
Utilization of the texture uniqueness cue in stereo

Xenophon Zabulis¹

Informatics and Telematics Institute,
Centre for Research and Technology Hellas
xenophon@iti.gr

1 Introduction

The cue to depth due to the assumption of texture uniqueness has been widely utilized in approaches to shape-from-stereo. Despite the recent growth of methods that utilize spectral information (color) or silhouettes to three-dimensionally reconstruct surfaces from images, the depth cue due to the texture uniqueness constraint remains relevant, as being utilized by a significant number of contemporary stereo systems [1, 2]. Certainly, combination with other cues is necessary for maximizing the quality of the reconstruction, since they provide of additional information and since the texture-uniqueness cue exhibits well-known weaknesses; e.g. at cases where texture is absent or at the so-called “depth discontinuities”. The goal of this work is to provide of a prolific, in terms of accuracy, precision and efficiency, approach to the utilization of the texture uniqueness constraint which can be, thereafter, combined with other cues to depth.

The uniqueness constraint assumes that a given pixel from one image can match to no more than one pixel from the other image [3, 4]. In stereo methods, the uniqueness constraint is extended to assume that, locally, each location on a surface is uniquely textured. The main advantages of the cue derived from the uniqueness constraint over other cues to depth are the following. It is independent from silhouette-extraction, which requires an accurate segmentation (e.g. [5]). It is also independent of any assumption requiring that cameras occur around the scene (e.g. [6]) or on the same baseline (e.g. [7, 8]). Moreover, it does not require that cameras are spectrally calibrated, such as in voxel carving / coloring approaches (e.g. [9, 10, 11]). The locality of the cue due to the uniqueness constraint facilitates multi-view and parallel implementations, for real-time applications [12, 13, 14].

Traditionally, stereo-correspondences have been established through a similarity search, which matched image neighborhoods based on their visual similarity [15, 16]. After the work in [17], volumetric approaches have emerged that establish correspondences among the acquired images after

backprojecting them onto a hypothetical surface. *Space-sweeping* methods [17, 18, 19, 20, 21, 22, 23], backproject the images onto a planar surface, which is “swept” along depth to evaluate different disparities. *Orientation-optimizing* methods [24, 25, 26, 27, 28], compare the backprojections onto hypothetical surface segments (surfels), which are evaluated at range of potential locations and orientations in order to find the location and orientation at which the evaluated backprojections match the best. The relation with the traditional, neighborhood-matching, way of establishing correspondences is the following. In volumetric methods, a match is detected as a location at which the similarity of backprojections is locally maximized. This location is considered as the 3D position of the surface point and the image points that it projects at as corresponding.

Orientation-optimizing methods compensate for the projective distortion in the similarity-matching process and have been reported to be of superior accuracy to window-matching and space-sweeping approaches [29, 1]. The reason is that the matching process is more robust when the compared textures are relieved from - the different for each camera - projective distortion. On the other hand, their computational cost is increased as many times as the number of evaluated orientations of the hypothetical surface segment.

The remainder of this chapter is organized as follows. In Section 2, relevant work is reviewed. In Section 3, the uniqueness cue and relevant theory are defined in the context of volumetric approaches and an accuracy-increasing extension to this theory is proposed. In Section 4, computational optimizations that accelerate the above method and increase its precision are proposed. In addition, the theoretical findings of Section 3 are employed to define a space-sweeping approach of enhanced accuracy, which is then combined with orientation-optimizing methods into a hybrid method. Finally in Section 5, conclusions are drawn and the utilization of the contributions of this work by stereo methods is discussed.

2 Related work

Both traditional and volumetric techniques optimize a similarity criterion along a spatial direction in the world or in the image, to determine the disparity or position of points on the imaged surfaces. In traditional stereo (e.g. [15, 16]) or space-sweeping [17, 18, 19, 20, 21, 22, 23], a single orientation is considered, typically the one frontoparallel to the cameras. Orientation-optimizing techniques [25, 26, 28, 24, 27] consider multiple orientations at an additional computational cost, but also provide an estimation of the normal of the imaged surface. As benchmarked in the literature [1, 2] and explained in Section 3.2, the accuracy of sweeping methods is limited in the presence of surface slant, compared to methods that account for surface orientation.

Although that space-sweeping approaches produce results of similar accuracy to the, traditional, neighborhood-matching algorithms [29, 1] they exhibit

decreased computational cost. Furthermore, the time-efficiency of sweeping methods is reinforced when implemented to execute in commodity graphics hardware [30, 31, 23]. Due to its “single-instruction multiple-data” architecture, graphics hardware executes the, essential for the space-sweeping approach, warping and convolution operations in parallel.

Regarding the shape and the size of sweeping surfaces, it has been shown [32] that projectively expanding this surface (as in [33, 32, 31, 30, 34]) exploits better the available pixel resolution, than implementing the sweep as a simple translation of the sweeping plane [17, 18, 19, 20, 21, 22, 23]. In this context, a more accurate space-sweeping method is proposed in Section 4.4. In orientation-optimizing approaches, the size of the hypothetical surface patches has been formulated as constant [25, 26, 28, 24, 27]. Predefined sets of sizes have been utilized in [28] and [24], in a coarse-to-fine acceleration scheme. However, the evaluated sizes were the same for any location and orientation of the patch, rather than modulated as their function.

Metrics for evaluating the similarity of two image regions (or backprojections) fall under two main categories. Photoconsistency [20, 35] and texture similarity [36, 21, 22, 23, 37], which is typically implemented using the SAD, SSD, NCC, or MNCC [38] operators. The difficulty using the photoconsistency metric is that radiometric calibration of the cameras is required, which is difficult to achieve and retain. In contrast, the NCC and MNCC metrics are not sensitive to the absence of radiometric calibration, since they compare the correlation of intensity values rather than their differences [39]. Finally, some sweeping-based stereo reconstruction algorithms match sparse image features [17, 18, 19] but are, thus, incapable of producing dense depth maps.

Global optimization approaches have also utilized the uniqueness approach [26, 40, 7, 41, 42, 43], but can yield local minima of the overall cost function and are much more difficult to parallelize than local volumetric approaches. As in local methods, the similarity operator is either an oriented backprojection surface segment (e.g. [26]) or an image neighborhood (e.g. [7]). Thus, regardless of how the readings of this operator are utilized by the reconstruction algorithm, the proposed enhancement of the hypothetical surface patch operator should only improve the accuracy of these approaches.

Finally, the assumption of surface continuity [3, 4] has been utilized for resolving ambiguities as well as correcting inaccuracies (e.g. [7, 41]). In traditional stereo, some approaches to enforce this constraint are to filter the disparity map [13], bias disparity values to be in coherence with neighboring [7], or require *inter-scanline* consistency [42]. The continuity assumption has been also utilized in 2D [44], but seeking continuity in the image intensity domain. The assumption has also been enforced to improve the quality of reconstruction, in post-processing; an abundance of approaches for 3D filtering of the results exists in the deformable models literature (see [45] for a review).

3 The texture uniqueness cue in 3D

In this section, the texture uniqueness cue is formulated volumetrically, or in 3D, and it is shown that this formulation can lead to more accurate reconstruction methods than the, traditional, 2D formulation. Next, the spatial extent over which textures are matched is considered and an accuracy-increasing extension to orientation-optimizing approaches is proposed. It is noted that, henceforth, it is assumed that images portray Lambertian surfaces, which can also be locally approximated by planar patches. Extension of these concepts beyond the Lambertian domain can be found in [46].

3.1 Uniqueness cue formulation

Let a calibrated image pair $I_{i=1,2}$, acquired from two cameras with centers $\mathbf{o}_{1,2}$ and principal axes $\mathbf{e}_{1,2}$; cyclopean eye is at $\mathbf{o} = \frac{\mathbf{o}_1 + \mathbf{o}_2}{2}$ and mean optical axis is $\mathbf{e} = \frac{\mathbf{e}_1 + \mathbf{e}_2}{2}$. Let also a planar surface patch \mathcal{S} , of size $\alpha \times \alpha$, centered at \mathbf{p} , with unit normal \mathbf{n} . Backprojecting I_i onto \mathcal{S} yields image $w_i(\mathbf{p}, \mathbf{n})$:

$$w_i(\mathbf{p}, \mathbf{n}) = I_i \left(P_i \cdot \left(\mathbf{p} + R(\mathbf{n}) \cdot [x' \ y' \ 0]^T \right) \right), \quad (1)$$

where P_i is the projection matrix of I_i , $R(\mathbf{n})$ is a rotation matrix so that $R(\mathbf{n}) \cdot [0 \ 0 \ 1]^T = \mathbf{n}$ and $x', y' \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$ local coordinates on \mathcal{S} .

When \mathcal{S} is tangent at a world surface, w_i are identities of the surface pattern (see Figure 1 left). Thus $I_1(P_1\mathbf{x}) = I_2(P_2\mathbf{x})$, $\forall \mathbf{x} \in \mathcal{S}$, and therefore their similarity is optimal. Otherwise w_i are dissimilar, because they are collineations from different surface regions. Assuming a voxel tessellation of space, the locations of surface points and corresponding normals can be recovered by estimating the positions at which similarity is locally maximized along the direction of the surface normal, or otherwise, exhibit a higher similarity value than their (two) neighbors in that direction. Such a location will be henceforth referred to as a *similarity local maximum*, or simply, local maximum. To localize the similarity local maxima, function $\mathbf{V}(\mathbf{p}) = s(\mathbf{p})\boldsymbol{\kappa}(\mathbf{p})$, is evaluated as:

$$s(\mathbf{p}) = \max_{\mathbf{n}} (\text{sim}(w_1(\mathbf{p}, \mathbf{n}), w_2(\mathbf{p}, \mathbf{n}))), \quad (2)$$

$$\boldsymbol{\kappa}(\mathbf{p}) = \arg \max_{\mathbf{n}} (s(\mathbf{p})). \quad (3)$$

where $s(\mathbf{p})$ the optimal correlation value at \mathbf{p} , and $\boldsymbol{\kappa}(\mathbf{p})$ the optimizing orientation. The best matching backprojections are $w'_{1,2} = w_{1,2}(\mathbf{p}, \boldsymbol{\kappa})$. Metric *sim* can be SAD, SSD, NCC, MNCC etc. To evaluate *sim*, a $r \times r$ lattice of points is assumed on \mathcal{S} . In addition, a threshold τ_c is imposed on s so that local maxima of small similarity value are not interpreted as surface occurrences.

The parameterization of \mathbf{n} requires two dimensions and can be expressed in terms of longitude and latitude, which define any orientation within a unit hemisphere. To treat equally different eccentricities of \mathcal{S} the orientation, $\mathbf{c} = [x_c \ y_c \ z_c]^T$, that corresponds to the pole of this hemisphere points to

\mathbf{o} ; that is $\mathbf{c} = \mathbf{p} - \mathbf{o}$ (see Figure 1 right). The parameterized orientations $\mathbf{n} = [x_i, y_i, z_i]^T$ are then:

$$\begin{aligned} x_i &= \frac{z_c \cdot x_c \cdot \cos \omega \cdot \sin \psi - y_c \cdot N_1 \cdot \sin \omega \cdot \sin \psi + x_c \cdot N_2 \cdot \cos \psi}{N_1 \cdot N_2} \\ y_i &= \frac{z_c \cdot y_c \cdot \cos \omega \cdot \sin \psi + x_c \cdot N_1 \cdot \sin \omega \cdot \sin \psi + y_c \cdot N_2 \cdot \cos \psi}{N_1 \cdot N_2} \\ z_i &= \frac{z_c \cdot \cos \psi - N_2 \cdot \cos \omega \cdot \sin \psi}{N_1}, \end{aligned} \quad (4)$$

where $N_1 = \sqrt{x_c^2 + y_c^2 + z_c^2}$, $N_2 = \sqrt{x_c^2 + y_c^2}$, $\omega \in [0, 2\pi)$, $\psi \in [0, \frac{\pi}{2})$. The corresponding rotation that maps $[0 0 1]^T$ to the particular orientation is $R = R_x \cdot R_y$, where R_x is the rotation matrix for a $\cos^{-1} z_k$ rotation about the xx' axis. If $x_k \neq 0$, then R_y is the rotation matrix for a $\tan^{-1}(\frac{y_k}{x_k})$ rotation about the yy' axis or, otherwise, R_y is the 3×3 identity matrix. The computational cost of the optimization for a voxel is $\mathcal{O}(Nr^2)$, where N is the number of orientations evaluated by \mathbf{n} .

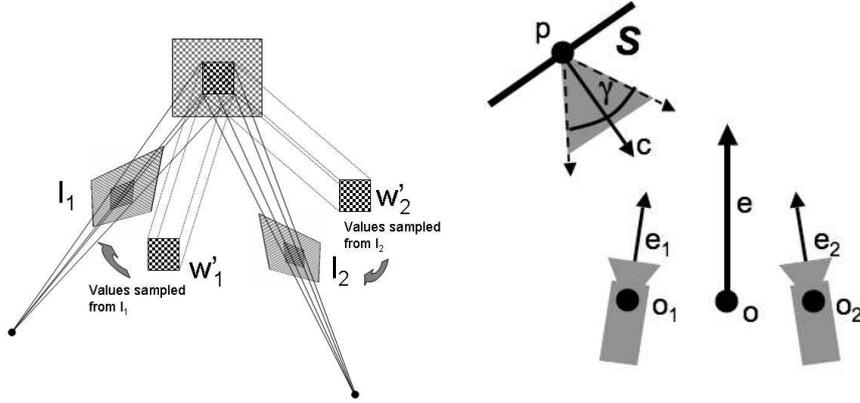


Fig. 1. *Left:* A surface is projectively distorted in images $I_{1,2}$, but the collineations $w'_{1,2}$ from a planar patch tangent to this surface are not. *Right:* Illustration of the geometry for Equation 4.

A reconstruction of the locations and corresponding normals of the imaged surfaces can be obtained by detecting the similarity local maxima that are due to the occurrence of a surface. These maxima can be detected as the positions where s is maximized along the direction of the surface normal [28]. An estimation of the surface normal is provided by κ since, according to Equation 3, κ should coincide with the surface normal. A suitable algorithmic approach to the computation of the above location is given by a 3D

version of the Canny edge detector [47]. In this version, the gradient is also 3-dimensional, its magnitude is given by $s(\mathbf{p})$ and its direction by $\kappa(\mathbf{p})$. The non-maxima suppression step of Canny’s algorithm performs, in essence, the detection of similarity local maxima since it rejects all voxels in \mathbf{V} that are *not* local maxima along the surface normal. A robust implementation of the above approach is achieved following the work in [48], but substituting the 3D gradient with \mathbf{V} .

3.2 Search direction and accuracy of reconstruction

Both optimizing for the orientation of the hypothesized patch \mathcal{S} and detecting similarity maxima in the direction of the surface normal increase the accuracy of final reconstruction. The claimed increase in accuracy for the uniqueness cue is theoretically expected due to the following proposition: the spatial error of surface reconstruction is a monotonically increasing function of the angle between the normal vector of the imaged surface and the spatial direction over which a similarity measure is optimized (proof in Appendix A). This proposition constitutes of a mathematical explanation of why space-sweeping approaches are less accurate than orientation optimizing methods. Intuitively, the inaccuracy is due to the fact that the backprojections on \mathcal{S} , which is oriented differently than the imaged surface, do not correspond to the same world points - except for the central point of \mathcal{S} .

The above proposition also explains why similarity local maxima are optimally recovered when backprojections are evaluated tangentially to the surface to be reconstructed. In addition to the above, when the search direction for local maxima is in wide disagreement with the surface normal, inhibition of valid maxima occurs, deteriorating the quality of reconstruction even further. The reason is that an inaccurate search direction may point to and, thus, suppress validly occupied neighboring voxels. When κ is more accurate, this suppression attenuates because κ points perpendicularly to the surface. The following experiments confirm that detecting similarity local maxima along κ and optimizing \mathbf{n} provides of a more fidelious reconstruction.

The first experiment utilizes computer simulation to show that this improvement of accuracy occurs even in synthetic images, where noise and calibration errors are absent. In the experiment, simulated was a binocular camera pair that imaged obliquely a planar surface. A planar patch \mathcal{S} , oriented so that its normal was equal to \mathbf{e} , was swept along depth. At each depth, the locations of the surface points that were imaged on \mathcal{S} , through the backprojection process, were calculated for each camera. Thus at a given depth, a point on \mathcal{S} indicates a pair of world points occurring someplace on the imaged surface. For each depth, the distances of such pairs of world points were summed. In Figure 2(a-c), the setup as well as the initial, middle and final position of the patch are shown. Figure 2(d) shows the sum of distances obtained for each depth value, for a $r \times r$, $r = 11$ grid on \mathcal{S} . According to the prediction, the minimum of this summation function does *not* occur at $\delta = 0$, which is the

correct depth. The dislocation of this minimum is the predicted depth error for this setup. The experiment shows that even in ideal imaging conditions, space-sweeping is *guaranteed* to yield some error if the scene includes any significant amount of slant.

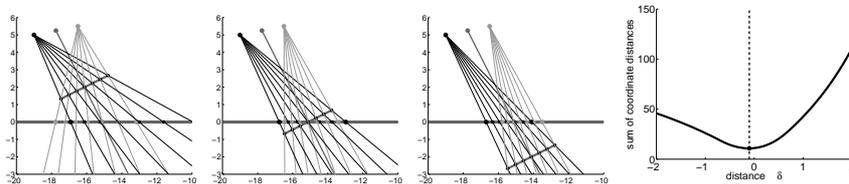


Fig. 2. Left to right: (a) initial, (b) middle and (c) final position of a hypothetical patch (magenta). Line pencils show the lines of sight from the optical centers to the imaged surface, through the patch. The middle line plots the direction of \mathbf{e} . The plot on the right (d), shows the sum of distances of points imaged through the same point on the patch as a function of δ ; it indicates that the maximum of similarity is obtained at $\delta \neq 0$ (< 0 , in this case).

In the second experiment (see Figure 3), a well-textured planar surface was reconstructed considering an \mathcal{S} which assumed: either solely the frontoparallel orientation - as in space-sweeping, or a set of orientations within a cone of opening γ around \mathbf{e} . Judging by the planarity of the reconstructed surface, the least accurate reconstruction was obtained by space-sweeping. Notice that in (c), due to the compensation for the projective distortion, backprojections $w'_{1,2}$ were more similar than in (b). As a result, higher similarity values were obtained and, thus, more local maxima exhibited a similarity value higher than threshold τ_c . Figure 3(d) is discussed in Section 3.3.

3.3 Optimizing accuracy in discrete images

In this subsection, the size, α , of \mathcal{S} and the corresponding image areas where \mathcal{S} projects are studied. A modulation of α is proposed to increase the accuracy of the patch operator, as it has been to date formulated [24, 25, 26, 28]. Finally, integration with the surface continuity assumption is demonstrated to alleviate the result from residual inaccuracies.

In discrete images, the number of subtended pixels by the projection of \mathcal{S} is analogous to the reciprocals of distance squared and the cosine of the relative obliqueness of \mathcal{S} to the cameras. Thus in Equation 3, when α is constant the greater the obliqueness the fewer the image pixels that the $r \times r$ image samples are obtained from. Therefore, there will always be some obliqueness after which the same intensity values will start to be multiply sampled. In this case, as obliqueness and/or distance increase the population of these intensities will tend to exhibit reduced variance, because they are being sampled from

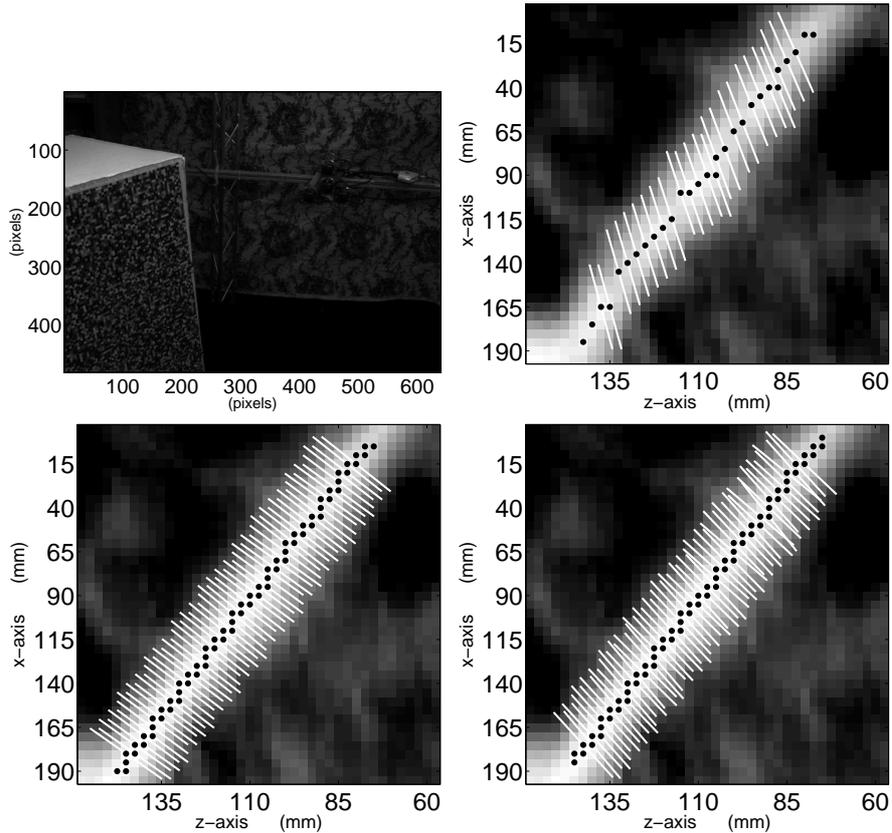


Fig. 3. Comparison of methods for similarity local maxima detection. Clockwise from top left: (a) Image from a horizontally arranged binocular pair (baseline 156mm), showing a XZ section in space at which \mathbf{V} was calculated three-ways: (b) plane-sweeping, (c) optimizing \mathbf{n} , and (d) updated assuming surface continuity (see Section 3.3). In (a), checker size was 1cm^2 and target was $\approx 1.5\text{m}$ from the cameras. In the maps (b-d), dark dots are local maxima, white lines are κ , voxel= 125mm^3 , $r = 21 \times 21$, $\alpha = 20\text{mm}$. In (c), $\gamma = 60^\circ$ and \mathbf{n} 's orientations were parameterized every 1° . Notice that the last two methods detect more local maxima, although that the same similarity threshold τ_c was used in all three conditions.

decreasingly fewer pixels. Thus, a bias in favor of greater slants and distances is predicted. Mathematically, because variance occurs in the denominator of the correlation function. Intuitively, because fewer image area supports now the similarity matching of backprojections on \mathcal{S} , and as a consequence, this matching becomes less robust.

To observe the predicted phenomenon, surface orientation was estimated and compared to ground truth. Experiments were conducted with both real

and synthetic images, to stress the point that the discussed inaccuracy cannot be attributed to noise or calibration errors and that, therefore, it must be contained in the information loss due to image discretization.

In the first experiment (see Figure 4), a binocular image pair was synthesized to portray a square, textured and planar piece of surface. Equations 2 and 3 were then evaluated for the central point on the surface. The similarity values obtained for each orientation of \mathbf{n} were arranged to form a longitude - latitude map, which can be read as follows. The longitude and latitude axes correspond to the dimensions defined by, respectively, modulating ψ and ω in Equation 4; coordinates $(0,0)$ correspond to \mathbf{c} . In the map, lighter pixels indicate a high similarity value and darker the opposite (henceforth, this convention is followed for all the similarity maps in this chapter). Due to the synthetic nature of the images, which facilitated a perfect calibration, a small amount of the predicted inaccuracy was observed.

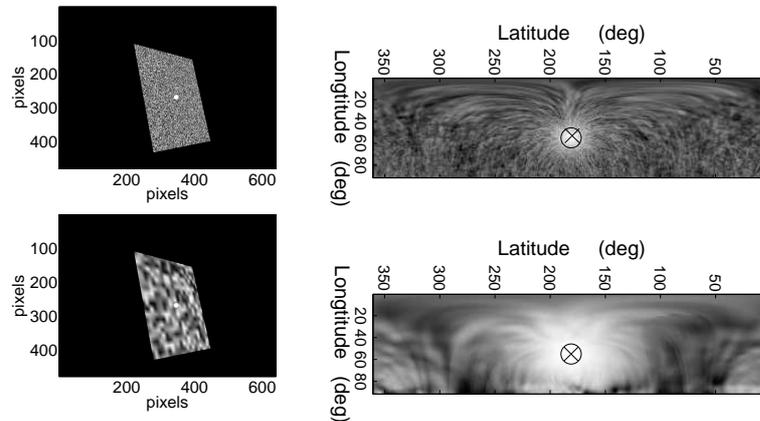


Fig. 4. Two textures rendered on a $251 \times 251 \text{ mm}$ planar surface patch (left) and the corresponding similarity values obtained by rotating a \mathcal{S} concentric with the patch (right). In the maps, camera pose is at $(0,0)$, crosses mark the maximal similarity value and circles mark ground truth. In the experiment, $\alpha = 100 \text{ mm}$, $r = 15$, $\gamma = 90^\circ$. The binocular pair was $\approx 1.5 \text{ m}$ from the patch and its baseline was 156 mm . The, angular, parameterization of \mathbf{n} was in steps of $.5^\circ$. The errors for the two conditions, measured as the angle between the ground truth normal and the estimated one, were and 2.218° (top) and 0.069° (bottom).

In the second experiment, calibration errors give rise to even more misleading local maxima and, also, the similarity value at very oblique orientations of \mathbf{n} ($> 60^\circ$) is observed to reach extreme positive or negative values. In Figure 5,

to indicate the rise of the spurious maximum at the extremes of the correlation map, the optimization was twice computed: once for $\gamma = 120^\circ$ and once for $\gamma = 180^\circ$. In both cases, the global maximum occurred at the extreme border of this map, thus corresponding to a more oblique surface normal - relative to the camera - than ground truth.

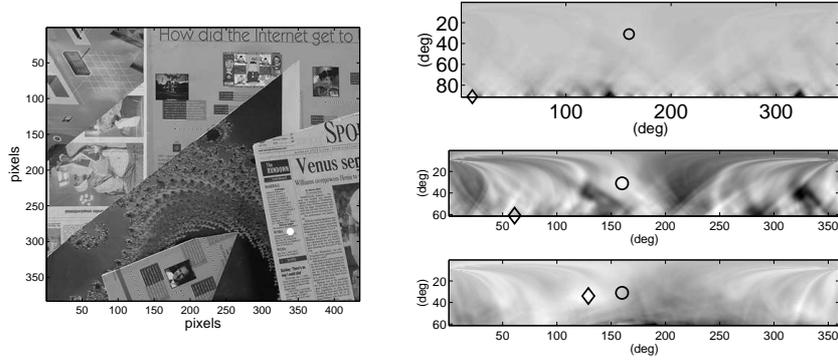


Fig. 5. Comparison of techniques. Repeating the experiment of Figure 4 for the first two frames of the “Venus” Middlebury Sequence and for two different γ s: 120° (top map) and 180° (middle map). In the experiment, $\alpha = 250$ length units, baseline was 100 length units and $r = 151$. The surface point for which Equations 2 and 3 were evaluated is marked with a circle (left). The projection of \mathcal{S} subtended an area of ≈ 50 pixels. The bottom map shows the increase in accuracy obtained by the size-modulation of \mathcal{S} with respect to obliqueness (see forward in text). Mapping of similarity values to intensities is individual for each map.

The above phenomenon can be alleviated if the size of the backprojection surface \mathcal{S} is modulated so that its image area remains invariant. In particular, the side of \mathcal{S} (or diameter, for a circular \mathcal{S}) is modulated as:

$$\alpha = \frac{\alpha_0 \cdot d}{d_0 \cdot \cos \omega}; \quad \omega = \cos^{-1} \left(\frac{\mathbf{v} \cdot \mathbf{n}}{|\mathbf{v}| \cdot |\mathbf{n}|} \right), \quad (5)$$

where $\mathbf{v} = \mathbf{p} - \mathbf{o}$, $d = |\mathbf{v}|$, ω is the angle between \mathbf{v} and \mathbf{n} and d_0 , α_0 initial parameters. Notice that even for a single location, size is still varied as \mathbf{n} is varied. Figures 5 and 6 show the angular and spatial improvement in accuracy induced by the proposed size-modulation. They compare the reconstructions obtained with patch whose size was modulated as above against those obtained with a constant-sized \mathcal{S} - as to date practiced in [24, 25, 26, 28]. A “side-effect” of the above modulation is that the larger the distance and the obliqueness of a surface, the lower the spatial frequency that is reconstructed at. This effect is considered as a natural tradeoff since distant and oblique surfaces are also imaged at lower frequencies.

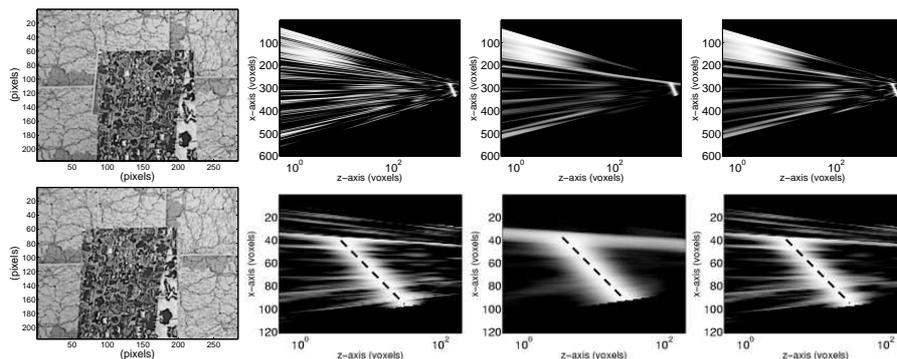


Fig. 6. Comparison of techniques. Shown is a stereo pair (left column) and three separate calculations of s across a vertical section, through the middle of the foreground surface. The bottom figures are zoom-in detail on the part that corresponds to the foreground and z -axes (horizontal in maps) are logarithmic. In the bottom figures, ground truth is marked with a dashed line. In the 2^{nd} column, a fine α was used, hence the noisy response at the background. Using a coarse α (3^{rd} column), yields a smoother response in greater distances, but diminishes any detail that could be observed in short range. In the 4^{th} column, α is projectively increased, thus, normalizing the precision of reconstruction by the area that a pixel images at that distance.

Assuming surface continuity has been shown to reduce inaccuracies due to noise or lack of resolution in a wide variety of methods and especially in global-optimization approaches (see Section 2). To demonstrate the compatibility of the proposed operator with these approaches and suppress residual inaccuracies, the proposed operator is implemented with feedback obtained from a surface-smoothing process. Once local maxima have been detected in \mathbf{V} , the computed κ s are updated as follows. For voxels where a similarity local maximum occurs, κ is replaced by the normal of the least-squares fitting plane through the neighboring occupied voxels. For an empty voxel, \mathbf{p}_e , the updated value of κ is $\sum_j \beta_i \kappa(\mathbf{p}_j) / \sum \beta_j$, where j enumerates the occupied voxels within \mathbf{p}_e 's neighborhood. After the update, local maxima are re-detected in the new \mathbf{V} . The results are more accurate, because similarity local maxima are detected along a more accurate estimation of the normal. Note that if \mathbf{n} 's optimization is avoided, the initial local maxima are less accurately localized and so are the updated κ s.

4 Increasing performance

Two techniques are proposed for increasing the performance of the proposed implementation of the uniqueness cue. The first aims the provision of high-precision results and the second the aims the reduction of its computational cost. In addition, a hybrid approach is proposed that combines

the rapid execution of space-sweeping with the increased accuracy of the proposed orientation-optimizing method. To enhance the accuracy of the space-sweeping part of the proposed approach, an enhanced version of space-sweeping is introduced that is based on the conclusions of Section 3.2.

4.1 Precision

In volumetric methods, the required memory and computation increase by a cubic factor as voxel size decreases and, thus, computational requirements are quite demanding in reconstructions of high-precision. The proposed technique refines the initial voxel-parameterized reconstruction to sub-voxel precision, given \mathbf{V} and the detected similarity local maxima as input. The local maxima are in voxel-parameterization and treated as a coarse approximation of the imaged surface. The method densely interpolates through the detected similarity local maxima. This interpolation is guided by \mathbf{V} , in order for the result to pass through the locations where similarity (s , or $|V|$) is locally maximized.

To formulate the interpolation, S_f henceforth refers to the 0 -*isosurface* of $\mathcal{G} = |\nabla \mathbf{V}|$, or otherwise to the set of locations at which $\mathcal{G} = 0$. At the vicinity of the detected local maxima, similarity is locally maximized and, thus, the derivative’s norm (\mathcal{G}) should be 0. The result is defined as the localization of S_f at the corresponding regions. Guiding the interpolation with the locations of S_f , utilizes the obtained similarity values to accurately increase the precision of the reconstruction and not, blindly, interpolate through the detected similarity local maxima.

The interpolation utilizes the Radial Basis Function (RBF) framework in [49] to approximate the isosurface. This framework requires pivots which will guide the interpolation, and which in the present case are derived by the detected local maxima. For each one of them, the values of \mathcal{G} at the locations $\mathbf{p}_{1,2} = \mathbf{p}_m \pm \lambda \boldsymbol{\kappa}$, where \mathbf{p}_m is the position of the local maxima, are estimated by trilinear interpolation. The pivots are assigned with values $\xi_{1,2} = c \cdot \mathcal{G}(\mathbf{p}_{1,2})$, where c is -1 for the closer of the two pivot points to the camera and 1 for the other. Values $\xi_{1,2}$ are of the opposite sign, to constrain the 0 -isosurface to occur in between them. The value of λ is chosen to be less than voxel size (i.e. 0.9) to avoid interference with local maxima occurring at neighboring voxels [50]. Function \mathcal{G} is approximated in high-resolution by the RBF framework and the isosurface is extracted by the Marching Cubes algorithm [51]. The result is represented as a mesh of triangles. The proposed approach is, in essence, a search for the zero-crossings of \mathcal{G} .

The computational cost of the above process is much less than the cost of evaluating \mathbf{V} at the precision that is interpolated. However, it is still a computationally demanding process of complexity $O(N^3)$, where N is the number of data points. Even though the optimization in [50], which reduces complexity to $O(N \log N)$, was adopted the number of data points in wide-area reconstructions can be quite large to obtain results in real-time. To, at least, parallelize the process, the reconstruction volume can be tessellated in

overlapping cubes and the RBF can be independently computed at each. No significant differences in the reconstruction were observed when fitting the RBF directly to the whole reconstruction and in the individual cubes, due the overlap of cubes. The partial meshes are finally merged as in [52].

4.2 Acceleration

Two hierarchical, coarse-to-fine iterative methods are proposed for the acceleration of the search for similarity local maxima.

The first, is an iterative coarse-to-fine search that reduces the number of evaluated \mathbf{n} 's in Equation 3. In this formula, the exhaustive search computes s for every \mathbf{n} within a cone of opening γ . At each iteration i : (a) the cone is canonically sampled and the optimizing direction $\boldsymbol{\kappa}_i$ is selected amongst the sampled directions, (b) the sampling gets exponentially denser, but (c) only the samples within the opening of an exponentially narrower cone around $\boldsymbol{\kappa}_{i-1}$ are evaluated. At each iteration, the opening γ_i of the cone is reduced as $\gamma_{i+1} = \gamma_i/\delta$, $\delta > 1$ (in our experiments $\delta = 2$). Iterations begin with $\boldsymbol{\kappa}_1 = \mathbf{c}$ and end when γ_i falls below a precision threshold τ_γ . For a voxel at \mathbf{p} , the parameterized normals are given by Equation 4, by modulating ψ to be in $[0, \gamma_i]$ and setting $\mathbf{c} = \mathbf{p} - \mathbf{o}$. In Figure 7, the accuracy of the proposed method is shown as a function of computational effort. As ground truth, the result of the exhaustive search was considered, which required 10800 invocations of the similarity function. It can be seen that after 3 iterations, which correspond to a speedup > 7 , the obtained surface normal estimation is inaccurate by less than 3° . After 7 iterations accuracy tends to be less than 1.5° (speedup ≈ 2). Given the correction of the surface normal in Section 3.3, the residual minute inaccuracies may be neglected without consequences for the quality of the reconstruction and the process is stopped at the 3^{rd} iteration. Also in practice, a speedup of ≈ 20 is obtained, since in our implementation only the 1^{st} iteration is performed if all samples are less than threshold.

The second method reduces the number of evaluated voxels, by iteratively focusing computational effort at the volume neighborhoods of similarity local maxima. It is based on a scale-space treatment of the input images. At each iteration, $\alpha_i = \alpha_0/2^i$ and $I_{1,2}$ are convolved with a Gaussian of $\sigma_i = \sigma_0/2^i$. Also, voxel volume is reduced by $1/2^3$ and correlation is computed only at the neighborhoods of the local maxima detected in the previous iteration. The effect of these modulations is that at initial scales correspondences are evaluated for coarse-scale texture and at finer scales utilize more image detail. Their purpose is to efficiently compare $w_{1,2}$ at coarse scales. At these scales, the projections of the points on \mathcal{S} in the image are sparse and, thus, even a minute calibration error causes significant miscorrespondence of their projections. Smoothing, in effect, decreases image resolution and, thus, more correspondences are established at coarse scales. In Figure 8, the method is demonstrated. No errors in the first iteration that led to a void in the final reconstruction were observed, utilizing 3 iterations of the above algorithm -

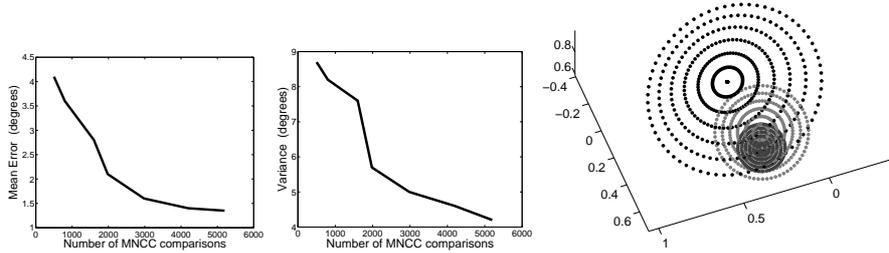


Fig. 7. Mean (left) and variance (middle) of the error of angular optimization as a function of the computational effort, measured in similarity metric invocations. In the experiment, the results of 10^3 estimations were averaged $\gamma_1 = 60^\circ$, $\delta = 2$ and $\tau_\gamma = 1^\circ$. The right plot illustrates the hierarchical evolution of considered orientations, as points on the unit sphere.

but of course this tolerance is a function of the available image resolution. In our experiments a speedup of ≈ 5 was observed on average.

4.3 Sphere-sweeping

In this subsection, the geometry of space-sweeping is revisited and a spherical parameterization of the sweeping surface is proposed and evaluated. The proposed approach substitutes the backprojection plane, in space-sweeping, with a spherical sector that projectively expands from the cyclopean eye outwards. Using this backprojection surface, a visibility ray \mathbf{v} departing from the optical center is always perpendicular to the backprojection surface for any eccentricity ϵ of in the field of view (FOV) (see Figure 9 left). Thus, the number of sampled image pixels per unit area of backprojection surface is maximized. In contrast, a frontoparallel planar surface is imaged increasingly slanted relatively to \mathbf{v} as ϵ moves to the periphery of the image and, therefore, a smaller accuracy is expected, based on the conclusion of Section 3.3 (see [32] for a proof).

The method is formulated as follows. Let a series of concentric and expanding spherical sectors S_i at corresponding distances d_i from the cyclopean eye (C). Their openings μ , λ in the horizontal and vertical direction, respectively, are matched to the horizontal and vertical FOVs of the cameras and tessellated by an angular step of c . Parameterization variables ψ and ω are determined as $\psi \in \{c \cdot i - \mu; i = 0, 1, 2, \dots, 2\mu/c\}$ and $\omega \in \{c \cdot j - \lambda; j = 0, 1, 2, \dots, 2\lambda/c\}$ and $\lceil \mu/c \rceil = \mu/c, \lceil \lambda/c \rceil = \lambda/c$. For both ψ and ω , value 0 corresponds to the orientation of the mean optical axis \mathbf{e} . To generate sectors S_i , a corresponding sector S_0 is first defined on a unit sphere centered at $O = [0\ 0\ 0]^T$. A point $p = [x\ y\ z]^T$ on S_0 is given by: $x = \sin(\psi)$, $y = \cos(\psi)\sin(\omega)$, $z = \cos(\psi)\cos(\omega)$. Its corresponding point p' on S_i is then:

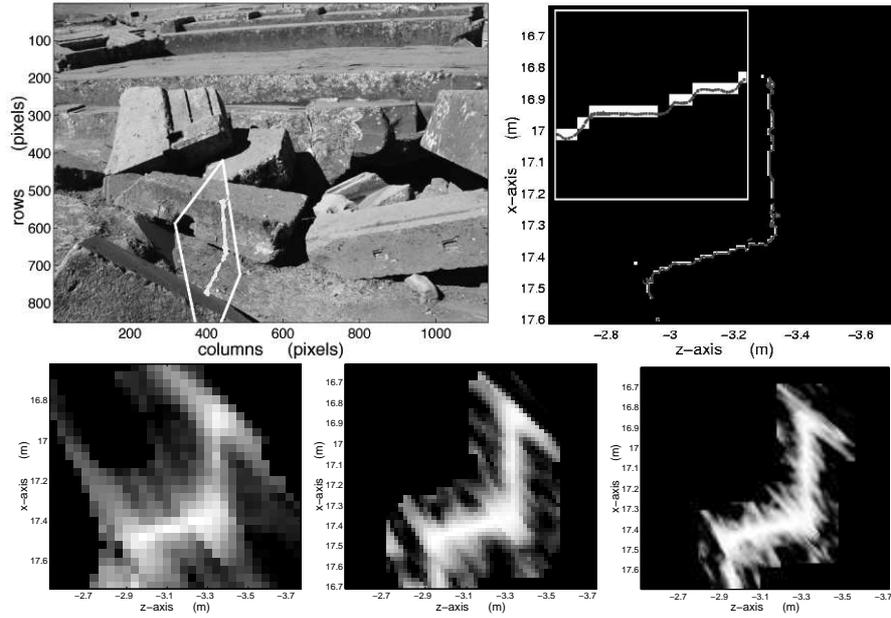


Fig. 8. Coarse-to-fine localization of similarity local maxima. A digital photograph from a $\approx 40\text{ cm}$ -baseline binocular pair and, superimposed, a section in space that perpendicularly intersects the imaged piece of rock. The 2D map on the right shows the result of local maxima detection across this section. Marked in white are the detected local maxima at voxel precision and with gray the result of their subvoxel approximation. These local maxima were then reprojected to the original image and marked (left). At the bottom, the three maps show the result of the coarse to fine computation of \mathbf{V} , for the same section in space. In the experiment, $\alpha_0 = 8\text{ cm}$, voxel = $(4\text{ cm})^3$, $r = 21$, $\sigma_0 = 5$.

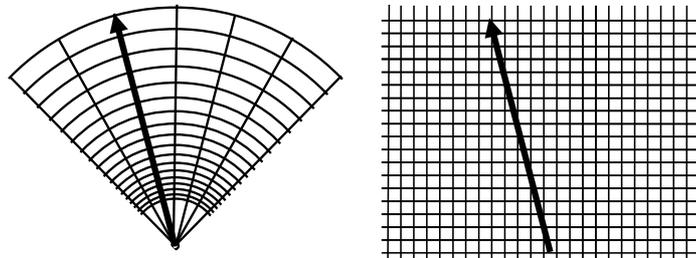


Fig. 9. Illustration of the sector (left) and voxel (right) based volume tessellations. Visibility is naturally expressed in the first representation, whereas in the second, traversing voxels diagonally is required.

$$p' = d_i [R_z(-\theta)R_y(-\phi)p + C], \quad (6)$$

where R_y and R_z are rotation matrices for rotations about the yy' and zz' axes, $\mathbf{v}_{1,2}$ are unit vectors on the principal axes of the cameras, $\mathbf{v} = (\mathbf{v}_1 + \mathbf{v}_2)/2$, and θ (longitude), ϕ (colatitude) \mathbf{v} 's spherical coordinates. Computational power is conserved, without reducing the granularity of the reconstructed depths, when parameterizing d_i on a disparity basis [53]: $d_i = d_0 + \beta^i$, $i = 1, 2, \dots, i_N$, where d_0 and i_N define the sweeping interval and β is modulated so that the farthest distance is imaged in sufficient resolution.

The rest of the sweeping procedure is similar to plane-sweeping and, thus, overviewed. For each S_i , the stereo images (≥ 2) are sampled at the projections S_i 's points on the acquired images, thus forming two $(2\mu/c \times 2\lambda/c)$ backprojection images. Backprojecting and locally comparing images is straightforwardly optimized by a GPU as a combination of image difference and convolution operations (e.g. [37, 54]). The highest local maximum along a ray of visibility is selected as the optimum depth. Correlation values are interpolated along depth to obtain subpixel accuracy.

Parameterizing the reconstruction volume into sectors instead of voxels provides of a useful surface parameterization, because the data required to compute visibility are already structured with respect to rays from the optical center. These data refer to a sector-interpretable grid (see Figure 9), but are structured in memory in a conventional 3D matrix. Application of visibility becomes then more natural, because the oblique traversal of a regular voxel space is sensitive to discretization artifacts. Finally, computational acceleration obtained by graphics-hardware is equally applicable to the resulting method as to plane-sweeping [32].

The proposed approach was compared to plane-sweeping on the same data and experimental conditions. Images were 480×640 pixels and target surfaces occurred from $1m$ to $3m$ from the cameras. In Figure 10, slices along depth that were extracted from the reconstructions are compared. Almost no effect between the two methods can be observed in the reconstructions, when obtained from the center of images (top row). The results differ the most when comparing reconstructions obtained from the periphery of images (rest of rows). In terms of reconstructed area, sphere-sweeping provided about $\approx 15\%$ more reconstructed points.

4.4 Combination of approaches

The methods presented in this chapter were combined into a stereo algorithm, which combines the efficiency of space-sweeping with the accuracy of orientation optimization. Results are shown in Figures 11, 12 and 13. The algorithm initiates by reconstructing a given scene with the method of sphere-sweeping (see Section 4.3). Then the proposed orientation-optimizing operator is employed and similarity local maxima are detected. Indicatively, performing the correlation step for the last example, $286sec$ were required on a Pentium

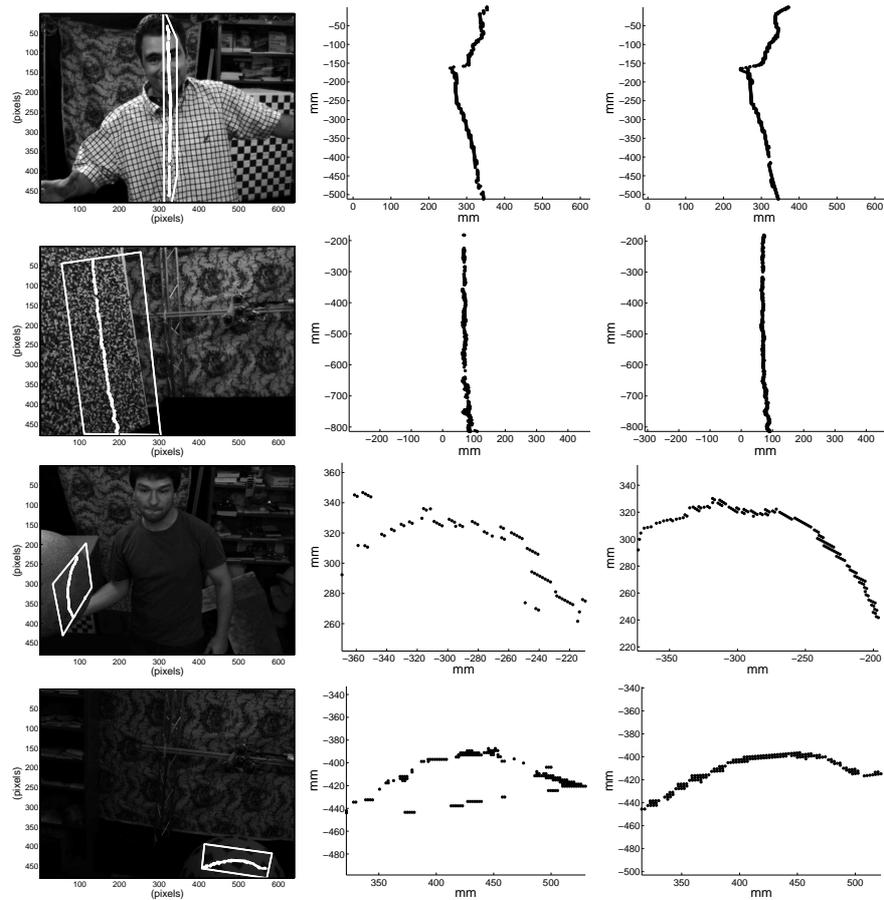


Fig. 10. Comparison of plane and sphere-sweeping. Each row shows an image from a 156 mm -baseline binocular pair (left) and a planar section (as in Figure 8) in space illustrating the reconstructions obtained from plane (middle) and sphere (right) sweeping. The reconstruction points shown are the ones that occur within the limits of the planar section. In the first two rows, the section intersects perpendicularly the reconstructed surface. In the 3^{rd} row, the section is almost parallel to the viewing direction and intersects the imaged sphere perpendicularly to its equator. In the last row, the section is almost parallel to image plane. The limits of the sections and the reconstructed points are reprojected on the images of the left row. The superiority of spherical sweeping is pronounced at the sections of the reconstructions that correspond to the periphery of the image.

3.2GHz, where image size was $\approx 10^6$ pixels, voxel size was 10 cm, $r = 21$, $\gamma = 60^\circ$.

A wide outdoors area reconstruction, in Figure 11, demonstrates the multiview expansion of the algorithm. For multiple views, at the end of each scale-iteration the space-carving rule [9] is applied to detect empty voxels and further reduce the computation at the next scale. At the last scale, the obtained \mathbf{V} s from each view are combined with the algorithm in [28].

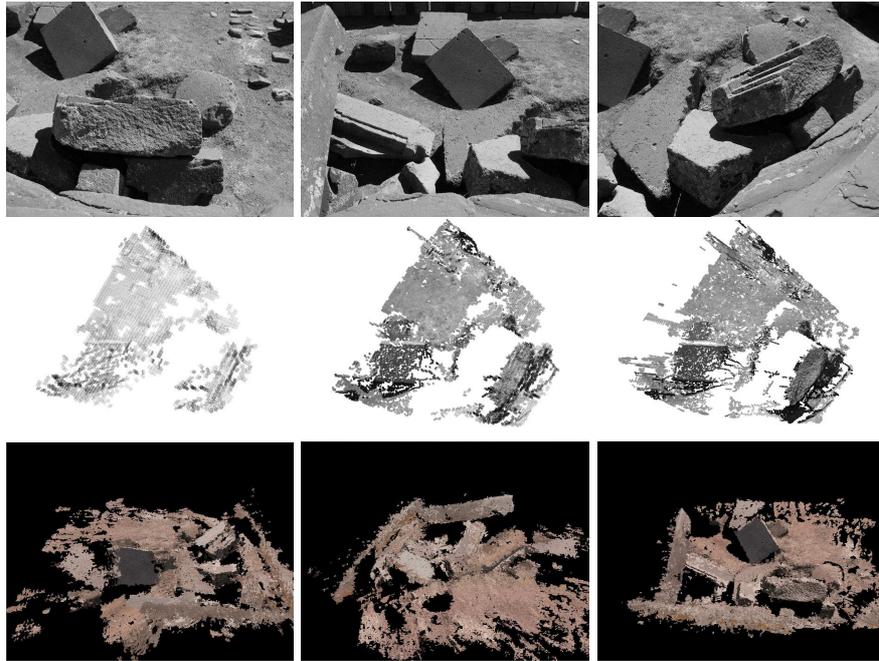


Fig. 11. Reconstruction results. Images from three 20 cm-baseline binocular pairs (1st row). Demonstration of coarse to fine spatial refinement of reconstruction for the 1st binocular pair (2nd row), using the approach of Section 4.2. Multiview reconstruction of the scene (3rd row).

5 Conclusion

This chapter is concerned with the depth cue due to the assumption of texture uniqueness, as one of the most powerful and widely utilized approaches to shape-from-stereo. The factors that affect the accuracy of the uniqueness cue were studied and the reasons that cast orientation-optimizing methods of superior accuracy to traditional and space-sweeping approaches were explained.

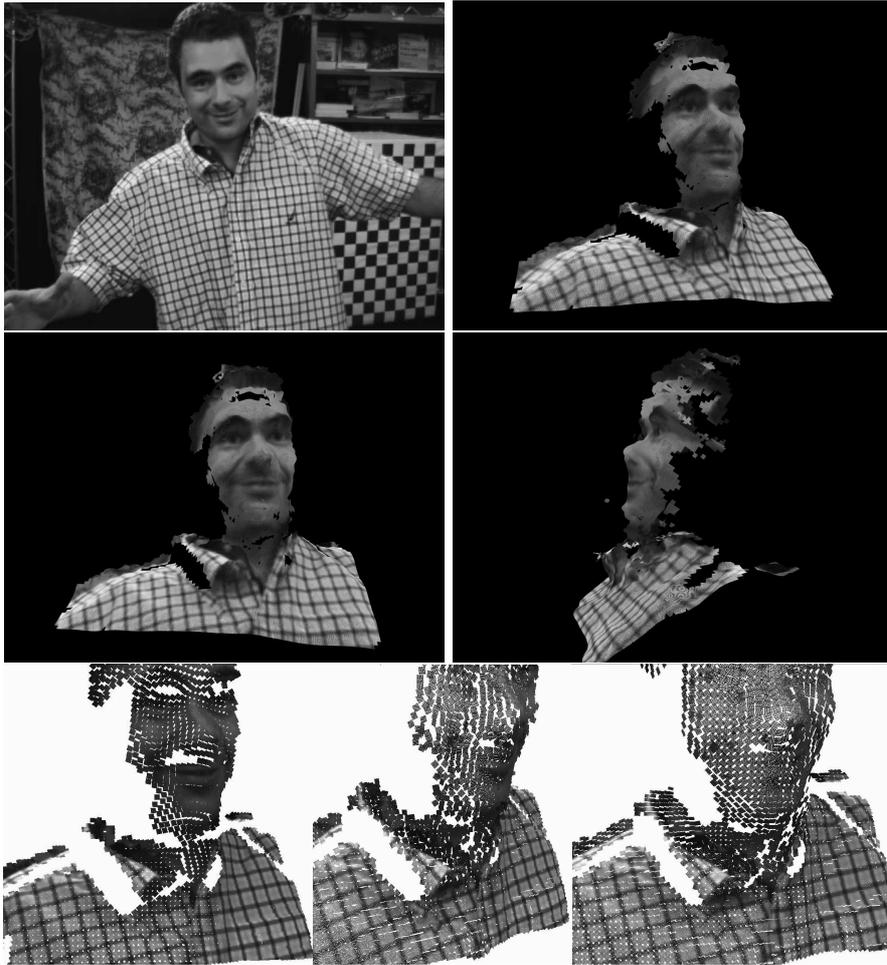


Fig. 12. Image from a binocular pair and reconstruction, utilizing the precision enhancement of Section 4.1. Output is a point cloud and image size was 640×480 . Last row compares comparison of patch-based reconstructions for plane-sweeping (left), orientation optimization (middle) and enforcement of continuity (right).

Furthermore, the proposed orientation-optimizing techniques improve the accuracy and precision orientation optimization as to date practiced.

Acceleration of the orientation-optimization approach is achieved by the introduction of two coarse-to-fine techniques that operate in the spatial and angular domain of the patch-based optimization. Finally, a hybrid approach is proposed that utilizes the rapid execution of a novel, accuracy-enhanced version of space-sweeping to obtain an initial approximation of the reconstruction

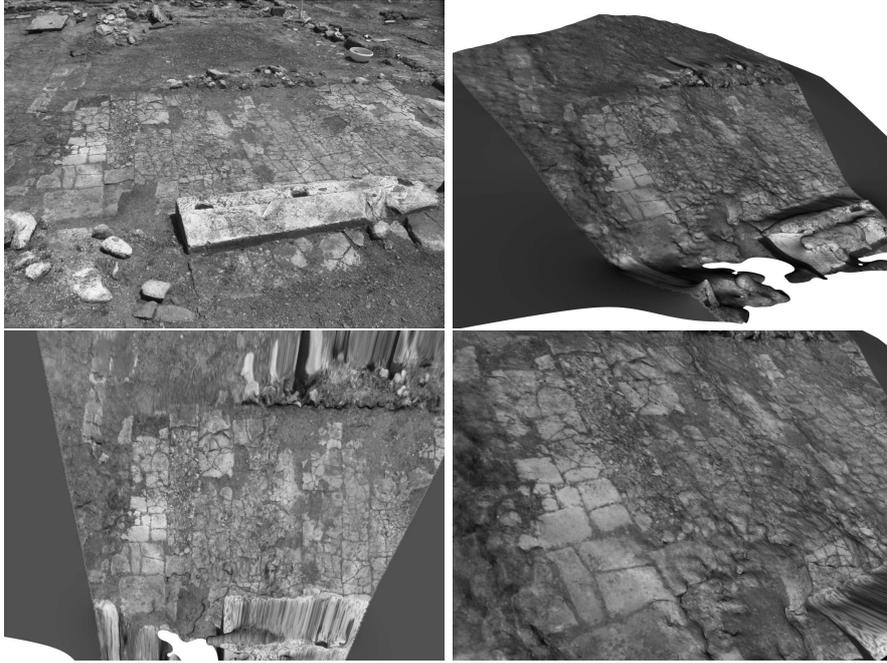


Fig. 13. Image from a 40 *cm*-baseline binocular pair and reconstruction, utilizing the precision enhancement of Section 4.1. The output isosurface is represented in a mesh and texture is directly mapped on it.

result. This result is then refined, based on the proposed techniques for the orientation-optimization of hypothetical surface patches.

The proposed extensions of to the implementation of the uniqueness cue to depth are integratable with diverse approaches to stereo. The size-modulation proposed in Section 3.3 is directly applicable to any approach that utilizes a planar surface patch, either in simple or multi-view stereo. Furthermore, utilization of the orientation-optimizing implementation of the uniqueness constraint has been demonstrated to be compatible with the assumption of surface continuity and, thus, applicable for global optimization approaches to stereo. A future research avenue of this work is the integration of the uniqueness cue with other cues to depth.

The next direction of this work is the utilization of parallel hardware for the real-time computation of wide area reconstructions, based on the fact that each of the proposed methods in this chapter is massively parallelizable. Computing the similarity s and orientation κ can be independently performed for each voxel and each orientation. Furthermore, detection of similarity local maxima can be also computed in parallel, if the voxel space is tessellated into overlapping voxel neighborhoods. Regarding the sweeping method that was

proposed, the similarity for each depth layer and for each pixel within this layer can also independently computed. Once the similarity values for each depth layer are available, the detection of the similarity-maximizing depth value for each column of voxels along depth can be also independently performed.

Biography

Xenophon Zabulis received his B.A., M.S. and Ph.D. degrees in Computer Science degree from the University of Crete in 1996, 1998 and 2001, respectively. He has worked as a Postdoctoral Fellow at the Computer and Information Science Department, at the interdisciplinary General Robotics, Automation, Sensing and Perception laboratory and at the Institute for Research in Cognitive Science, both at the University of Pennsylvania. He is currently a Research Fellow at the Institute of Informatics and Telematics - Centre of Research and Technology Hellas.

A Appendix

It is shown that s is maximized at the locations corresponding to the imaged surface only when values of s are computed from collineations that are parallel to the surface. Definitions are initially provided. Let:

- Cameras at T, Q that image a locally planar surface,
- Sweeping direction \mathbf{v} , which given a base point (e.g. the cyclopean eye) defines a line \mathcal{L} , which intersects the imaged surface at K . The two general types of possible configurations are shown in Figure 14.
- The hypothetical backprojection planar patch \mathcal{S} , where the acquired images are backprojected. Patch \mathcal{S} is on \mathcal{L} , centered at $D \in \mathcal{L}$ and oriented as \mathbf{v} .
- Function $b(X_1, X_2)$, $X_{1,2} \in R^3$ be the intersection of the line through X_1 and X_2 with the imaged surface. The surface point that is imaged at some point A on \mathcal{S} is $b(A, T)$, where T is the optical center. In the figure, $B = b(A, Q)$ and $C = b(A, T)$.
- Point O the orthogonal projection of A on the surface.
- θ and ϕ the acute angles formed by the optical rays through A , from T and Q .

Assuming texture uniqueness, the backprojection images of B and C are predicted to be identical only when B and C coincide. Thus the distance $|BC|$ for some point on \mathcal{S} is studied, assuming that when $|BC| \rightarrow 0$ correlation of backprojections is maximized. As seen in Figure 14,

- $|BC|$ is either $|OB| + |OC|$ (middle) or $|OC| - |OB|$ (left),
- $|AO| = |AB| \sin \theta = |AC| \sin \phi$, $|OC| = |AC| \cos \phi$, $|OB| = |AB| \cos \theta$, and

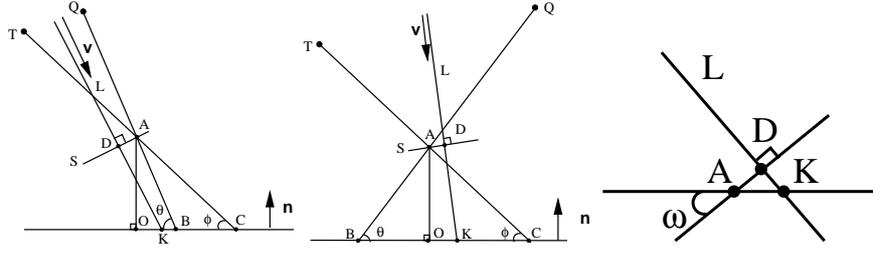


Fig. 14. Proof figures. See corresponding text.

- $\phi < \theta < \pi/2$.

Thus, $|BC|$ is either $|AO| \frac{\tan \theta + \tan \phi}{\tan \theta \tan \phi}$ or $|AO| \frac{\tan \theta - \tan \phi}{\tan \theta \tan \phi}$. Both quantities are positive. The first because $\theta, \phi \in (0, \pi/2]$ and the second because $\theta > \phi$ (see left on the above figure),

Therefore the monotonicity of $|BC|$ as a function of $\delta = |KD|$ is fully determined by $|AO|$. In the special case, where θ or ϕ is $\pi/2$, say θ , $|BC| = |OC| = |AC| \cos \phi = |AO| / \tan \phi$ and $|BC|$ is a monotonically increasing function of δ . Thus, similarity of the backprojection images on \mathcal{S} is indeed maximized when D coincides with K . This case corresponds to $\psi = 0$ (see forward in text).

Thus, when $\mathbf{v} \neq \mathbf{n}$, it is only for point D that $b(D, T)$ and $b(D, Q)$ will coincide when $\delta = 0$. The geometry for all other points on \mathcal{S} is shown in Figure 14 (right). From the figure, it is shown that in this case $b(A, T)$ and $b(A, Q)$ coincide only when $\delta > 0$ ($|KD| > 0$). For all the rest of the points the depth error is:

$$|KD|^2 = r^2 \tan^2 \omega \quad (7)$$

which shows that the error is determined not only by r ($|AD|$) but also from the ‘‘incidence angle’’ $\psi = \frac{\pi}{2} - \omega$ between \mathbf{v} and \mathbf{n} . Equation 7 shows that when $\mathbf{v} = \mathbf{n}$ ($\omega = 0$) the similarity of backprojections is maximized at the location of the imaged surface (when $\delta = 0$). In contrast, when $\omega \rightarrow \frac{\pi}{2}$ (or, when the surface is imaged from an extremely oblique angle) reconstruction error (KD) tends to infinity.

Since Equation 7 holds for every point on \mathcal{S} , it is concluded that the error in reconstruction is a monotonically increasing function of the angle between the incidence angle ψ or the search direction relative to the imaged surface.

References

1. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-2-3):7–42, 2002.

2. M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.
3. D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.
4. D. Marr and T. Poggio. A computational theory of human stereo vision. In *Royal Society of London Proceedings Series B*, volume 204, pages 301–328, 1979.
5. A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.
6. K. M. Cheung, T. Kanade, J. Y. Bouguet, and M. Holler. A real time system for robust 3D voxel reconstruction of human motions. In *IEEE Computer Vision and Pattern Recognition*, volume 2, pages 714 – 720, 2000.
7. V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *European Conference in Computer Vision*, volume 1, pages 379–393, 2002.
8. M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, 1993.
9. K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):197–216, 2000.
10. W. Culbertson, T. Malzbender, and G. Slabaugh. Generalized voxel coloring. In *Vision Algorithms: Theory and Practice*, pages 100–115, 1999.
11. G. Slabaugh, B. Culbertson, T. Malzbender, M. Livingston, I. Sobel, M. Stevens, and R. Schafer. Methods for volumetric reconstruction of visual scenes. *International Journal of Computer Vision*, 3(57):179–199, 2004.
12. J. Lanier. Virtually there. *Scientific American*, 284(4):66–75, 2001.
13. J. Mulligan, X. Zabulis, N. Kelshikar, and K. Daniilidis. Stereo-based environment scanning for immersive telepresence. *IEEE Circuits and Systems for Video Technology*, 14(3):304–20, 1999.
14. N. Kelshikar, Zabulis X., K. Daniilidis, V. Sawant, Sinha S., T. Sparks, S. Larsen, H. Towles, K. Mayer-Patel, H. Fuchs, J. Urbanic, K. Benninger, R. Reddy, and G. Huntoon. Real-time terascale implementation of tele-immersion. In *International Conference in Computer Science*, pages 33–42, 2003.
15. N. Ayache. *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*. MIT Press, Cambridge MA, 1991.
16. A. Trucco, E. and Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, New Jersey, 1998.
17. R. T. Collins. A space-sweep approach to true multi-image matching. In *IEEE Computer Vision and Pattern Recognition*, pages 358–363, 1996.
18. J. Bauer, K. Karner, and K. Schindler. Plane parameter estimation by edge set matching. In *26th Workshop of the Austrian Association for Pattern Recognition*, pages 29–36, 2002.
19. C. Zach, A. Klaus, J. Bauer, K. Karner, and M. Grabner. Modelling and visualizing the cultural data set of Graz. In *Virtual reality, archeology, and cultural heritage*, 2001.
20. K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.
21. C. Zhang and T. Chen. A self-reconfigurable camera array. In *Eurographics Symposium on Rendering*, 2004.

22. T. Werner, F. Schaffalitzky, and A. Zisserman. Automated architecture reconstruction from close-range photogrammetry. In *CIPA - International Symposium*, 2001.
23. C. Zach, A. Klaus, B. Reitinger, and K. Karner. Optimized stereo reconstruction using 3D graphics hardware. *Workshop of Vision, Modelling, and Visualization*, pages 119–126, 2003.
24. A. Bowen, A. Mullins, R. Wilson, and N. Rajpoot. Light field reconstruction using a planar patch model. In *Scandinavian Conference on Image Processing*, pages 85–94, 2005.
25. R. Carceroni and K. Kutulakos. Multi-View scene capture by surfel sampling: From video streams to Non-Rigid 3D motion, shape & reflectance. *International Journal of Computer Vision*, 49(2-3):175–214, 2002.
26. O. Faugeras and R. Keriven. Complete dense stereovision using level set methods. In *European Conference in Computer Vision*, pages 379–393, 1998.
27. A. S. Ogale and Y. Aloimonos. Stereo correspondence with slanted surfaces: critical implications of horizontal slant. In *IEEE Computer Vision and Pattern Recognition*, volume 1, pages 568–573, 2004.
28. 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.
29. D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 131–140, 2001.
30. M. Li, M. Magnor, and H. P. Seidel. Hardware-accelerated rendering of photo hulls. *Eurographics*, 23(3), 2004.
31. R. Yang, G. Welch, and G. Bishop. Real-time consensus-based scene reconstruction using commodity graphics hardware. In *Pacific Graphics*, 2002.
32. 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 *International Conference on Image Processing*, 2006.
33. M. Pollefeys and S. Sinha. Iso-disparity surfaces for general stereo configurations. In *European Conference on Computer Vision*, pages 509–520, 2004.
34. V. Nozick, S. Michelin, and D. Arqus. Image-based rendering using plane-sweeping modelisation. In *International Association for Pattern Recognition - Machine Vision Applications*, pages 468–471, 2005.
35. S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2):151–173, 1999.
36. R. Szeliski. Prediction error as a quality metric for motion and stereo. In *International Conference in Image Processing*, volume 2, pages 781–788, 1999.
37. 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 *IEEE International Symposium on 3D Data Processing, Visualization and Transmission*, pages 534–541, 2004.
38. H. Moravec. Robot rover visual navigation. *Computer Science: Artificial Intelligence*, pages 105–108, 1980/1981.
39. J. Mulligan and K. Daniilidis. Real time trinocular stereo for tele-immersion. In *ICIP*, pages 959–962, Thessaloniki, Greece, 2001.
40. S. Paris, F. Sillion, and L. Qu. A surface reconstruction method using global graph cut optimization. In *Asian Conference in Computer Vision*, 2004.

41. K. Junhwan, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *International Conference in Image Processing*, volume 2, pages 1033–1040, 2003.
42. Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, 1985.
43. I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs. A maximum likelihood stereo algorithm. *CVIU*, 63(3):542–567, 1996.
44. P. Mordohai and G. Medioni. Dense multiple view stereo with general camera placement using tensor voting. In *IEEE International Symposium on 3D Data Processing, Visualization and Transmission*, pages 725–732, 2004.
45. J. Montagnat, H. Delignette, and Ayache N. A review of deformable surfaces: topology, geometry and deformation. *Image and Vision Computing*, 19:1023–1040, 2001.
46. H. Jin, S. Soatto, and A. Yezzi. Multi-view stereo beyond lambert. In *IEEE Computer Vision and Pattern Recognition*, pages 171–178, 2003.
47. J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
48. O. Monga, R. Deriche, G. Malandain, and J. P. Cocquerez. Recursive filtering and edge tracking: two primary tools for 3d edge detection. *Image and Vision Computing*, 9(3):203–214, 1991.
49. G. Turk and J. F. O’Brien. Modelling with implicit surfaces that interpolate. *ACM Transactions on Graphics*, 21(4):855–873, 2002.
50. J. C. Carr, R. K. Beatson, J.B. 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 *ACM - Special Interest Group on Graphics and Interactive Techniques*, pages 67–76, 2001.
51. W. Lorensen and H. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):169–169, 1987.
52. G. Turk and M. Levoy. Zippered polygon meshes from range images. In *ACM - Special Interest Group on Graphics and Interactive Techniques*, pages 311–318, 1994.
53. J. X. Chai, X. Tog, S. C. Chan, and H. Y. Shum. Plenoptic sampling. In *ACM - Special Interest Group on Graphics and Interactive Techniques*, pages 307–318, 2000.
54. C. Zach, K. Karner, and H. Bischof. Hierarchical disparity estimation with programmable 3D hardware. In *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pages 275–282, 2004.