Breakdancers* sequence (view 3, frame 0) using original depth maps, corrupted depth maps and maps resulting from usage of fine partition, coarse partition and final reconstruction (both partitions)

Depth maps for synth.	PSNR (dB)	SSIM
Original	31.99	0.941
Corrupted	31.45	0.936
Fine partition	31.51	0.937
Coarse partition	31.50	0.937
Final reconstruction	31.50	0.937

Table 4 PSNR and mean SSIM of synthesized view for the *Ballet* sequence (view 3, frame 0) using original depth maps, corrupted depth maps and maps resulting from usage of fine partition, coarse partition and final reconstruction (both partitions)

Depth maps for synth.	PSNR (dB)	SSIM
Original	30.17	0.946
Corrupted	28.95	0.929
Fine partition	29.32	0.936
Coarse partition	29.01	0.933
Final reconstruction	29.17	0.935

Synthesis results are presented when using the original depth maps of views 2 and 4, corrupted versions of the maps and reconstructed versions at several stages. The tested reconstruction stages are those resulting from depth fitting with coarse partitions (500 regions) only, with fine partitions (2,000 regions) only and final reconstruction using both partition layers. As observed in Table 3, both PSNR and SSIM values of synthesis with maps from final reconstruction is only marginally superior to those of corrupted maps in spite of the significant visual quality differences among them. When comparing final reconstruction against the intermediate reconstructions (with fine and coarse partitions only), PSNR and SSIM results are approximately equivalent.

A similar setup to that of *Breakdancers* is used for visual evaluation of the proposed algorithm on *Ballet*. Synthesis results for view 3 are provided in Fig.7 wherein the original, corrupted and reconstructed maps of view 2, shown in Fig. 5b, c, f, along with their view 4 counterparts are employed. The blocking effects visible around the head and torso of the ballerina in Fig. 7b are seen to be significantly reduced when using the reconstructed maps in Fig.7c. Due to the severity of the inflicted depth corruption, artifacts are still visible in the background and amidst some foreground areas, however, object boundaries in Fig. 7c are in conformance with those attained from the original maps shown in Fig. 7a.

A setup analogous to that described for Table 3 is used in determining the PSNR and mean SSIM values for the *Ballet* sequence summarized in Table 4. As expected, usage of reconstructed maps outperforms those of corrupted maps in synthesis. Nevertheless, results from final reconstruction are slightly inferior to those obtained when using maps reconstructed with fine partitions only. These synthesis results come in spite of the superiority, in terms of depth map PSNR and SSIM, of the final reconstruction maps over the maps obtained with fine partition only for views 2 (see Table 2) and 4.

View synthesis results for the *Akko & Kayo* sequence (view 48, frame 0) are presented in Fig. 8. For this and the following horizontally-aligned multiview sequences, the original depths are substituted by their median values within 32×32 blocks in order to form the coarse depth estimates. Given the coarse estimate of Fig. 8b, the proposed method adequately aligns depth discontinuities to the objects in the scene, in this case, the head contour perceived in

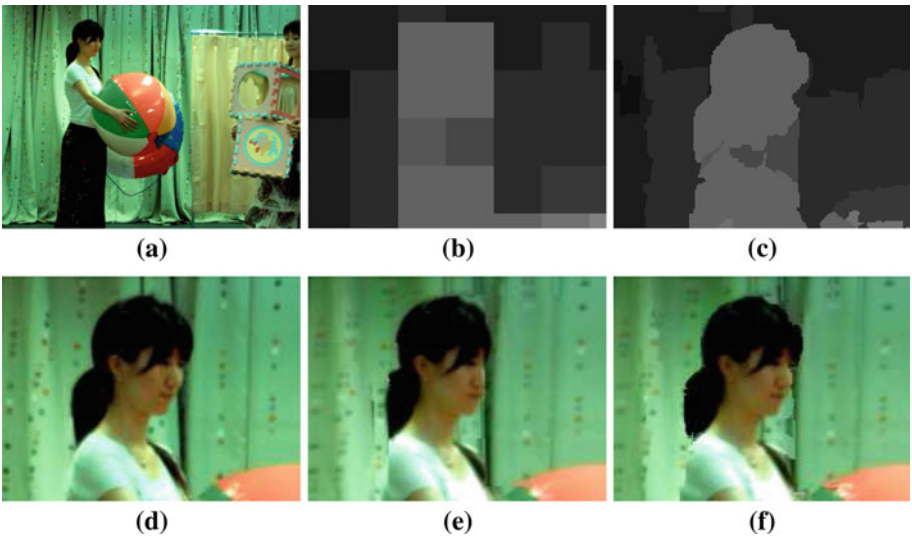

Fig. 8 **a** Image from the *Akko & Kayo* sequence (view 48, frame 0) and detail crops of the head area from the **b** coarse (corrupted) depth estimate and **c** reconstructed depth estimate from view 47. Detail crops of **d** the captured view and synthesized view 48 with **e** coarse depth estimate and **f** reconstructed depth estimate

Table 5 PSNR and mean SSIM of synthesized view for the *Akko & Kayo* sequence (view 48, frame 0) using original depth maps, corrupted depth maps and maps resulting from usage of fine partition, coarse partition and final reconstruction (both partitions)

Depth maps for synth.	PSNR (dB)	SSIM
Original	28.07	0.837
Corrupted	26.88	0.828
Fine partition	27.33	0.831
Coarse partition	27.14	0.825
Final reconstruction	27.38	0.830

Fig. 8c. Misalignments are still appreciable in the neck area and background curtains, primarily due to the low-quality of the coarse depth input. View synthesis results illustrate the quality difference between coarse and reconstructed depth estimates. Portions of the head area and back can be seen missing from Fig. 8e while correctly reproduced in Fig. 8f.

Objective quality metrics for the synthesized results for *Akko & Kayo* are reported in Table 5. PSNR of synthesis with maps from final reconstruction is superior to those of maps from fine and coarse partitions and approximately 0.5 dB above synthesis with corrupted maps. In terms of SSIM, the final reconstruction outperforms the corrupted version as well. Due to the severity of inflicted depth map corruption, synthesis form reconstructed versions are still inferior in quality to that from the relatively low-quality original maps.

For the Balloons sequence (view 4, frame 0) of Fig. 9, reconstruction can be seen to properly align depth discontinuities to the outer object contours. Nevertheless, depth differences of the background region above the performers shoulders with respect to the foreground are subtle in the coarse estimate of Fig. 9b and lost in reconstruction depicted in Fig. 9c. View synthesis with the coarse estimates can be seen to contain artifacts around the outer contours of the performers hand and arm, see Fig. 9e, which are not present in the image synthesized

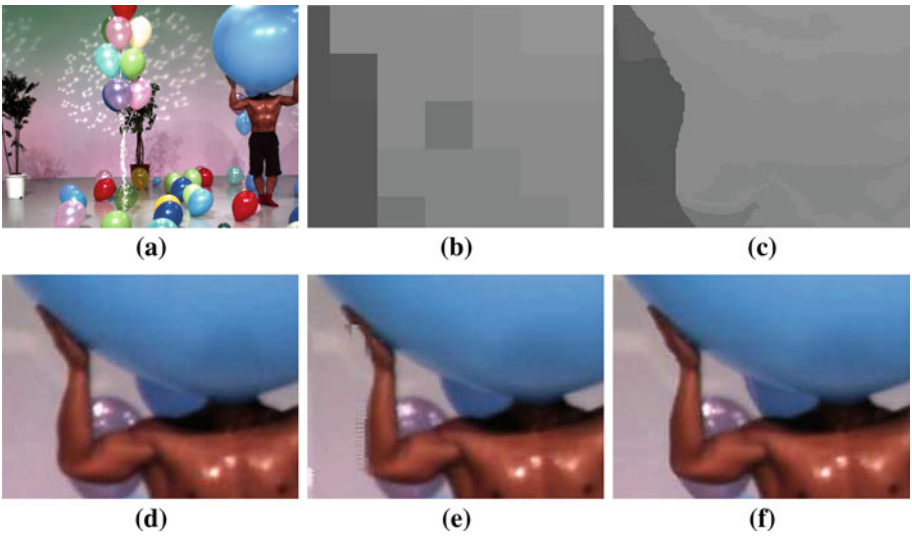

Fig. 9 **a** Image from the *Balloons* sequence (view 4, frame 0) and detail crops of the performer from the **b** coarse (corrupted) depth estimate and **c** reconstructed depth estimate from view 3. Detail crops of **d** the captured view and synthesized view 4 with **e** coarse depth estimate and **f** reconstructed depth estimate

Table 6 PSNR and mean SSIM of synthesized view for the *Balloons* sequence (view 4, frame 0) using original depth maps, corrupted depth maps and maps resulting from usage of fine partition, coarse partition and final reconstruction (both partitions)

Depth maps for synth.	PSNR (dB)	SSIM
Original	29.89	0.963
Corrupted	25.86	0.942
Fine partition	26.49	0.945
Coarse partition	27.71	0.954
Final reconstruction	27.85	0.956

with reconstructed estimates, Fig. 9f. This image is similar in quality to the captured view of Fig. 9d.

PSNR and mean SSIM values for the *Balloons* sequence in Table 6 reveal the superiority of synthesis with the final reconstructed maps over those obtained from using either fine or coarse partitions only. Significant gains of approximately 2 dB are achieved with respect to the corrupted maps, reflecting the subjective quality gains ascertained from visual inspection of the synthesized images. Unlike the case for the *Akko & Kayo* sequence, results from using the coarse partitions are superior to those from using fine partitions for *Balloons*.

Synthesis results for the *Poznan Street* sequence (view 4, frame 0) are shown in Fig. 10. Visual inspection of coarse and reconstructed depth maps demonstrate improvements although some depth and object misalignments are still perceivable in Fig. 10c due to erroneous segmentation in both of the adopted fine and coarse partitions. Synthesis results show significant artifacts around the car's contours in Fig. 10e which are greatly reduced by using reconstructed depths in Fig. 10f.

Lastly, PSNR and mean SSIM values of synthesized views for *Poznan Street* are presented in Table 7. Recorded PSNR values resulting from usage of final reconstruction maps are slightly inferior to those of corrupted maps. The mean SSIM values, on the other hand,

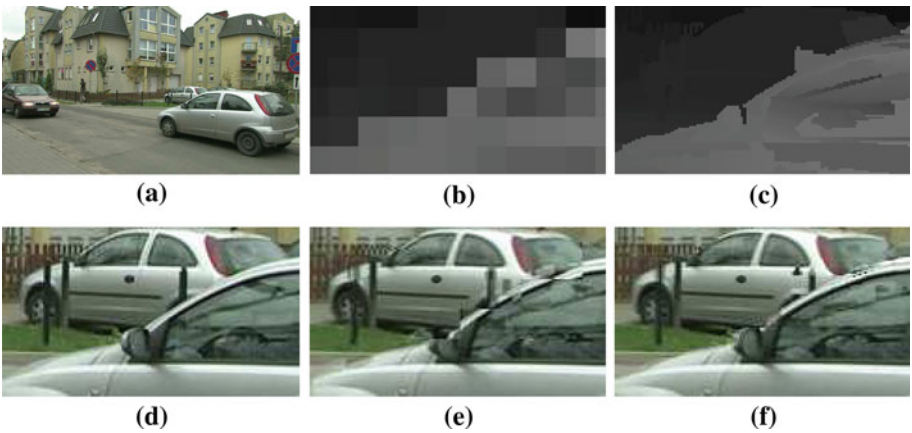

Fig. 10 **a** Image from the *Poznan Street* sequence (view 4, frame 0) and detail crops of the car from the **b** coarse (corrupted) depth estimate and **c** reconstructed depth estimate from view 3. Detail crops of **d** captured view and synthesized view 4 with **e** coarse depth estimate and **f** reconstructed depth estimate

Table 7 PSNR and mean SSIM of synthesized view for the *Poznan Street* sequence (view 4, frame 0) using original depth maps, corrupted depth maps and maps resulting from usage of fine partition, coarse partition and final reconstruction (both partitions)

Depth maps for synth.	PSNR (dB)	SSIM
Original	35.14	0.882
Corrupted	32.84	0.861
Fine partition	32.74	0.864
Coarse partition	31.88	0.857
Final reconstruction	32.64	0.861

present a more faithful representation of the perceived visual quality noticeable in Fig. 10. Synthesis results with the fine partitions are superior in terms of SSIM to those form the corrupted maps, while results with the final reconstructed maps are equal in SSIM to the corrupted ones. In spite of the differences among these objective quality metrics, neither reflect the significant subjective quality gains in synthesis obtained by employing reconstructed depth maps over corrupted ones.

5 Conclusion

This paper presents a novel depth map reconstruction method which takes as inputs a single view and a coarse depth estimate. Color-based hierarchical image partitions are built and represent the set of regions over which depth map estimates may be refined while respecting the proposed color borders within the partitions. The use of a hierarchy grants the method greater robustness to segmentation errors common within higher level (coarser) image partitions. Reconstructed estimates are determined through plane-fitting procedures which extend well beyond strong discontinuity areas and are capable of modeling gradual depth transitions as well. Results demonstrate the potential of the algorithm in correctly recovering from severely corrupted depth estimates such as those encountered when employing off-the-shelf range sensor systems with coarse and/or low resolution measurements. As future work we

intend to evaluate algorithm performance when applied towards the adaptation and alignment of non-registered pairs of color images and depth maps.

Acknowledgments This work was partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under grants 300223/2008-0, 470940/ 2010-7 and 310375/2011-8, and by DPP/Universidade de Brasília.

References

Call for proposals on 3d video coding technology. ISO/IEC JTC1/SC29/WG11 MPEG 2011/N12036, 2011.

Domanski, M., Grajek, T., Klimaszewski, K., Kurc, M., Stankiewicz, O., Stankowski, J., et al. (2009). *Poznan multiview video test sequences and camera parameters*. ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050.

Dorea, C., Pardas, M., & Marques, F. (2009). Trajectory tree as an object-oriented hierarchical representation for video. *IEEE Transactions on Circuits and Systems for Video Technology*, *19*(4), 547–560.

Dorea, C., & de Queiroz, R. L. (2011). Depth map reconstruction using color-based region merging. In *ICIP*.

Ekmekcioglu, E., Velisavljevic, V., & Worrall, S. T. (2009). Edge and motion adaptive median filetring for multi-view depth enhancement. In *Proceesings of PCS*.

Garrido, L., & Salembier, P. (1998). Region based analysis of video sequences with a general merging algorithm. In *EUSIPCO*.

Huihai, L., Woods, J. C., & Ghanbari, M. (2007). Binary partition tree analysis based on region evolution and its application to tree simplification. *IEEE Transactions on Image Processing, 16*(4), 1131–1138.

Kauff, P., Atzpadin, N., Fehn, C., Muller, K., Schreer, O., Smolic, A., et al. (2007). Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. *Image Communications, 22*, 217–234.

Kolb, A., Barth, E., Koch, R., & Larsen, R. (2009, March). Time-of-flight sensors in computer graphics. *Eurographics (state-of-the-art report)* (pp. 119–134).

Merkle, P., Smolic, A., Muller, K., & Wiegand, T. (2007). Multi-view video plus depth representation and coding. In *Proceedings of PCS*.

Muller, K., Smolic, A., Dix, K., Merkle, P., Kauff, P., & Wiegand, T. (2008). View synthesis for advanced 3D video systems. *EURASIP Journal on Image and Video Processing, 2008*, 1–11.

Oh, K. J., Yea, S., Vetro, A., & Ho, Y. S. (2009). Depth reconstruction filter and down/up sampling for depth coding in 3D video. *IEEE Signal Processing Letters, 16*(9), 747–750.

Reference software of depth estimation and view synthesis for FTV/3DV: ISO/IEC JTC1/SC29/WG11, M15386. October 2008.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision, 47*(1–3), 7–42.

Tanimoto Lab at Nagoya University. (2008). http://www.tanimoto.nuee.nagoya-u.ac.jp/.

Torres-Mendez, L. A., & Dudek, G. (2004). Reconstruction of 3D models from intensity images and partial depth. In *Proceecings of American Association for Artificial Intelligence*.

Wang, L., Jin, H., Yang, R., & Gong, M. (2008). Stereoscopic inpainting: Joint color and depth completion from stereo images. In *IEEE Proceedings of CVPR*.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error measurement to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612.

Yang, Q., Yang, R., Davis, J., & Nister, D. (2007). Spatial-depth super resolution for range images. In *IEEE Proceedings of CVPR*.

Zhao, Y., Chen, Z., Tian, D., Zhu, C., & Yu, L. (2010). Suppressing texture-depth misalignment for boundary noise removal in view synthesis. In *Proceedings of PCS*.

Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S., & Szeliski, R. (2004). High-quality video view interpolation using a layered representation. *ACM SIGGRAPH and Transactions on Graph, 23*, 600–608.

Author Biographies

Camilo Dorea received the B.S. degree from the University of Brasilia, Brazil, in 1997, the M.S. degree from the University of Maryland at College Park, USA, in 1999, both in electrical engineering, and the Ph.D. degree in telecommunications from the Technical University of Catalonia (UPC), Barcelona, Spain, in 2007. From 2007 to 2008 he was with Thomson Corporate Research at Princeton, NJ, USA. In 2009 he joined the Department of Computer Science at the University of Brasilia where he is currently Assistant Professor. His research interests include video segmentation and analysis, video coding, and mutiview and 3D processing.

Ricardo L. de Queiroz received his Engineer degree from Universidade de Brasilia, Brazil, in 1987, his M.Sc. from Universidade Estadual de Campinas, Brazil, in 1990, and Ph.D. from The University of Texas at Arlington, in 1994, all in electrical engineering. From 1990 to 1991, he was with the DSP research group at Universidade de Brasilia, as a research associate. He joined Xerox Corp. in 1994, where he was a member of the research staff until 2002. From 2000 to 2001, he was also an adjunct faculty member at Rochester Institute of Technology. He joined the Electrical Engineering Department at Universidade de Brasilia in 2003, and in 2010, he became a full professor in the Computer Science Department. He has published over 140 articles in journals and conferences and has contributed chapters to books as well. He also holds 46 issued patents. He is an elected member of the IEEE Signal Processing Society's Image, Video and Multidimensional Signal Processing (IVMSP) and Multimedia Signal Processing (MMSP) Technical Committees. He is a past editor for the *EURASIP Journal on Image and Video Processing*, *IEEE Signal Processing Letters*, *IEEE Transactions on Image Processing*, and *IEEE Transactions on Circuits and Systems for Video Technology*. He has been appointed an IEEE Signal Processing Society Distinguished Lecturer for the 2011–2012 term. His research interests include image and video compression, multirate signal processing, and color imaging. He is a senior member of IEEE, and a member of the Brazilian Telecommunications Society and the Brazilian Society of Television Engineers.