An efficient and memory-conserving implementation of multi-view stereo for wide-area reconstruction

Xenophon Zabulis, Nikos Grammalidis, and Georgios D. Floros

Abstract

This paper deals with the automatic stereo reconstruction of wide-area scenes. Its particular goal is a computationally efficient method that can be performed on a personal computer, despite the large amount of data involved in the reconstruction of wide-area scenes. Robustness is considered in terms of the accuracy of the final reconstruction, as well as, in the context of simplifying the image acquisition process for the end-user.

1. Introduction

The large number of images and broad extent of the reconstruction volume in wide-area stereo raise demanding computational requirements in processing power and representation capacity, respectively. At the same time, the increase in resolution of cameras raises corresponding demands, regarding the detail of 3D reconstructions that can be obtained with an off-the-shelf camera. Our goal is the ability to process the large amount of data required for reconstructing a wide-area scene on a conventional computer, while fully utilizing multiple images of arbitrary high resolution.

In this context, the application of stereo methods that simultaneously consider all input views becomes quite complicated. The difficulties are due to the inability to fit the images and 3D representation in memory, which are required by global operations such as visibility reasoning. In the literature, wide-area scenes are treated by individually processing stereo views and registering the partial results on a common reference frame. A view-combination step is applied subsequently to stereo, to cope with the effects of view overlap and registration inaccuracy.

Space-sweeping is an efficient stereo algorithm that provides adequately complete reconstructions for well-textured surfaces (see Sec. 2). It has been utilized in multi-view stereo e.g. [12, 20], as a means of coping with large amounts of data, in a timely fashion. For the same purpose several hardware-accelerated implementations for the GPU exist, e.g. [25, 14, 11]. In this paper, an extension to this algorithm is proposed that targets its acceleration and its execution within an arbitrary small amount of memory capacity; the extension is algorithmic and, thus, equally applicable to a conventional or hardware-accelerated implementation. Our additional interest in memory-conservation is twofold. First, because processing of high-resolution images typically exceeds the available memory capacity. Second, because state-of-the-art graphics hardware has even less onboard memory than a personal computer and, thus, the processing of high resolution images in the GPU is not straightforward.

To cope with the large number of reconstruction points, we employed a volumetric and local view-combination approach. In order to fit a finely interpolating surface to the voxelized results of view combination, we considered surface-fitting approaches and compared them with respect to accuracy of fit and robustness to outliers.

An additional requirement in wide-area stereo is the estimation of camera motion. Principal approaches to handheld camera motion are based on tracking features in consecutive images of a sequence to estimate camera motion. Images that are acquired from viewpoints densely arranged in space and time, i.e. with a video camera [19] or multiple ones [20], facilitate robust camera motion estimation, despite inaccuracies in feature tracking. To simplify image acquisition we attempted a sparser arrangement of less viewpoints. This attempt was supported by the utilization of (a) high resolution images (8Mpix) and (b) an experimental conclusion showing that more discriminant features can support a sparser and less constrained "walkthrough" of camera locations.

The remainder of this paper is organized as follows. In Sec. 2 related work is reviewed and in Sec. 3, the proposed extension to the space-sweeping is presented. A memory-conserving method to efficiently combine the reconstruction results from space-sweeping is proposed in Sec. 4. We demonstrate our contributions with experiments in Sec. 5, along with an experimental evaluation of conditions that facilitate robust camera motion estimation. Summary and directions for future research are provided in Sec. 6.

2. Related work

The *space-sweeping* approach to stereo reconstruction was introduced in [7], where the depth of sparse image features (edges) was recovered, by corresponding them across multiple images with parallel optical axes. The method utilizes a hypothetical plane that is translated along the "depth" dimension. The image features are backprojected from both images onto this, "reference", plane. At a given depth, the features whose backprojections are collocated are considered as corresponding and as to occur at that depth. The algorithm was formulated for dense stereo [25, 14, 17], as follows. At a given depth, the acquired images are backprojected on the hypothetical surface and locally compared as to visual similarity, by e.g. a correlation metric. The comparison evaluates the similarity of pairs of patches centered at the same coordinates in the backprojected images. It then stores the result for each such pair in a 2D similarity image, also at the same coordinates. For the series of similarity values of a pixel along depth, the depth at which similarity is maximized is an estimate of the depth of the surface in the direction of that pixel. The output is image M, which is called a *depth map* because each of its pixels indicates the estimated depth for the corresponding spherical coordinate. Curved alternatives of the sweeping plane have been proposed in [21, 28], to increase accuracy at image periphery.

Space-sweeping is not the most accurate approach to stereo reconstruction. A weakness of this method is that it exhibits reduced accuracy at the reconstruction of oblique surfaces relative to the sweeping direction, due to perspective distortion. To cope with this issue, multiple sweeping directions have been proposed in [10]. In order to compensate for projective distortion and achieve better matches, the backprojections of the acquired images are evaluated on multiple planar surfaces that assume a range of orientations in space and the most fitting are selected [9, 4, 26] . On the other hand, considering multiple orientations significantly increases the amount of computation.

Several methods have been proposed for the combination of partial reconstructions acquired by individual stereo views (or LIDAR scanners); a comprehensive review of "surface fusion" methods can be found in [20]. Most relevant to this paper is the work in [8], which is based on a volumetric representation, that implicitly defines the reconstruction surface as the zero level-set of a functional defined in 3D space. The method was utilized in [12] for the merging of individual stereo results. In [27], this functional was determined by the similarity value of stereo-matching and the Radial Basis Function (RBF) framework in [5] was utilized to finely approximate the pursued zero level-set. In a different approach [18], an error-minimization strategy that fits a surface into an unorganized set of points was proposed.

Finally, a component of a convenient approach to wide-area stereo is the estimation of camera motion, in order to avoid the use of marker points or additional localization modalities. This paper partially adopts the approach in [1, 19], which robustly utilizes the tracking of Harris features [13] in a video sequence to estimate camera motion. Guided by the result in [24], SIFT features [16] are utilized instead, to establish correspondences across.

3. Extension of space-sweeping

The proposed extension to space-sweeping achieves the following two goals. First, the ability to execute the algorithm for a large image given an arbitrary small capacity of memory. Second, the computational acceleration of the technique. Before proceeding, notation is introduced.

The version of the technique introduced in [28] is adopted, in which the backprojection plane is substituted by a spherical sector S that projectively expands from the cyclopean eye $\vec{o}$ outwards. Thus, a line of sight $\vec{v}$ departing from the cyclopean eye $\vec{o}$ is always perpendicular to the backprojection surface and, as a result, the method provides more accurate results in the periphery of the image.

The concentric instances of the backprojection sector S at depth values d_δ, where $\delta \in \{1, 2, 3, ..., n\}$ are noted as S_δ. The set of d_δ's values is called depth range $\mathcal{D}$. These values are exponentially increasing, so that the images' depth granularity is fully exploited, while a minimum number of depth values is evaluated [6]. Points on S_δ are parameterized by an angular step of c and determined by spherical coordinates ψ and ω. To generate sectors S_i, a corresponding sector S_0 is first defined on a unit sphere. A point $\vec{p} = [x\, y\, z]^T$ on S_0 is given by: $x = sin(\psi)$, $y = cos(\psi)sin(\omega)$, $z = cos(\psi)cos(\omega)$. Its corresponding point $\vec{p'}$ on S_i is then:

$$\vec{p'} = d_i \left[R_z(-\psi)R_y(-\omega)\vec{p} + \vec{o}\right], \quad (1)$$

where R_y and R_z are rotation matrices for rotations about the yy' and zz' axes. The backprojection images are locally compared with a $w \times w$ correlation kernel $\mathcal{K}$, which yields a similarity score s. The strongest local maximum of s along a line of sight indicates the estimated depth. The requirement of locality for this maximum introduces robustness to spurious maxima and textureless regions. It, however, requires that the previous and next values of s are recorded for each depth and, therefore, thrice the memory capacity.

3.1. Memory conservation

Although that the pair of high resolution images may fit in memory, the use of auxiliary representations for depth, 3D coordinates, and color of reconstruction points quickly reduce the available capacity.

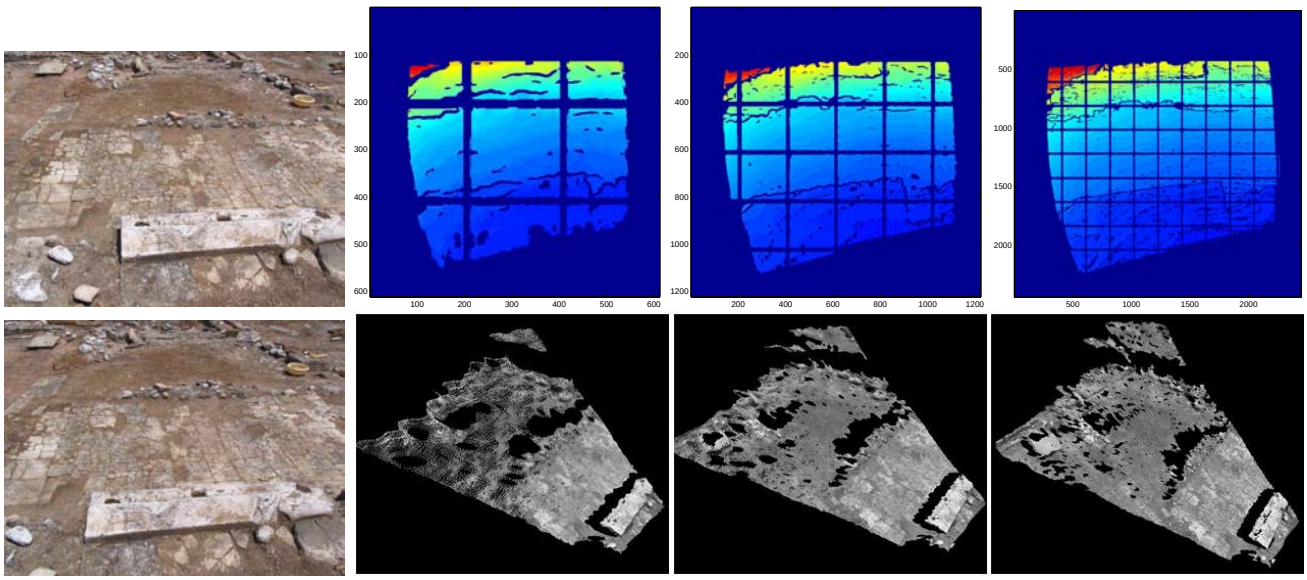

Figure 1. Illustration of the proposed extension to space-sweeping for 3 iterations.

Memory conservation is achieved by tessellating the backprojection surface $\mathcal{S}$. Each such segment is independently swept along depth and the results are merged into a single depth map. An individual depth map is obtained for each segment and, at the end, these depth maps are concatenated to produce the total depth map for the image pair. The number of segments is determined by the available memory capacity and can be arbitrary large.

The tessellation procedure partitions $\mathcal{S}$ into $\kappa \times \kappa$ spherical segments of equal size, which are henceforth called sectors. It is performed along the two spherical parameterization coordinates ψ and ω, which respectively correspond to image width and height. Sectors overlap at their boundaries, so that the support samples for the completion of kernel $\mathcal{K}$ are available and the computation of s at these locations is possible. The amount of angular overlap, so that each point in the depth map is computed only once, is $c\,\lfloor w/2 \rfloor$.

3.2. Coarse-to-fine acceleration

The proposed acceleration technique is coarse-to-fine and iterative, with N being the - a priori defined - number of iterations. Iterations initiate processing from coarser samplings of the stereo images and proceed with finer ones; at the last iteration the original image is processed. The method utilizes depth maps from coarser scales to accelerate the computation of finer ones, by restricting the evaluated depth range.

At iteration i, the image area is $\sqrt{\beta^{i-N}}$ of the original image area, meaning that in each iteration image rows and columns are reduced by a factor of β. Thus, β determines the rhythm of the coarse-to-fine progression and throughout this paper was 2. Prior to subsampling, images are Gaussianly smoothed for aliasing according to β. The result of the process is a pyramid of each stereo image. Stereo images at iteration i are noted as $I^i_{1,2}$ and their corresponding depth map M_i.

Since the resolution of images is reduced at coarser scales, two more modulations are introduced to sample images more efficiently. First, c is modulated so that the density of points on the backprojection surface adjusts in accordance to image size: $c_i = \beta^{N-i}c$. Second, the number of the - exponentially distributed - depth values in $\mathcal{D}$ is increased by a factor of β at each iteration: $n_i = \beta^{i-N}n$. The above two modulations guarantee that all pixel samples will be utilized, while images $I^i_{1,2}$ are not oversampled. Thus, c is set so that at the finest scale, each pixel is sampled at least once.

The algorithmic procedure is as follows. In the first iteration, the top of the pyramid is accessed and a low-resolution M_1 is rapidly obtained, due to the small number of points on S. Due to the above memory-conservation technique, S was initially tessellated into $\kappa \times \kappa$ sectors. In each iteration thereafter, each sectors is recursively tessellated into $\beta \times \beta$ child sectors, so that the entire partitioning scheme over all iterations generates a tree-structure of sectors, where in the i^{th} iteration, $(\kappa\beta^{i-1}) \times (\kappa\beta^{i-1})$ partitions are processed.

After the first iteration, the depth range within which a sector is swept is restricted, reducing the amount of the required computation. Instead of evaluating the full depth range $\mathcal{D}$, the limits for each sector are derived from the depth values obtained for the parent sector in the previous iteration. These limits are the extreme values of depth in the region of M_{i-1} that corresponds to the current sector.

Acceleration stems from the fact that the area in the depth map of a child segment is only the $\sqrt{\beta}$ of the area of its parent sector. It is therefore quite likely to exhibit an even narrower depth range than its parent sector - or in the worst case the same.

During the coarse-to-fine progression, the required memory capacity to process any sector at any iteration is constant and equal to the memory required for the processing of the $\kappa \times \kappa$ root sectors (of the first iteration). Therefore, the technique does not demand for more memory as the precision of reconstruction increases.

The method is demonstrated in Fig. 1, for 3 iterations. On the left, the original stereo pair is shown. The rest of the figure, shows the depth maps (top) and 3D reconstructions (bottom) in each iterations (left to right corresponds to coarse-to-fine). The depth maps show the $(\kappa\beta^{i-1}) \times (\kappa\beta^{i-1})$ tessellation for each iteration ($\kappa = 3$, $\beta = 2$).

4 Application to wide-area multi-view stereo

The combination of individual reconstructions is performed by a local volumetric approach, so that the data can be partitioned. The result of this operation is 3D points and normals, which are utilized by a surface-fitting approach to output a mesh-represented surface.

4.1 View combination

A particular difficulty in view combination for wide-area stereo is that typical memory capacity is insufficient to store all the 3D points. An approach utilized in the literature is to sequentially fuse depth maps obtained from consecutive stereo pairs of a video sequence, thus retaining at any time a single reconstruction in memory [10]. Since we wish to utilize only a few snapshots of the scene, we attempt the utilization of all available views, given the limitation of memory capacity.

To overcome the increased demand in memory capacity that the processing of the entire volume would demand, a local view combination method is applied [26]. The method utilizes all available views and its accuracy was tested in outdoor scenes in [29], but due to memory constraints reconstructions were restricted in smaller spatial extents and coarser resolutions. The method defines a functional for each voxel in space that is based on the score of correlation matching and extracts a reconstruction at the local maxima of this functional. It can be therefore evaluated independently for each voxel.

Depending on available memory capacity, the reconstruction volume is tessellated into arbitrary small rectangular parallelepipeds, called "boxes". By processing each box individually, an arbitrary large reconstruction can be evaluated, even with a small memory space. Between boxes there is an overlap, facilitating semi-local operations in the synthesis of multiple views (e.g. local plane-fitting etc), as well as, the detection of local maxima of the functional near box boundaries.

The adopted method selects a single correlation maximum along an estimated normal of the reconstructed surface to combine views. Therefore the output of the procedure for each box is much smaller, in terms of memory capacity than the input. In addition, the recovery of the surface normal from the stereo images, instead of estimating it from the 3D points, is an additional information that is valuable in surface-fitting.

4.2 Surface fitting

With the reconstruction points significantly reduced, the fitting of surfaces throughout the whole reconstruction volume is more feasible. Based on the concept in [8], we considered two more recent approaches [5, 18] towards the fitting of an interpolating surface through scattered points (or a "point cloud"). We note that due to utilized view-combination combination approach, for each 3D point we have an estimation of its normal as well.

Although that the above methods provide some treatment of outliers and surface normals, they were developed in the context of laser scanning and are prone to the magnitude of errors in wide-area stereo. The work in [5] provides a treatment in the availability of surface normals, however minute errors in the direction of these normals create artifacts of significantly greater magnitude. The method of [18] provides of better detail with significant time-efficiency, but cannot exploit surface normals and is sensitive to outliers. To improve our results, an outlier extraction process is proposed that besides the estimated locations of 3D points considers their estimated normals too.

The proposed extension, requires that neighboring points on a surface should have continuous normals as well. The normal $\vec{\nu_i}$ of point $\vec{p_i}$ is available from the utilized stereo reconstruction method. For each point $\vec{p_i}$, we compute the vectors from itself to its k-nearest neighbors $\vec{v_{ij}} = \vec{p_j} - \vec{p_i}$, where $\vec{p}_{j=1,2\ldots,k}$ the k neighbors of $\vec{p_i}$. We then compute the angle α between the $\vec{\nu_i}$ and $\vec{v_{ij}}$ via their dot product $\vec{\nu_i} \cdot \vec{v_{ij}} = |\vec{\nu_i}||\vec{v_{ij}}| \cos\alpha$. If the surface is planar, α should be $\pi/2$. More generally, if the surface is assumed as locally planar then

$$|\alpha - \pi/2| < \tau_\alpha \tag{2}$$

should hold. The above geometry is illustrated in Fig. 2. In the figure, p_i is an outlier point and p_j is one of its k-nearest neighbors. Solid arrows indicate (estimated) surface normals and dashed ones vectors $\vec{v_{ij}}$. Our assumption is that for a surface point, the above condition should hold for the majority of its k-nearest neighbors. To model the fact that certain outliers occur at great distance from the veridical

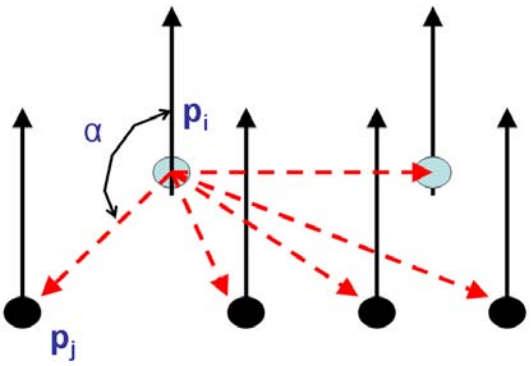

Figure 2. Illustration of the geometry of outlier extraction (see text).

surfaces proximity is accounted by assuming that the average distance, of an outlier to its k-nearest neighbors, is greater than threshold τ_d.

To avoid dealing with threshold values τ_d and τ_α they are both liberally set (e.g. $\tau_a = 20^\circ$, $\tau_d = 5$ voxels) and we depend on the combined value of the two assumptions above for the final result. To efficiently retrieve the k-nearest neighbors of $\vec{p}_i$ a kd-tree ($k = 3$) that is sorted according to distance [2], is utilized.

5 Experiments

The main experiment in this section regards the reconstruction of an outdoors $\approx 15\,m \times 30\,m$ scene, from 40 images acquired with a Canon Powershot SLR camera. The image resolution was 3272×2456 color pixels, with 2^{16} levels of gray. In the first 20 frames, the motion of the camera was smooth. In the rest of the frames, images were acquired along wider baselines, as the observer moved around the scene with greater affinity. In Fig. 3, some sample images of the sequence are shown.

In the remainder of this section, the proposed extension to space-sweeping is experimentally evaluated. Next, an experiment in camera motion estimation compares accuracy and robustness of the procedure with respect to the type of tracked image features. Finally, comparative results in surface fitting and outlier suppression conclude this section.

5.1 Sphere-sweeping extension

Stereo views were comprised of image pairs acquired at adjacent positions. The sphere sweeping method was utilized to obtain a reconstruction from each view. The visual angle of the spherical sectors was $90^\circ \times 60^\circ$. To guarantee that each image pixel was to be sampled at least once, a depth map of a (final) resolution of $\approx 10^7$ pixels was utilized. The number of reconstructed points was about 10% less, on average. The depth range $\mathcal{D}$ was comprised of 960 depth levels, exponentially distributed in depth, $w = 15$, $N = 5$, and $\kappa = 3$.

Due to memory limitations we were not able to execute the conventional version of the space-sweeping algorithm in the above conditions. For half the depth map resolution, or $1/4$ of the pixels, it lasted more than 4 hours, on a Pentium computer with 1GB of RAM. Since the complexity of the algorithm is linear to the number of pixels, the estimated time for the resolution we failed to evaluate is about 16 hours. Utilizing the proposed sphere-sweeping extension for the *full* image resolution the computation time was constrained in ≈ 2.5 hours, depending on the stereo pair (note that the exact amount of computation, is content-dependent). With respect to the conventional version the expected speedup is ≈ 7. Due to the conservative ($\beta = 2$) coarse-to-fine progression, cases where an error in coarse scale would result to the omition of part of the reconstruction did not occur. Finally, we confirmed that the fine scale result was identical to that which would have been produced by the conventional version of the method, by performing the latter for smaller parts of stereo pairs and comparing the results.

5.2 Camera motion estimation

Camera motion estimation is based on the work in [1], as elaborated for projective geometry in [19]. The Bundle Adjustment procedure in [15] was optionally invoked to improve the estimated camera motion and structure, as a final post-processing step.

In the experiment, image feature detection and matching was tested in two conditions: utilizing Harris and SIFT features. In the Harris condition, matching was correlation-based using a neighboring constraint which was based on [30] and extended in [22] to disambiguate between similar matches. In the SIFT condition, matching was as in [16]. It is noted that the computation of the SIFT features for the images required more memory than available. The image was, thus, tessellated into patches, which were independently processed, and features from each patch were merged. To avoid blocking artifacts, the patches were adequately overlapping. Duplicate features, due to block overlap, were removed in merging.

The recovered camera motions for the Harris and SIFT conditions are shown in Fig. 4. The first observation from the experiment was that in the SIFT condition greater baselines of camera motion were tolerated. In the 40-frame sequence of the experiment, the Harris condition returned very inaccurate results after a specific frame (20^{th}), where camera motion was relatively large. Increasing the spatial

Figure 3. Sample images from the acquired sequence (frame numbers from left: 0, 14, 21, 34).

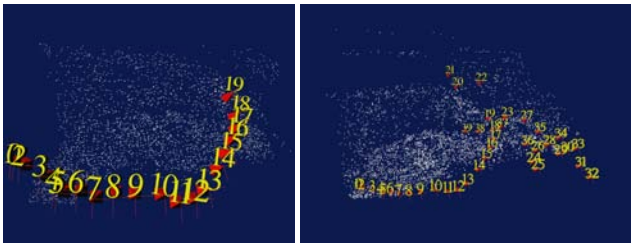

Figure 4. Estimated camera trajectory using Harris (left) and SIFT features (right).

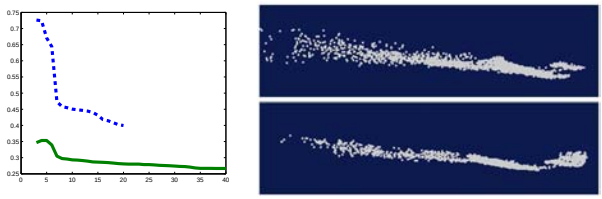

Figure 5. Comparison of reconstruction accuracy for the established correspondences in the Harris and SIFT conditions (see text).

range of search in the Harris condition did not alleviate of the problem. In contrast, an even greater increase in error was observed, due to the occurrence of more spurious matches. In Fig. 7, the viewpoints of some camera locations are indicated into a colorized top view of the total reconstruction.

The second observation regards the number and accuracy of the corresponded features. In the SIFT condition more correspondences (about 22% on average) were established across frames. The mean reprojection error (measured in pixels) is shown in Fig. 5(left), as a function of time (number of input frame). The Harris plot (dashed line) terminates earlier than the SIFT plot, due to the error at the frame 20. Accuracy can be also visually confirmed in Fig. 5(right), which shows the reconstructed 3D locations of the established correspondences, in a side-view and for the two conditions (top: Harris, bottom: SIFT). By observing the spatial variance of these points and the considering approximately planar structure of the ground, it is concluded that SIFT-based correspondences approximate better the imaged surface.

We observed the same behavior in several sequences without in any case the Harris condition performing better. The consequent tolerance to relatively large baselines or angle changes is important, because we wish to reconstruct scenes from a only few snapshots instead of a video sequence.

5.3 Surface-fitting

After combining (see Sec. 4) the results from the 40 stereo views of the SIFT condition, we compared the surface-fitting methods in Fig. 6.

The results from the RBF interpolation [5] show that although this method constructs smooth surfaces, the accuracy of the reconstruction is seriously affected from the presence of outlier points. In the experiments, we initially utilized the *errorbar* method of this approach to introduce tolerance against outliers and errors in surface normal estimations. However, the interpolating surface would create protrusions in order to interpolate through, or near, outlier points. We then tested the *reduced* method of the approach, which was the one to perform best (shown in Fig. 6). The technique attempts to find a better fit by reducing the number of pivot points, however oversimplifies the result or exhibits the same artifacts as above. Another issue with the RBF approach was that the method would necessarily output a single interpolating surface and, thus, introduce surface in regions where input points would not occur.

Ohtake's method [18] yielded more accurate reconstructions in significantly shorter period ($\approx 10\times$) than [5], but was prone to outliers. In their occurrence it created "holes" or small pyramidic structures that protruded from the main surface, to fit through remaining outliers. The proposed, in Sec. 4.1, preprocessing step applied to the above approach, is shown in Fig. 6, and it is observed that the proposed outlier suppression technique improves the result fitting pro-

cess, produces less holes, and reconstructs a greater portion of the scene. The detail images show the round structure (on the right), which is shown in the bottom-right of frame #14 in Fig. 3.

6 Summary and future work

In this paper an approach to wide-area multi-view stereo is proposed that is automatic and can be performed with of-the-shelf computational resources and materials. Our goal is the acquisition of images from an arbitrary large area and the automatic generation of the 3D model of the scene, in a personal computer. This paper contributes on the time-efficient and memory conserving reconstruction of the scene. In order to support a convenient image acquisition process, it tests the possibility of acquiring a few snapshots with a high-resolution camera. In this effort, SIFT features facilitate a more affine trajectory of the camera.

We point out that, besides acceleration, the extension of Sec. 3 is useful for the stereo-processing of high-resolution images, in the limited memory capacity of the GPU. A time-consuming part of such implementations, is the transfer of data from RAM to the GPU's working memory. Given our data-partitioning memory-conservation approach, a depth-first access tree-structure of sectors (see Sec. 3), is proposed. This way the algorithm first completes the processing of an image region until the finest scale and, then, proceeds to the next.

A future research direction concerns larger-scale reconstructions and the case that even the reduced capacity reconstruction, inputted to the method of Sec. 4.2, is too large to be processed by the surface fitting approach. A possible treatment would be partially fit a surface in each box and then adopt the work in [23] to "zip" the partial surface-fittings.

Given the progress in automatic panorama extraction [3], a relaxation of the constraint for sequential ordering of input images in camera motion estimation could be attempted. Regarding memory-constraints in this motion estimation, a way to adopt bundle adjustment for an arbitrary number of frames is pursued. The method in [15], sufficed for the experiments in this paper, but we were unable to apply it in twice the data.

We left out of the scope of this paper the procedure of texture mapping the output mesh surfaces. Currently a scheme of multiple Z-buffers is utilized, to determine the images that a 3D point is visible and assign it with the appropriate color. However, inaccuracies in camera motion estimation and photometric equalization of images give rise to "seams" in the final result and require further research. Fig. 7 shows the visual result from such a coloring or points, from a top view. For reference with the camera motion estimation procedure, marked are also the viewpoints and numbers of some frames. The large white structure on the bottom is the one shown in Fig. 1. In addition, conventional approaches to mapping texture on the triangles of a mesh (e.g, VRML), require that all input textures are present in memory. Since this is difficult for many images a more efficient approach, which possibly utilizes only necessary image segments, is warranted.

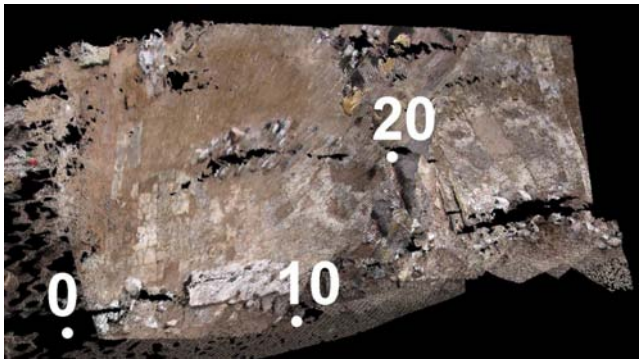

Figure 7. Colorized reconstruction of the scene for the SIFT condition.

References

[1] P. Beardsley, A. Zisserman, and D. Murray. Sequential updating of projective and affine structure from motion. *IJCV*, 23(3):235–259, 1997.

[2] J. L. Bentley. K-D trees for semidynamic point sets. In *Annual symposium on Computational geometry*, pages 187–197, New York, NY, USA, 1990. ACM.

[3] M. Brown and D. G. Lowe. Recognising panoramas. In *ICCV*, 2006.

[4] R. Carceroni and K. Kutulakos. Multi-View scene capture by surfel sampling: From video streams to Non-Rigid 3D motion, shape & reflectance. *IJCV*, 49(2-3):175–214, 2002.

[5] J. C. Carr, R. K. Beatson, J. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3D objects with radial basis functions. In *SIGGRAPH*, pages 67–76, 2001.

[6] J. X. Chai, X. Tog, S. C. Chan, and H. Y. Shum. Plenoptic sampling. In *SIGGRAPH*, pages 307–318, 2000.

[7] R. T. Collins. A space-sweep approach to true multi-image matching. In *IEEE CVPR*, pages 358–363, 1996.

[8] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH*, volume 30, pages 303–312, 1996.

[9] O. Faugeras and R. Keriven. Complete dense stereovision using level set methods. In *ECCV*, pages 379–393, 1998.

[10] D. Gallup, J. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys. Real-time plane-sweeping stereo with multiple sweeping directions. In *CVPR*, pages 1–8, 2007.

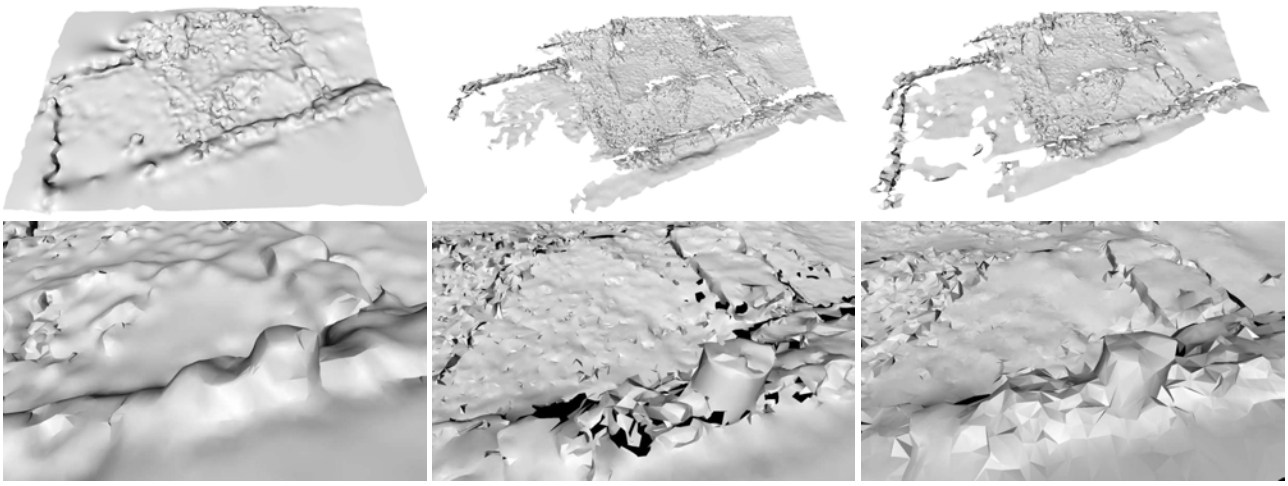

Figure 6. Comparison of surface fitting methods; panoramic view (top) and detail (bottom). Left to right: RBF, Ohtake, proposed extension. In the electronic document, images are zoomable.

[11] I. Geys, T. P. Koninckx, and L. J. Van Gool. Fast interpolated cameras by combining a gpu based plane sweep with a max-flow regularisation algorithm. In *3DPVT*, pages 534–541, 2004.

[12] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *CVPR*, pages 2402–2409, 2006.

[13] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conf.*, pages 147–151, 1988.

[14] M. Li, M. Magnor, and H. P. Seidel. Hardware-accelerated rendering of photo hulls. *Eurographics*, 23(3), 2004.

[15] M. Lourakis and A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Technical Report TF 340, ICS-FORTH, 2004.

[16] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[17] V. Nozick, S. Michelin, and D. Arqus. Image-based rendering using plane-sweeping modelisation. In *IAPR Machine Vision Applications*, pages 468–471, 2005.

[18] Y. Ohtake, A. Belyaev, and H.-P. Seidel. An integrating approach to meshing scattered point data. In *ACM Symposium on Solid and Physical Modeling*, pages 61–69, 2005.

[19] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *IJCV*, 59:207–232, 2004.

[20] M. Pollefeys, D. Nister, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, and H. Towles. Detailed real-time urban 3D reconstruction from video. *IJCV*, 2007.

[21] M. Pollefeys and S. Sinha. Iso-disparity surfaces for general stereo configurations. In *ECCV*, pages 509–520, 2004.

[22] T. Tola. Multi-view 3D reconstruction of a scene containing independently moving objects. Master's thesis, Middle East Technical University, Ankara, Turkey, 2005.

[23] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *SIGGRAPH*, pages 311–318, 1994.

[24] T. Tuytelaars and L. V. Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *BMVC*, pages 421–425, 1998.

[25] R. Yang, G. Welch, and G. Bishop. Real-time consensus-based scene reconstruction using commodity graphics hardware. In *Pacific Graphics*, 2002.

[26] X. Zabulis and K. Daniilidis. Multi-camera reconstruction based on surface normal estimation and best viewpoint selection. In *IEEE International Symposium on 3D Data Processing, Visualization and Transmission*, pages 733–40, 2004.

[27] X. Zabulis and G. Kordelas. Efficient, precise, and accurate utilization of the uniqueness constraint in multi-view stereo. In *IEEE International Symposium on 3D Data Processing, Visualization and Transmission*, 2006.

[28] X. Zabulis, G. Kordelas, K. Mueller, and A. Smolic. Increasing the accuracy of the space-sweeping approach to stereo reconstruction, using spherical backprojection surfaces. In *ICIP*, 2006.

[29] X. Zabulis, A. Patterson, and K. Daniilidis. Digitizing archaeological excavations from multiple views. In *3DIM*, 2005.

[30] Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report 2273, INRIA, 1994.