A Quasi-Dense Approach to Surface Reconstruction from Uncalibrated Iages

Maxime Lhuillier, Long Quan

▶ **To cite this version:**

HAL Id: hal-00091032
https://hal.science/hal-00091032

Submitted on 5 Sep 2006

A Quasi-Dense Approach to Surface Reconstruction from Uncalibrated Images

Maxime LHUILLIER[1] and Long QUAN[2]

[1] LASMEA-UMR 6602 UBP/CNRS, 24 av. des Landais, 63177 Aubière, France.

[2] Dep. of Computer Science, HKUST, Clear Water Bay, Kowloon, Hong Kong SAR.

E-mails: [1]lhuillie@lasmea.univ-bpclermont.fr   [2]quan@cs.ust.hk

Abstract

This paper proposes a quasi-dense approach to 3D surface model acquisition from uncalibrated images. Firstly, correspondence information and geometry are computed based on new quasi-dense point features that are re-sampled sub-pixel points from a disparity map. The quasi-dense approach gives more robust and accurate geometry estimations than the standard sparse approach. The robustness is measured as the success rate of full automatic geometry estimation with all involved parameters fixed. The accuracy is measured by a fast gauge-free uncertainty estimation algorithm. The quasi-dense approach also works for more largely separated images than the sparse approach, therefore requires fewer images for modeling. More importantly, the quasi-dense approach delivers a high density of reconstructed 3D points on which a surface representation can be reconstructed. This fills the gap of insufficiency of the sparse approach for surface reconstruction, essential for modeling and visualization applications. Secondly, surface reconstruction methods from the given quasi-dense geometry are also developed. The algorithm optimizes new unified functionals integrating both 3D quasi-dense points and 2D image information including silhouettes. Combining both 3D data and 2D images is more robust than the existing methods using only 2D information or only 3D data. An efficient bounded regularization method is proposed to implement the surface evolution by level-set methods. Its properties are discussed and proved for some cases. As a whole, a complete automatic and practical system of 3D modeling from raw images captured by hand-held cameras to surface representation is proposed. Extensive experiments demonstrate the superior performance of the quasi-dense approach with respect to the standard sparse approach in robustness, accuracy, and applicability.

Keywords: 3D reconstruction, surface reconstruction, structure from motion, 3D modeling, matching, uncertainty, variational calculus, level-set method.

1 Introduction

Passive vs. active methods: Three-dimensional model acquisition has always been one of the fundamental research topics in computer vision. Active 3D scanners are currently the dominant technology for capturing digital object models for applications. Their geometric accuracy has continually improved. But they remain expensive, and, more importantly, they suffer from a number of technical limitations. They are invasive and some materials such as hair can not be scanned. They are also not "scalable" to objects of different sizes, especially large ones and outdoor scenes. In comparison, passive image-based modeling from collections of images captured by handheld cameras offers several advantages. It needs only low-cost hardware, it can be applied to objects of any size, and also it preserves the appearance information from original photographs while maintaining perfect geometric alignment.

Sparse and dense approaches: There are two approaches for reconstruction from images. For uncalibrated images, the standard is based on the sparse points of interest developed in the last decade [10, 20, 37, 12, 41, 2, 26, 21]. This sparse approach is sufficient for computing or tracking camera positions, but not for representing the scene or the objects in the scene as it merely reconstructs sparsely distributed 3D points. For calibrated images (the sparse approach can be used for calibration purpose to start a dense method), dense stereo methods including the traditional direct stereo matching and more recent volumetric approaches of simultaneously reconstructing the object and computing the dense correspondence [46, 25, 9, 24] are the usual approaches to reconstruction. The main disadvantages of the best dense stereo methods are that they only reconstruct smoothed layers of disparities, special configurations in handling multiple views (often in one half-space looking at the other half-space, and for images of small baselines), and they are very expensive in terms of time and memory [24]. In practice, to handle uncalibrated images, a combination of these two methods is a natural choice [37, 41]. For instance, from a sparse geometry, dense stereo matching algorithms are run for some selected pairs in [40] and all in [38]. The dense reconstruction is triangulated and texture-mapped to obtain the final models. The surface models obtained are often partial and the surface triangulation is simply inherited from a 2D triangulation in one image plane, which means that the surface topology can not be properly handled.

Quasi-dense approach: Motivated by the insufficiency of the existing sparse and dense approaches, we develop in this paper a quasi-dense approach to surface reconstruction from a sequence of uncalibrated images. This gives a more robust and accurate geometry estimation, a quasi-dense geometry, using fewer images. It fills the gap of insufficiency of the sparse approach by delivering a high density of 3D points from uncalibrated images that makes a surface reconstruction tractable.

In addition to presenting a complete system of 3D modeling from raw images captured from hand-held cameras, the main contributions of this paper are threefold:

- The introduction of point features as the re-sampled sub-pixel points from the quasi-dense disparity map to densify the feature points, thereby overcoming the sparseness of the points of interest method.

- An automatic quasi-dense geometry computation from uncalibrated images. Compared with the standard sparse approach, the quasi-dense approach not only gives more robust and accurate reconstruction results, but it works for largely separated images. More importantly, it produces a high density of points that can be used for direct surface reconstruction. A fast gauge-free estimation algorithm is also developed for an efficient evaluation of the reconstruction accuracy.

- New surface reconstruction algorithms integrating both 3D data points and 2D image informations. This is possible because of a unified functional based on a minimal surface formulation. We believe that the new functional has far fewer local minima than those derived from 2D data alone and that this will result in more stable and more efficient algorithms. For the efficient evolution of surfaces, we also propose a bounded regularization method based on level-set methods. Its stability is also proved.

Paper organization: Section 2 introduces the quasi-dense sub-pixel point features and correspondences and describes their computation. Section 3 describes the whole procedure of the estimation of the quasi-dense geometry by stressing two-view analysis and fast gauge-free uncertainty estimation. Section 4 presents the surface reconstruction by integrating the quasi-dense 3D information and 2D image information. Section 5 gives more implementation details and experimental results, and Section 6 concludes the paper. Sections 3 and 4 are extensions to our previous conference papers [28, 29].

2 Quasi-dense correspondences

In this section, we introduce and define the concept of quasi-dense point correspondences as our "point" features. We also describe the computation procedures.

2.1 Quasi-dense pixel correspondences by match propagation

We start with the standard sparse matching algorithm between two images to detect the points of interest [30, 18] in each image. Points of interest are naturally reliable point features sparsely distributed in each image space. They are also tractable for widely separated images having larger disparities. A ZNCC (Zero-Mean Normalized Cross Correlation) method, followed by cross validation, is used to match these points of interest in two images. This gives a list of sparse point correspondences that contains inevitable errors.
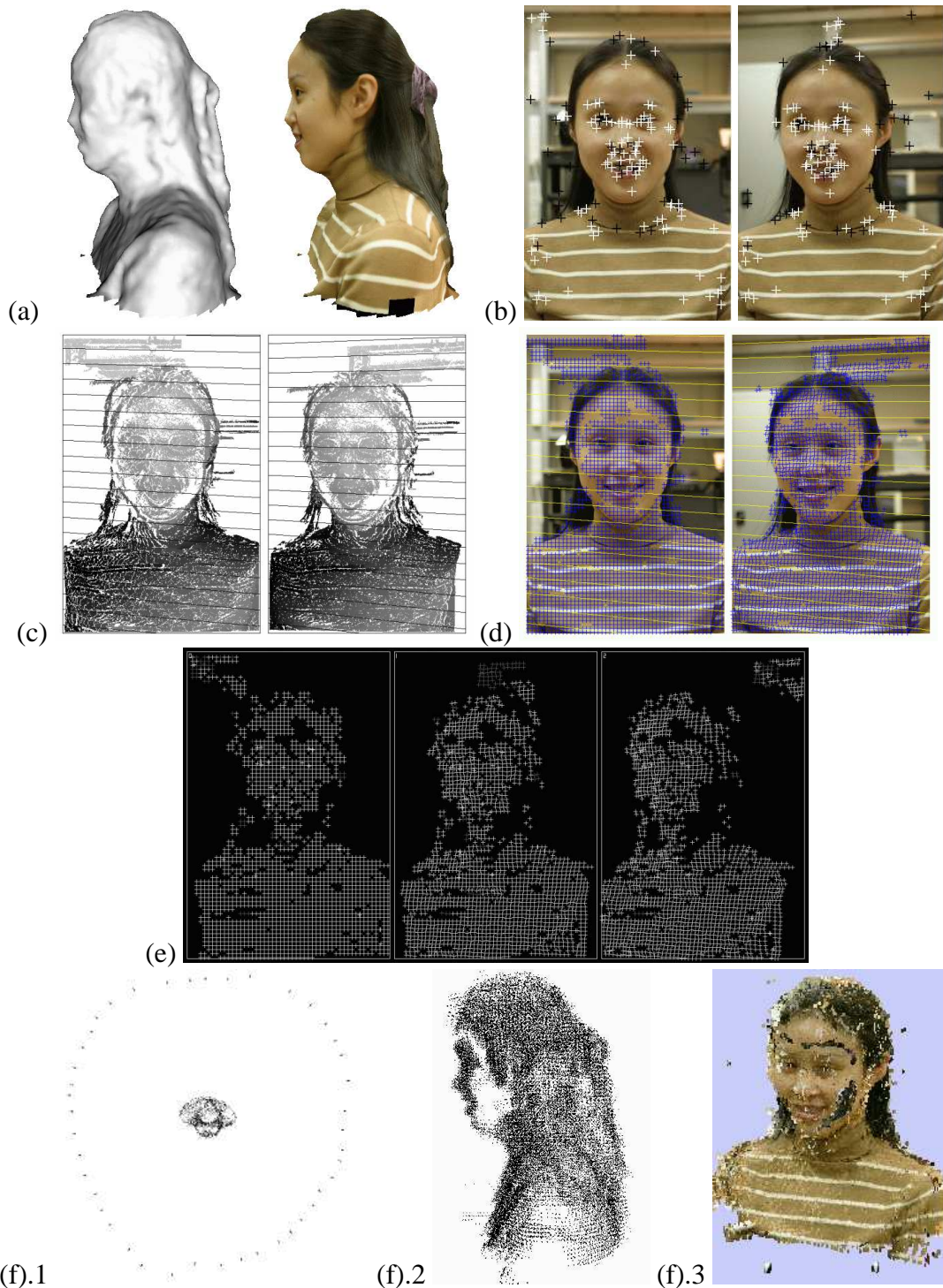
Figure 1: The goal to reach is (a) the reconstructed surface geometry (Gouraud shaded and textured mapped). Also, key steps of the quasi-dense approach are given. (b) The initial sparse correspondence points of interest for a pair of images. Correspondence outliers (still unknown at this stage) are marked in black. (c) The quasi-dense disparity map by two propagations and the estimated epipolar geometry. (d) The re-sampled quasi-dense correspondence points. (e) The refined quasi-dense correspondences in a triplet of images. Inliers (resp. outliers removed by the three-view geometry) are marked in white (resp. gray). (f) The Euclidean quasi-dense geometry: foreground object, background and the camera poses, each camera is displayed with a small black pyramid. (f).1 a top view of the whole geometry. (f).2 a close-up view of the face in point cloud. (f).3 each 3D quasi-dense point is displayed with a small patch of texture.

4

The standard sparse approach uses a robust statistical method to remove correspondence outliers by fitting the underlying fundamental matrix [63, 54] to the list of sparse point correspondences between the two images. Instead, we "densify" the correspondences by match propagation. We first sort this list of point correspondences using the correlation score. These sorted point correspondences are called seed points. At each step of the propagation, we choose the best corresponding pixels scored by ZNCC from the current list of seed points. Then, in the immediate spatial neighborhood of the seed points, we look for new potential matches and add the bests to the current list of seed points according to a combination of local constraints such as correlation, gradient disparity, and confidence. The matching uniqueness and the ending of the process are guaranteed by choosing only new matches that have not yet been selected. A more detailed description of match propagation and its properties can be found in [27]. In [3, 61, 56, 49], the matches of seed points are used as ground control points to densify the disparity map for stereo matching algorithms.

It is important to note that ours is a very efficient algorithm both in time and space, and at each time only the best match is selected. This drastically limits the possibility of bad matches. For instance, the seed selection step seems very similar to many existing methods [63, 54] for matching points of interest using correlation, but the crucial difference is that we need only take the most reliable ones rather than trying to match a maximum number of them. In extreme cases, only one good seed match is sufficient to provoke an avalanche of correspondences in the textured images. This makes our algorithm much less vulnerable to bad seeds. The same is true for propagation. The risk of bad propagation is considerably diminished by the best-first strategy over all matched boundary points.

This best-first match propagation approach produces denser, but not completely dense, pixel correspondences that we call quasi-dense pixel correspondences. One example for a real pair of images is illustrated in Figures 1.c and 3.a.

2.2   Quasi-dense (sub-pixel) point correspondences by homographic re-sampling

Quasi-dense pixel correspondences are not directly used as our "point" features in subsequent computations. Instead, we re-sample these pixel correspondences into what we call quasi-dense point correspondences.

On one hand, the re-sampling is motivated by the fact that the quasi-dense pixel correspondences give an irregular distribution of clusters of pixels, which is not suitable for geometry computation. Many clustered pixels do not create strong geometric constraints while making the estimation cost high. Re-sampling produces not only a reduced set and more uniform distribution of matched points in the images, but it also creates matching points with sub-pixel accuracy. On the other hand, the re-sampling is equally motivated by the necessity of post-match regularization to improve match reliability by integrating local geometric constraints, since the quasi-dense pixel correspondences

5

may still be corrupted by wrong correspondences.

We assume that the scene or object surface is at least locally smooth. Therefore, instead of directly using global geometric constraints encoded by a fundamental matrix, we first use local smoothness constraints encoded by local planar homography : the quasi-dense pixel correspondences are regularized by locally fitting local homographies to them.

The first image plane is initially divided into a regular square grid of 8×8 pixels. This size is a trade-off between the sampling resolution and regularization stability. For each square patch, all pixel correspondences inside it from the quasi-dense pixel correspondences are used to tentatively fit a plane transformation. The most general linear plane transformation is a homography represented by a homogeneous 3×3 non singular matrix. Four matched pixels, no three of them collinear, are sufficient to estimate a plane homography. In practice, an affine transformation encoded by 6 d.o.f. using three matched pixels rather than a homography is preferred as the local perspective distortion is often mild between images. Because of unavoidable matching errors and the points not lying on the dominant local plane (e.g. at the occluding contours), the putative transformation for a patch cannot be estimated using standard least squares estimators. The Random Sample Consensus (RANSAC) [11] is used for a robust estimation of the transformation, $\mathbf{H}$. Finally, for each confirmed patch correspondence, a pair of corresponding points, $\mathbf{u}_i \leftrightarrow \mathbf{H}_i\mathbf{u}_i$, with sub-pixel accuracy is created by selecting a representative center point of the patch in the first image $\mathbf{u}_i$ and its homography-induced corresponding point, $\mathbf{H}_i\mathbf{u}_i$, in the second image. The set of all corresponding points created this way is called the "quasi-dense correspondences". In practice, we also add to the quasi-dense point correspondences all corresponding points of interest within the patch and validated by the homography of the patch. Usually these points of interest have longer tracks along the sequence than other points obtained by propagation. This definition of quasi-dense correspondences is illustrated in Figure 2, and an example from a real image pair is given at (c) of Figure 4.

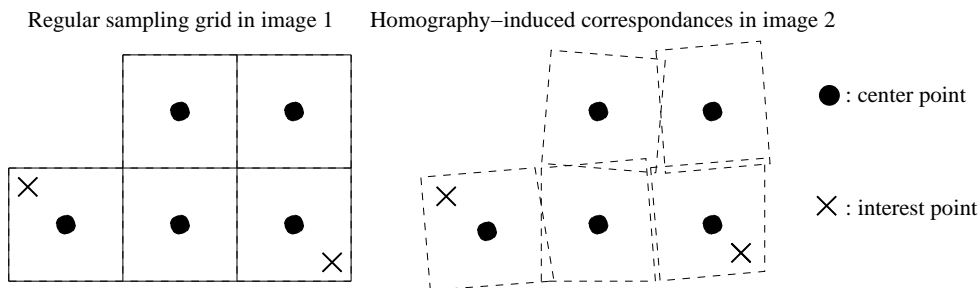Figure 2: For each corresponding patch, the re-sampled points include the center point of the patch and all points of interest within the patch and their homography-induced correspondences in the second image.

These re-sampled corresponding points are not only more suitable for geometric computation thanks to their more uniform distribution in images, but they also are more reliable as the robust local

homography fitting significantly singled out match errors contained in the original quasi-dense pixel correspondences, as illustrated at the middle and left of Figure 3.

3   Estimation of a Quasi-Dense Geometry

The geometric estimation of a sequence of uncalibrated images, including both camera positions and the 3D reconstruction of scene points, is now standard for sparse points of interest [20, 10]. Mostly, we will apply some of the standard algorithms to our new "point" features, the quasi-dense point correspondences. But we also propose two new algorithms. The first is the core two-view quasi-dense correspondence and geometry method that turns out to be much more robust and accurate than the sparse methods. The second is a fast gauge-free uncertainty estimation, necessary for our development. A description of the whole optimization procedure including three-view projective and N-view projective and Euclidean geometry parameterization and estimation is briefly given in Appendix A.2 for the completeness of the system.

3.1   Robust two-view quasi-dense correspondence and geometry

The two-view geometry of a rigid scene is entirely encoded by the fundamental matrix. The standard strategy is to recover geometry using sparse matching [63, 54] within a random sampling framework. We propose two procedures for fundamental matrix estimation in our quasi-dense approach: the first is the *constrained propagation* that grows only those satisfying the current epipolar constraint. The second is the *unconstrained propagation*. The unconstrained propagation is motivated by the fact that the estimation might be local, biased toward the areas with a high density of matches (for example, either merely the background or the foreground or a dominant plane) if the initial distribution of the matched points is not uniform across images. This bias due to the irregular distribution of the points in image space is well-known and discussed in [20]. The final strategy that combines these procedures and overcomes the disadvantage of each is given as follows:

1. Detect feature "points of interest" in each image; establish the initial correspondences between the images by computing normalized correlation; Sort the validated correspondences by correlation score and use them to initialize a list of seed matches for match propagation;

2. Unconstrained propagation from all the seed points using a best-first strategy without the epipolar constraint to obtain quasi-dense pixel correspondences represented as a disparity map;

3. Re-sample the quasi-dense disparity map by local homographies to obtain the quasi-dense (sub-pixel) point correspondences; Estimate the fundamental matrix using a standard robust algorithm on the re-sampled points, i.e., the quasi-dense correspondences;

4. Constrained propagation from the same initial list of seeds using a best-first strategy with the epipolar constraint by the computed fundamental matrix;

5. Again re-sample the obtained quasi-dense disparity map to get the final quasi-dense point correspondences; Reestimate the fundamental matrix with the final quasi-dense correspondences.

The result of this procedure is a list of quasi-dense sub-pixel correspondences satisfying the epipolar constraint. These sub-pixel matches are usually more reliable, denser, and more evenly distributed over the whole image space than the standard sparse points of interest. Figure 1 shows the major steps of the computation for a typical pair of images. The computational time for a pair of images is more costly than a standard sparse method, but it is limited about 10-15s for 512×512 images on a P4 2.4Ghz. Figure 3 illustrates the incremental robustification of correspondences in different steps. We have also tested the strategy of combining a sparse geometry and a constrained propagation, and have found that the domain of the final propagation tends to be reduced and results in the undesirable local estimates discussed earlier.
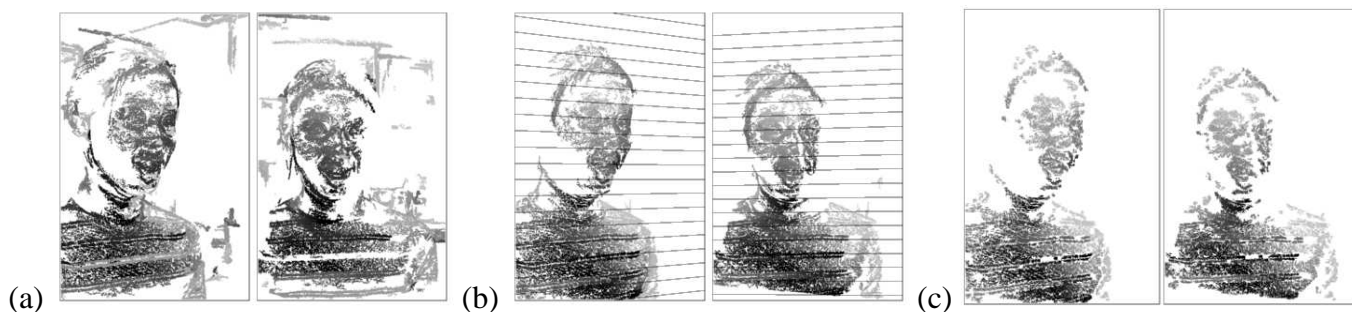

(a) (b) (c)

Figure 3: The quasi-dense disparity map, (a) after the first unconstrained propagation, (b) after the second constrained propagation, (c) after robust local homography fitting that removes many wrong pixel correspondences for the second constrained propagation.

Another advantage of this strategy is that it works for more largely separated image pairs than those acceptable by the standard sparse approach for the simple reason that the number of matched interest points dramatically decreases with an increasing geometric distortion between views. However, we do not compare our approach with specific sparse methods such as affine invariant regions [59, 32] and points [44, 53] matching methods. Figure 4 shows comparative results between the sparse and the quasi-dense methods for a widely-separated pair for which the standard sparse method fails in computing a wrong fundamental matrix. Also, in our experiments, we may use as few as about 20 images to make a full turn of the object, which might be impossible for the standard sparse approach.

Figure 4: (a) Initial sparse correspondences by cross-correlation for a pair of images with large disparities. Only 31 matches in white out of 111 are correct. (b) Failure of the standard sparse method. Many correspondence points in black are obviously incorrect. (c) Successful quasi-dense estimation.

3.2   Fast gauge-free uncertainty estimation

To assess the accuracy of the final 3D reconstruction obtained using the global bundle adjustment described in Appendix A, the covariance matrix should be estimated both for camera positions and for each reconstructed point. As the underlying Hessian matrix is of extremely large size and singular in the gauge-free situation, we develop a fast gauge-free covariance estimation inspired by recent work [58, 20].

The covariance matrix could be estimated as the inverse of Hessian, $\mathbf{H}^{-1} = (\mathbf{J}^T\mathbf{J})^{-1}$, up to a common noise level, σ^2, which can also be estimated if $\mathbf{H}$ is not singular, i.e. if there is no gauge freedom. For numerical efficiency, the current bundle optimization has been carried out with an over-parametrized free gauge. We need to solve two major problems: the first is that the final $\mathbf{H}$ after optimization is now singular due to free gauge; and the second is that $\mathbf{H}$ is excessively large in size. The singularity of $\mathbf{H}$ could have been easily handled by a direct SVD based pseudo-inverse if it were not excessively large in size.

Normal covariance matrix   When $\mathbf{H}$ is non-singular and has a specific sparse structure as in our case, it can be block-diagonalized into $\mathbf{H} = \mathbf{T}\mathbf{A}\mathbf{T}^T$. Then, the pseudo-inverse of $\mathbf{H}$ can be efficiently computed as

$$\mathbf{H}^+ = \mathbf{H}^{-1} = (\mathbf{T}\mathbf{A}\mathbf{T}^T)^+ = (\mathbf{T}^{-T})\mathbf{A}^+\mathbf{T}^{-1},$$

exactly as in the basic reduction technique used in photogrammetry [4, 20, 10, 35, 34, 58].

Now the final resulting $\mathbf{H}$ after minimization is singular due to the free gauge. Though it is still formally possible to compute $\mathbf{H}^*$ as $\mathbf{T}^{-T}\mathbf{A}^+\mathbf{T}^{-1}$, it is no more the pseudo-inverse $\mathbf{H}^+$ of $\mathbf{H}$. We need to clarify its underlying statistical meaning. The choice of coordinate fixing rules is a gauge fixing [58]. Each choice of gauge, locally characterized by its tangent space, determines an oblique covariance matrix. The interpretation of $\mathbf{H}^*$ computed above is therefore an oblique covariance matrix at the particular solution point we have chosen by a first-order perturbation analysis around the maximum

9

likelihood solution [34]. It has been shown that all these oblique covariance matrices at a given solution point from different gauges are geometrically equivalent in the sense that they all have the same "normal" component in the orthogonal space to the gauge orbit. This normal component is called the normal covariance [34]. We choose this uncertainty description as it is convenient and does not require the specification of gauge constraints. It also gives a lower bound on all covariances defined at that point on the gauge orbit.

Fast computation To compute this more significant normal covariance, we need to project any oblique covariance onto the orthogonal space to the gauge orbit tangent space, i.e. $\mathbf{Cov} = \mathbf{PH^*P}^T$, where $\mathbf{P}$ is the projector to $Im(\mathbf{H})$ in the direction of $Ker(\mathbf{H})$. The major difficulty is handling the very large size of $\mathbf{H}$ to make the projection computable.

If $\mathbf{N}$ is an orthonormal basis of $Ker(\mathbf{H})$, then $\mathbf{P} = \mathbf{I} - \mathbf{NN}^T$. Since $dim(Ker(\mathbf{H})) = 7$, $\mathbf{N}$ is a thin matrix of dimensions $O(p+c) \times 7$, where p is the number of points and c is the number of cameras. Using the approximated Hessian $\mathbf{H} = \begin{pmatrix} \mathbf{C} & \mathbf{M} \\ \mathbf{M}^\top & \mathbf{S} \end{pmatrix}$, where $\mathbf{C}$ (resp. $\mathbf{S}$) is an inversible block-diagonal and sub-hessian of camera (resp. structure) parameters, we have $\mathbf{H} = \mathbf{TAT}^T = \begin{pmatrix} \mathbf{I} & \mathbf{Y} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{Y}^\top & \mathbf{I} \end{pmatrix}$ where $\mathbf{Y} = \mathbf{MS}^{-1}, \mathbf{Z} = \mathbf{C} - \mathbf{MS}^{-1}\mathbf{M}^\top$. Thus,

$$Ker(\mathbf{H}) = \begin{pmatrix} \mathbf{I} \\ -\mathbf{Y}^\top \end{pmatrix} Ker(\mathbf{Z})$$

and $\mathbf{N}$ is efficiently computed by a SVD of the matrix $\mathbf{Z}$ of small dimensions $O(c) \times O(c)$.

Using the notation $\mathbf{K} = \mathbf{H^*N}$, the computation of normal covariance $\mathbf{Cov}$ is given by

$$\mathbf{Cov} = \mathbf{PH^*P}^\top = \mathbf{H^*} - \mathbf{KN}^\top - \mathbf{NK}^\top + \mathbf{N(N^\top K)N}^\top.$$

The matrices $\mathbf{N}$ and $\mathbf{K}$ are very thin. Their width is only 7; the matrix $\mathbf{N}^\top\mathbf{K}$ is 7×7. The calculation complexity of all diagonal blocks of $\mathbf{Cov}$ for camera and point covariances is $O(c+p)$, provided that $\mathbf{K}$ and the corresponding diagonal blocks of $\mathbf{H^*}$ are computed.

The diagonal blocks of $\mathbf{H^*}$ are computed in time $O(pc^2 + c^3)$ and in memory $O(i + c^2)$ with i the number of 2D points [20, 4]. For $\mathbf{K}$, let $\mathbf{N}_c, \mathbf{N}_s$ be vectors such that $\mathbf{N}^\top = (\ \mathbf{N}_c^\top \quad \mathbf{N}_s^\top \ )$ and the height of $\mathbf{N}_c$ (resp. $\mathbf{N}_s$) is the same as that of $\mathbf{C}$ (resp $\mathbf{S}$). It is easy to verify that

$$\mathbf{K} = \begin{pmatrix} 0 \\ \mathbf{S}^{-1}\mathbf{N}_s \end{pmatrix} + \begin{pmatrix} \mathbf{I} \\ -\mathbf{Y}^\top \end{pmatrix} \mathbf{Z}^+(\mathbf{N}_c - \mathbf{YN}_s).$$

This calculation is feasible because of the small size of $\mathbf{Z}^+$ ($O(c) \times O(c)$) and the diagonal structure of $\mathbf{S}$. Time and space complexities are only $O(i + c^2)$.

Figure 5 illustrates one example of the computed uncertainty ellipsoids. We finish this fast gauge-free covariance computation by making the following observations:
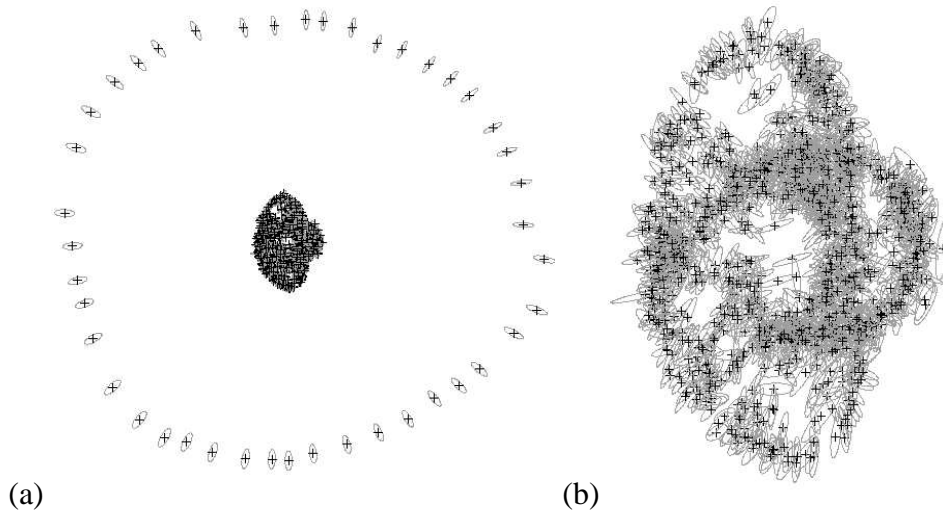
<center>(a)   (b)</center>

Figure 5: The 90% confidence ellipsoids zoomed by 4 for reconstructed points and camera centers. Only one out of 50 point ellipsoids are displayed. (a) a top view of the cameras and the object. (b) a close-up view of the only object on the right.

- The gauge theory has been reintroduced into computer vision in [58], which gives a very general exposition on issues related to the gauge-free covariance matrices. Hartley and Zisserman pointed out in [20] that a condition is necessary to have $\mathbf{H}^* = \mathbf{H}^+$. Unfortunately, this condition is not satisfied. Morris et al. [34] have justified the computation of $\mathbf{H}$ by defining a "geometrically" equivalent class of covariance matrices. However, they carried out only a small-scale bundle problem.

- This fast computation of uncertainty can also be extended to the computation of any gauged oblique covariance matrix with any oblique projection having a small kernel dimension as the projector still verifies $\mathbf{P} = \mathbf{I} - \mathbf{B}\mathbf{C}^{\top}$, and $\mathbf{B}$ and $\mathbf{C}$ are very thin matrices like $\mathbf{N}$.

- We also prove some invariant properties of the normal covariance matrix. The point and camera center ellipsoids are invariant with respect to rigid transformation (resp. scaling factor) if the scaling factor (resp. rigid transformation) is fixed.

4 Surface Reconstruction

In addition to increased robustness and accuracy, the most significant aspect of the quasi-dense approach is that it reconstructs a high density of 3D points on which we can build a surface representation of the objects. Surface-based representations as natural extensions of the point-based geometry are indispensable for most current modeling and visualization applications. This section describes the surface reconstruction algorithms from the quasi-dense 3D points.

<center>11</center>

Many surface reconstruction algorithms have been proposed for different data. For only 2D images and camera geometry, the recent volumetric methods [46, 25, 9, 24] are the most general image-based approaches, but they are not robust enough. For densely scanned 3D point data, Szeliski et al. [51] used a particle-based model of deformable surfaces; Hoppe et al. [22] presented a signed distance for implicit surfaces; Curless and Levoy [7] described a volumetric method; and Tang and Medioni [52] introduced a tensor voting method. Most recently, Zhao et al. [64] developed a level-set method based on a variational method of minimizing a weighted minimal surface. Similar work to [64] has also been reported by Whitaker [60] using a MAP framework. For depth data obtained from stereo systems, it is more challenging than that from scanned 3D data as the stereo data are usually much sparser and less regular. Fua [14] used a system of particles to fi t the stereo data. Kanade et al. [36] and Fua and Leclerc [13] proposed a deformable mesh representation to match the multiple dense stereo data. These methods that perform reconstruction by deforming an initial model or tracking the discretized particles to fi t the data points are both topologically and numerically limited compared to modern dynamic implicit surface approaches.

The insuffi cient 3D reconstruction from images and the diffi culties of obtaining surface data from only images have motivated us to develop a new approach that integrate both quasi-dense 3D points and all available 2D image information including image correlation, photo-consistency, and silhouette information if it is available. We propose a variational approach with new functionals integrating 3D points and 2D image information. An effi cient bounded regularization method to implement the surface evolution by level-set is also developed.

4.1   Problem statement and general approach

Given a set of calibrated 2D images and a set of quasi-dense 3D points derived from the given images, the goal is to reconstruct a surface representation of the objects in the scene.

The problem is different from surface reconstruction from a set of calibrated images as addressed in [9, 46, 25] in which only 2D images are used without any 3D information. It is also different from surface reconstruction from scanned 3D data without 2D image information [22, 51, 7, 64, 52].

The general methodology that we follow is a variational approach inspired by the work of Faugeras and Keriven [9], Caselles et al. [6, 5], Zhao et al. [64], and many others. Intrinsic functionals as a kind of weighted minimal surface are defi ned to integrate both 3D point data and 2D image data. The object surfaces are represented as a dynamic implicit surface, $u(\mathbf{x}) = 0$ in $\mathbf{R}^3$, which evolves in the direction of the steepest descent provided by the variation calculation of the functional we minimize. The intrinsic nature of the functionals (i.e., independent of any surface parametrization) makes the implementation of the surface evolution by the level-set method possible, which in turn handles the surface topology changes.

Our contribution is threefold: First, we show that the accuracy of the reconstructed 3D points is sufficient for the 3D modeling application. Second, we introduce new intrinsic functionals that take into account both 3D data points and 2D original image information, unlike previous works that consider either only 2D image information [9] or only scanned 3D data [64]. By doing this, we compensate for the lack of reconstructed 3D points with 2D information. The new functionals are also expected to have a much smaller number of local minima and better convergence than a pure 2D approach [9]. Third, we propose a bounded regularization method that is more efficient than the usual full regularization methods and give a proof of its stability.

4.2 Defining the functionals

By analogy to 2D geodesic active contours [6] whose mathematical properties have been established, the weighted minimal surface formulation was introduced by Caselles et al. [5] and Kichenassamy et al. [23] for 3D segmentation from 3D images, i.e., the 3D surfaces they seek are those minimizing the functional $\int \int w ds$ using the weight $w = g(\nabla I)$ where ds is the infinitesimal surface element and g is a positive and decreasing function of the 3D image gradient ∇I.

Faugeras and Keriven [9] developed a surface reconstruction from multiple images by minimizing the functional $\int \int w ds$ using a weighting function w that measures the consistency of the reconstructed objects reprojected onto 2D images. This measure is usually taken as a function of the correlation functions $\rho(\mathbf{x}, \mathbf{n})$ between pairs of 2D images, i.e., $w(\mathbf{x}, \mathbf{n}) = g(\rho(\mathbf{x}, \mathbf{n}))$. The correlation function is dependent not only on the position $\mathbf{x}$ of the object surface, but also on its orientation $\mathbf{n}$. A potentially general and powerful reconstruction approach was therefore established. But the existence and uniqueness of the solution for the proposed functional have not yet been elucidated.

In the different context of surface reconstruction from sufficiently dense and regular sets of scanned 3D point data, Zhao et al. [64] proposed to minimize the functional $\int \int w ds$ using a new weighting function w as the distance function of any surface point $\mathbf{x}$ to the set of 3D data points. Given a set of data points $\mathcal{P}$ and $d(\mathbf{x}, \mathcal{P})$ the Euclidean distance of the point $\mathbf{x}$ to $\mathcal{P}$, the weighting function is simply $w(\mathbf{x}) = d^p(\mathbf{x}, \mathcal{P})$. The method gives interesting results with good 3D data points.

In our surface reconstruction, we have both 3D data points and 2D image data. It is interesting to observe that the variational formulation mentioned above in different contexts is based on the minimal surface. This makes it possible to define a unifying functional taking into account data of a different nature. Thus, we first propose to minimize the functional $\int \int w ds$ using a new weighting function for the minimal surface formulation consisting of two terms $w(\mathbf{x}, \mathbf{n}) = d^p(\mathbf{x}, \mathcal{P}) + \lambda e(\mathbf{x}, \mathbf{n}, I)$ where the first $d(\mathbf{x}, \mathcal{P})$ is the 3D data attachment term that allows the surface to be attracted directly onto the 3D points; and the second $e(\mathbf{x}, \mathbf{n}, I)$ is a consistency measure of the reconstructed object in the original 2D image space. The consistency measure might be taken to be any photo-consistency or correlation

function. The minimizing functional is given by

$$p(\mathbf{x}) = \int \int (d^p(\mathbf{x}, \mathcal{P}) + \lambda e(\mathbf{x}, \mathbf{n}, I)) ds.$$

Silhouette information might also be a useful source of information for surface construction [50]. It is not suffi cient on its own as it gives only an approximate visual hull, but it is complementary to other sources of information. If used, it amends the distance function of the weighting function as

$$d(\mathbf{x}, \mathcal{P} \cup \mathcal{S}) = min(d(\mathbf{x}, \mathcal{P}), \epsilon + d(\mathbf{x}, \mathcal{S})),$$

where d is the 3D Euclidean distance function; $\mathcal{P}$ is the set of 3D points; $\mathcal{S}$ is the surface of the intersections of the cones defi ned by the silhouettes, i.e., the visual hull; and ϵ is a small constant favoring 3D points over the visual hull in the neighborhood of 3D points. An adequate initialization is also proposed to optimize the functional derived from this weighting function.

4.3   Solving the variational problem

The solutions of the minimizing functional are given by a set of PDEs: the Euler-Lagrange equation designated $\nabla p = 0$, and obtained from the functional $p = \int \int w ds$ to be minimized. The Euler-Lagrange equation is often impossible to solve directly. One common way is to use an iterative and steepest-descent method by considering a one-parameter family of smooth surfaces $\mathbf{x}(t) : (u, v, t) \mapsto (x(u, v, t), y(u, v, t), z(u, v, t))$ as a time-evolving surface $\mathbf{x}$ parametrized by time t. The surface moves in the direction of the gradient of the functional p with the velocity $-\nabla p$, according to the flow $\mathbf{x}_t = \frac{\partial \mathbf{X}(u,v,t)}{\partial t} = -\nabla p$. This is the Lagrangian formulation of the problem that describes how each point on the dynamic surface moves in order to decrease the weighted surface. The fi nal surface is then given by the steady state solution $\mathbf{x}_t = 0$. The problem with this approach is that it does not handle the topology change [47]. However, it is important to notice that though the derivation has been based on a parametrization, the various quantities including the velocity for the steepest descent flow are intrinsic, i.e., independent of any chosen parametrization that makes the computation possible. This paves the way for the well-known and powerful level-set formulation [39, 47] that regards the surface as the zero level-set of a higher dimensional function. As the flow velocity $-\nabla p$ is intrinsic (it has been demonstrated for a general w depending also on the surface normal in [9]), we may easily embed it into a higher dimensional smooth hyper-surface $u(t, \mathbf{x}) = 0$ which evolves according to $u_t = -(\nabla p \cdot \mathbf{n})||\nabla u||_2$, and the normal $\mathbf{n} = -\frac{\nabla u}{||\nabla u||_2}$. Topological changes, accuracy, and stability of the evolution are handled using the proper numerical schemes developed by Osher and Sethian [39].

4.4 A bounded regularization method

The Bounded Regularization Method   The Euler-Lagrange expression ∇p might be complicated if the weighting function $w(\mathbf{x}, \mathbf{n})$ also depends on the surface normal [9]. It seems that the complication by this dependency on the surface normal is rather unnecessary in practice [16]. We therefore assume a weighting function independent of the surface normal. Thus, the expression $\nabla p \cdot \mathbf{n}$ consists simply of two terms like the geodesic active contour case, $\nabla w \cdot \mathbf{n} + w\nabla \cdot \mathbf{n}$, in which the first is the data attachment term and the second the regularization term. Using $\mathbf{n} = -\frac{\nabla u}{||\nabla u||_2}$ on the level-set function, the surface evolves according to

$$\frac{\partial u}{\partial t} = \nabla w \nabla u + w||\nabla u||_2 H,$$

where $H = div\frac{\nabla u}{||\nabla u||_2}$ is the sum of the two principal curvatures (twice the mean curvature). When w is taken to be the correlation function, it is the simplified version of [9] presented in [16]. And when w is taken to be the 3D distance function, it is the first method proposed in [64]. However the curvature-based regularization $w||\nabla u||_2 H$ over-smooths, resulting in a loss of geometric details and in slow convergence as the time step has to be $\Delta t = O(\Delta x^2)$ for a stable solution.

In [65], a convection model is also proposed to ignore the regularization term, $w||\nabla u||_2 H$, to speed up the procedure, but this is only envisageable for applications where data quality is sufficient, for instance, for synthetic and high-quality scanned data [65].

Motivated by the need for regularization of noisy data and the inefficiency of the curvature-based regularization, we propose an intermediate bounded regularization method. It has a "bounded" regularization term, $min(w, w_{max})||\nabla u||_2 H$, instead of the "full" regularization term, $w||\nabla u||_2 H$. The corresponding evolution equation is given as:

$$\frac{\partial u}{\partial t} = \nabla w \nabla u + min(w, w_{max})||\nabla u||_2 H.$$

The following remarks can be made:

- The fully regularized surface evolution is obtained when $w_{max} \geq ||w||_\infty$.

- The unregularized surface evolution is obtained when $w_{max} = 0$.

- As $0 \approx w \leq w_{max}$ in the vicinity of the steady surface for any w, it is expected that the fully regularized and the bounded regularized evolutions behave in the same manner in this region.

Efficiency of the bounded regularization method   The efficiency of our proposed bounded regularization method is evaluated by estimating the maximum time step Δt_{max} for stability computation. We are currently unable to quantify Δt_{max} of the bounded regularization method for the general

15

curvature-based regularization, but we are able to prove it for a simplified isotropic regularization using a Laplacian operator. Replacing the curvature-based regularization by the isotropic regularization for Δt_{max} calculation is a heuristic motivated by the fact that the curvature/anisotropic regularization term, $||\nabla u||_2 H = ||\nabla u||_2 div \frac{\nabla u}{||\nabla u||_2}$, and the Laplacian/isotropic one $\triangle u$ are equal when $||\nabla u||_2 = 1$ is enforced periodically, which is the case often in practice to avoid too flat and too steep variations of u. It is therefore tempting to simplify the evolution equation to

$$\frac{\partial u}{\partial t} = \nabla w \nabla u + min(w, w_{max}) \triangle u.$$

Assuming that the stability condition is the same for curvature-based and Laplacian-based regularizations, the stability, $||u^{n+1}||_\infty \leq ||u^n||_\infty$, is achieved if $\Delta t \leq \Delta t_{max}$ with

$$\Delta t_{max} = \frac{\Delta x^2}{6w_{max} + ||\Delta x(|d^{0x}w| + |d^{0y}w| + |d^{0z}w|)||_\infty},$$

where d^{0x}, d^{0y}, and d^{0z} are the centered differences at the grid point in the three axes. The proof is given in Annex B.

We choose w_{max} to be proportional to Δx for our bounded regularization method, i.e., fixing $w_0 = \frac{w_{max}}{\Delta x}$, and obtain

$$\Delta t_{max} = \frac{\Delta x}{6w_0 + ||||\nabla w||_1||_\infty}.$$

Under this condition, the complexity of Δt_{max} is given by $\Delta t_{max} = \Theta(\Delta x)$, the same for the bounded regularized and unregularized evolutions, much better than $\Delta t_{max} = \Theta(\Delta x^2)$ for the fully regularized evolution.

No previous work to our knowledge provides such a stability analysis for an evolution equation with both convection and regularization terms. In practice, the time step, $\Delta t = \Delta t_{max}$, is always used for surface evolution in all our examples with the bounded and curvature-based regularization.

5 Implementation and experiments

Experimental data are available at `maxime.lhuillier.free.fr` or `www.cs.ust.hk/˜quan`

5.1 Quasi-dense geometry estimation

5.1.1 Comparative experiments

Representative real examples of the quasi-dense reconstruction (QUASI) are given and compared with the standard sparse methods (SPARSE) to demonstrate the superior performance of QUASI both in accuracy and robustness. The results on synthetic sequences are presented in [28] due to space limitations.

Implementation of the sparse methods   The first method simply tracks all points of interest detected in each individual image. The second is a mixture of sparse and quasi-dense methods: it assesses points of interest from individual images by geometry that is computed from the quasi-dense algorithm and re-evaluates the whole geometry only from these matched points of interest. In the following, SPARSE indicates the best result of these two methods.

Reconstruction accuracy   is measured by considering the bundle adjustment as the maximum likelihood estimates, if we assume that the image points are normally distributed around their true locations with an unknown standard deviation σ. The confidence regions for a given probability can therefore be computed from the covariance matrix of the estimated parameters. The covariance matrix is defined only up to the choice of the gauge [58, 33, 35] and the common unknown noise level σ^2. The noise level σ^2 is estimated from the residual error as $\sigma^2 = r^2/(2e - d)$, where r^2 is the sum of the e squared re-projection errors, d is the number of independent parameters of the minimization: $d = 1 + 6c + 3p - 7$ (1 is the common focal length, c is the number of cameras, p is the number of reconstructed points and 7 is the gauge freedom choice). We use our fast gauge-free uncertainty estimation method presented in Section 3.2 to compute the normal covariance matrix $\mathbf{H}^+$ from the oblique covariance matrix $\mathbf{H}^*$ in the coordinate system of the camera in the middle of the sequence and with the scale unit equal to the maximum distance between camera centers. We choose a 90% confidence ellipsoid for any 3D position vector, either camera position or 3D point. The maximum of semi-axes of the 90% confidence ellipsoid is computed as the uncertainty bound for each 3D position. The camera uncertainty is characterized by taking the mean of all uncertainty bounds of camera positions $\overline{\mathbf{x}}_{c_i}$ as the number of cameras is moderate. The point uncertainty is characterized by computing the rank 0 (the smallest uncertainty bound $\mathbf{x}_0$), rank $\frac{1}{4}$ ($\mathbf{x}_{\frac{1}{4}}$), rank $\frac{1}{2}$ (median $\mathbf{x}_{\frac{1}{2}}$), rank $\frac{3}{4}$ ($\mathbf{x}_{\frac{3}{4}}$) and rank 1 (the largest uncertainty bound $\mathbf{x}_1$) of the sorted uncertainty bounds, as the number of points is very high. The uncertainty of the focal length f is given by the standard deviation σ_f. We give detailed experimental results for some typical real sequences. The Lady 1 sequence (20 images at 768×512) has a more favorable lateral motion in close-range. The uncertainties given in Table 2 and Figure 7 for QUASI are smaller than for SPARSE, and three to six times smaller for focal length and camera positions. Similar conclusions hold for all the sequences shown in Figure 14 for which SPARSE succeeds. The Lady 2 sequence (43 images at 408×614) is captured with an irregular but complete tour around the central person. Table 3 and Figure 1 show the results. The Garden-cage sequence (34 images at 640×512) was captured by a hand-held still camera (Olympus C2500L) with an irregular but complete tour using a rather short focal length to increase the viewing field for the larger background. The Garden-cage sequence contains a close-up of a bird cage and a background of a house, and tree, with a very profound viewing field. SPARSE methods failed because some triplets of consecutive images do not have sufficiently matched points of interest. The QUASI method gives the uncertainties

listed in Table 4 and 90% ellipsoids shown in Figure 8. As the images were captured with the smallest focal length available, the camera's non-linear distortion became non-negligible. After a first round of Euclidean bundle adjustment, a second adjustment by adding one radial distortion parameter ρ for all cameras is carried out. We find that $\rho = -0.086$. This result is similar to that obtained with a very different method proposed in [8] for the same camera but different images: $\rho = -0.084$. The corridor sequence from Oxford University (11 images at 512×512 resolution) has a lateral forward motion along the scene which does not provide strong geometry, but favors the SPARSE method as it is a low textured polyhedric scene in which matched points of interest are abundant and spread well over the scene. With almost 40 times redundancy in the number of points, camera position and focal length uncertainties for QUASI are two to four times smaller than for SPARSE. However, the point uncertainties are almost of the same order of magnitude for the majority of points. As the camera direction and path are almost aligned with the scene points, many points on the far background of the corridor are almost at infinity. Not surprisingly with the actual fixing rules of the coordinate choice, they have extremely high uncertainty bound along the camera direction for both methods as illustrated in Figure 6.

According to the discussions in Section 3.2, $\mathbf{H}^+$ gives the normal covariance matrix while $\mathbf{H}^*$ is an oblique covariance matrix at a given solution point as was previously used in [35, 20]. The normal covariance matrix should be the '*smallest*' one in the sense of the matrix trace, the lower bound of all oblique covariance matrices. We want empirically to demonstrate this by comparing these different covariance matrices. In all cases, the main uncertainty changes are those of camera centers which are bigger for the normal covariance matrix than for the oblique one, while the trace of the whole normal covariance matrix is slightly smaller than that of the oblique one as expected. This suggests that the normal covariance matrix describes a better distribution of uncertainties between cameras and structures.

Reconstruction robustness   To measure the reconstruction robustness, we consider the success rate of reconstruction for all tested sequences in this paper as illustrated in Table 6. The robustness of QUASI with respect to the sampling rate of the sequence is also experimented. For the Lady 2 sequence making a complete tour around the object, SPARSE fails for a sequence of 43 images, but succeeds for a sequence of 86 images with only 1827 sparse points. QUASI succeeds until a subsequence of 28 images with 25339 quasi-dense points. A typical pair of this subsequence is shown in Figures 4 and 3. It is clear that QUASI has superior robustness: whenever a sequence is successful for SPARSE, it is equally successful for QUASI, while SPARSE fails for many sequences (including those not shown in this paper). Furthermore, even when SPARSE is successful, it is sometimes only the mixed SPARSE that is successful. Recall that the SPARSE method was defined as the best result of a pure sparse and a mixed sparse-quasi method. We also notice that our QUASI method requires

Corridor	#3D points	σ	f	σ_f	$\overline{\mathbf{x}}_{c_i}$	$\mathbf{x}_0$	$\mathbf{x}_{\frac{1}{4}}$	$\mathbf{x}_{\frac{1}{2}}$	$\mathbf{x}_{\frac{3}{4}}$	$\mathbf{x}_1$	Tr(Cov)
QUASI	16976	0.41	714	4.36	7.0e-4	.014	.070	.13	.38	15700	4.59e7
SPARSE	427	0.52	761	17.3	1.7e-3	.016	.056	.12	.32	106	2254
QUASI	16976	0.41	714	4.36	65	2.6	20	28	31	6809	1.52e7
SPARSE	427	0.52	761	17.3	3.5	0.24	1.09	1.71	2.06	81	1829
QUASI	16976	0.41	714	4.40	2.0e-2	1.5e-2	6.9e-2	0.12	0.33	22	1171
SPARSE	427	0.52	761	17.4	9.0e-2	2.7e-2	6.2e-2	0.12	0.31	10	362

Table 1: Uncertainty measures for the Corridor sequence: the mean of the uncertainty bounds of camera centers and the rank-k of the sorted uncertainty bounds of points, calculated from the oblique (resp. normal) covariance matrix at top (resp. middle and bottom). Middle (resp. bottom): they are physically implausible (resp. plausible) for SPARSE and QUASI without (resp. with) deleting points that are further away than 5 from the middle of the camera movement (camera length= 1).

Lady 1	#3D points	σ	f	σ_f	$\overline{\mathbf{x}}_{c_i}$	$\mathbf{x}_0$	$\mathbf{x}_{\frac{1}{4}}$	$\mathbf{x}_{\frac{1}{2}}$	$\mathbf{x}_{\frac{3}{4}}$	$\mathbf{x}_1$	Tr(Cov)
QUASI	26823	.53	849	2.26	6.1e-4	1.2e-3	4.2e-3	5.4e-3	6.1e-3	1.9e-2	5.24
SPARSE	383	.54	866	13.6	3.8e-3	5.2e-3	9.8e-3	1.1e-2	1.3e-2	2.6e-2	186.269
QUASI	26823	.53	849	2.26	3.7e-3	6.8e-4	4.0e-3	4.9e-3	5.6e-3	1.8e-2	5.21
SPARSE	383	.54	866	13.6	1.1e-2	1.0e-3	4.4e-3	5.1e-3	6.1e-3	2.1e-2	186.267

Table 2: Uncertainties for Lady 1 from the oblique (top) and normal (bottom) covariance matrix.

Lady 2	#3D points	σ	f	σ_f	$\overline{\mathbf{x}}_{c_i}$	$\mathbf{x}_0$	$\mathbf{x}_{\frac{1}{4}}$	$\mathbf{x}_{\frac{1}{2}}$	$\mathbf{x}_{\frac{3}{4}}$	$\mathbf{x}_1$	Tr(Cov)
QUASI	55410	.50	1294	1.65	1.1e-3	7.1e-4	1.8e-3	2.0e-3	2.3e-3	0.14	2.85
QUASI	55410	.50	1294	1.65	5.1e-3	6.1e-4	1.7e-3	2.0e-3	2.3e-3	0.14	2.83

Table 3: Uncertainties for Lady 2 from the oblique (top) and normal (bottom) covariance matrix.

Garden	#3D pts	σ	f	σ_f	ρ	σ_ρ	$\overline{\mathbf{x}}_{c_i}$	$\mathbf{x}_0$	$\mathbf{x}_{\frac{1}{4}}$	$\mathbf{x}_{\frac{1}{2}}$	$\mathbf{x}_{\frac{3}{4}}$	$\mathbf{x}_1$	Tr
QUASI	50161	.46	732	.27	-0.086	1.1e-4	3.8e-4	5.2e-4	1.9e-2	4.4e-2	.12	2.6	126
QUASI	50161	.46	732	.27	-0.086	1.1e-4	6.1e-3	4.3e-4	1.9e-2	4.4e-2	.12	2.6	124

Table 4: Uncertainties for Garden-cage from the oblique (top) and normal (bottom) covariance matrix.

	#cameras (resulution)	#3D points	matching and 2-views	3-views and merge
Corridor	11 (512 × 512)	16976	2 min	4 min
Lady 1	20 (768 × 512)	26823	4	6
Lady 2	43 (408 × 614)	55411	8	14
Garden-cage	34 (640 × 512)	50161	9	17

Table 5: Computation times in minutes for the QUASI method with a P4 2.4 Ghz.

	oxford	garden	bust	man1	apple	man2	lady1	man3	girl	lady2	man4	man5
Q	done	done	done	done	done	done	done	done	done	done	done	done
S	done	fail	done	fail	done	fail	done	fail	fail	fail	fail	fail

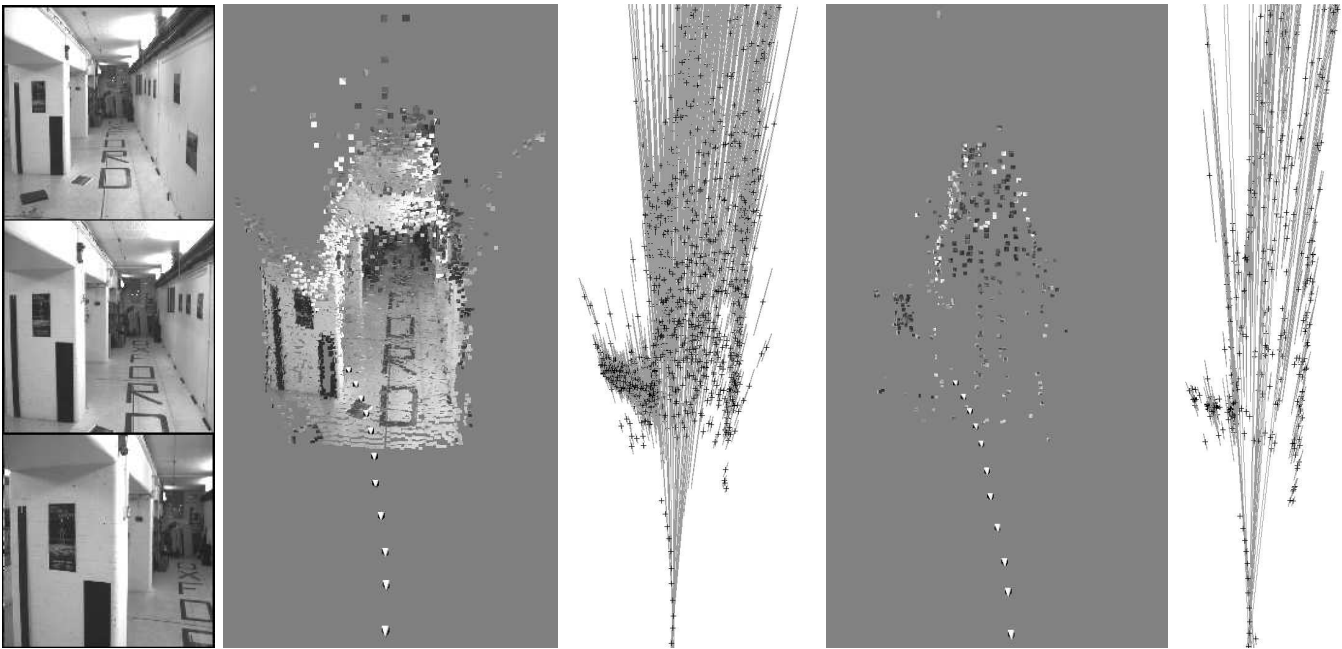Table 6: Automatic success rate of reconstruction between Q(uasi) and S(parse).

19

Figure 6: ¿From left to right: three of the 10 Corridor images, QUASI and SPARSE reconstructions for Corridor and their 90% confidence ellipsoids viewed on a horizontal plane. Only one out of 10 ellipsoids for QUASI are displayed.
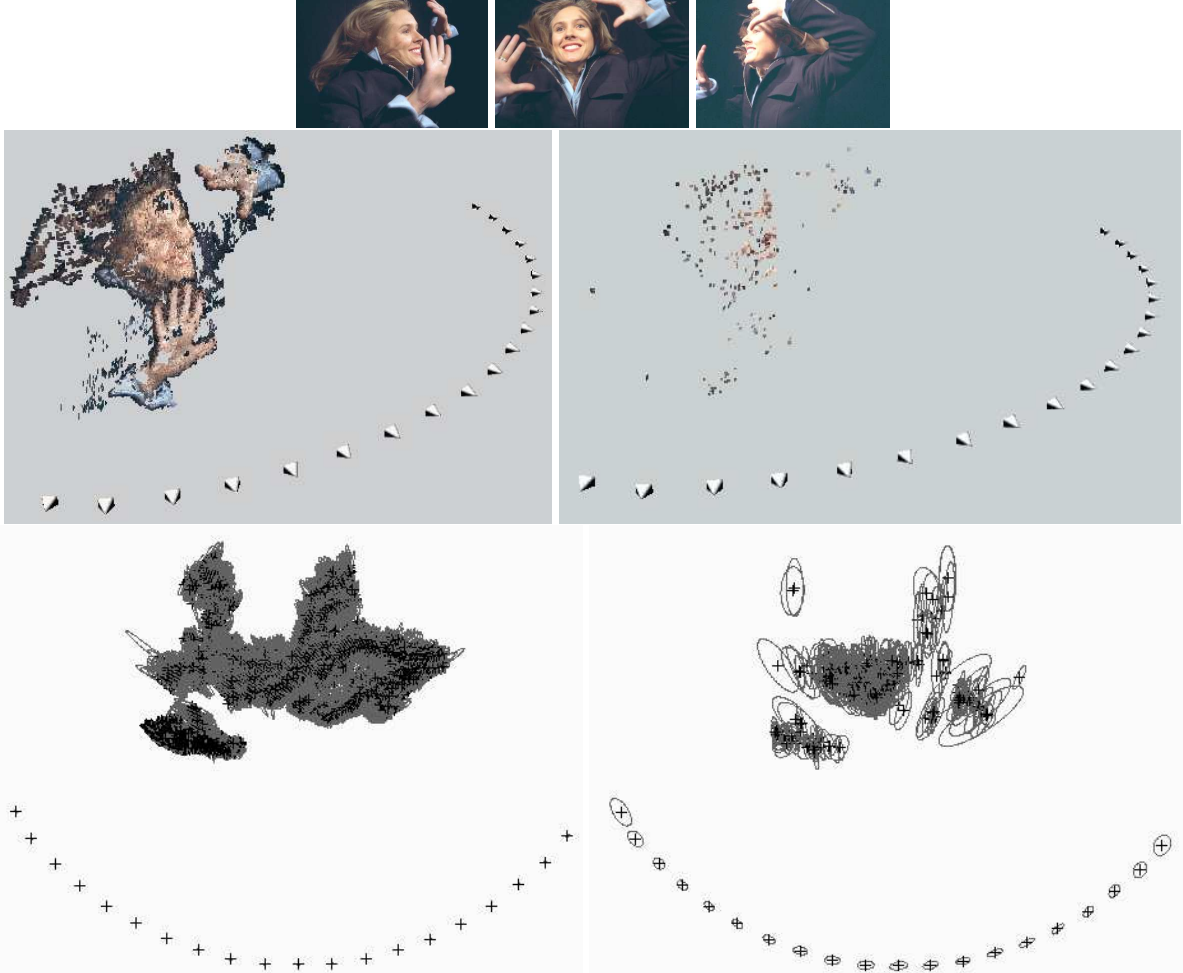

Figure 7: ¿From top to bottom and left to right: three of the 20 Lady 1 images, QUASI and SPARSE reconstruction for Lady 1, and their 90% ellipsoids (zoomed by 4) viewed on a horizontal plane.

Figure 8: Left: three of the 34 Garden-cage images. Right: top view of the 90% confidence ellipsoids. The small square shaped connected component at the center is the reconstructed bird cage while the visible crosses forming a circle are camera positions.

only about 30 to 35 frames for a complete tour around an object. This is far fewer than the 50 to 100 frames necessary for SPARSE methods such as [38, 41, 1], which are more suitable for video sequences.

5.2   Surface reconstruction from quasi-dense geometry

The surface reconstruction method is limited to smooth and closed objects. The outdoor scenes such as the garden-cage example are not handled. To model a complete object, we usually make a full turn around the object by capturing about 30 to 35 images to compute the geometry of the sequence.

5.2.1   Surface initialization from quasi-dense 3D points

The reconstructed 3D points are segmented into the foreground object and the background. The background includes obvious outliers like isolated and distant points from the majority of points. The points of the foreground object are obtained as the largest connected component of the graph neighborhood of all points such that the distance between any two "edge" points of this graph should be smaller than a multiple of the uncertainty median of the points. The surface initialization is then obtained as follows. The foreground object points are regularly sliced into sections along the major direction of the point cloud. A 2D-convex hull is computed for each section and these convex hulls

are used to defi ne the successive sections of a truncated cone as the bounding volume of the object. The initialization of all examples shown in this paper is automatically obtained using this method. One example of the initialization for the Bust sequence is shown on the left of Figure 10. We note that the initialization procedures proposed in [64, 65] cannot be applied here because of the big holes without 3D points, especially at the object bottom.

Also, all 3D points are rescaled into a $150 \times 150 \times 150$ voxel space in all examples by applying a similarity transformation. The resulting voxel size is of the same order of magnitude as the uncertainty median of the 3D points.

5.2.2 Description of different methods

The following surface evolution methods are tested and compared in Section 5.2.3.

BR3D is the bounded regularization method by taking the weighting function w to be the 3D distance from the set $\mathcal{P}$ of the reconstructed 3D points: $w(\mathbf{x}) = d(\mathbf{x}, \mathcal{P})$. The number of iterations is always 100 with $w_0 = 0.1$.

BR2D is the bounded regularization method by taking the weighting function w to be the image correlation function ρ. More details are given in Section 5.2.3.

BR3D+BR2D sequentially applies the BR3D and BR2D methods. Fifty iterations are used with BR2D.

BR3D2D uses the weighting function w as a combination of a 3D distance function and a 2D image consistency measure using a bounded regularization method: $w(\mathbf{x}) = d(\mathbf{x}, \mathcal{P}) + \epsilon e(\mathbf{x}, I)$, where $e = 0.2\sqrt{\sigma_r^2(\mathbf{x}) + \sigma_g^2(\mathbf{x}) + \sigma_b^2(\mathbf{x})}$ and $\sigma(\mathbf{x})$ is the standard deviation of the reprojected voxel in each of three color channels in $[0, 1]$. The consistency measure e is similar to the photo-consistency of the space-carving method. The basic idea is to avoid surface evolution in the immediate neighborhood of the reconstructed points where the surface previously obtained by BR3D is assumed to be correct. It also inflates the surface elsewhere and stops in the surface portions having inconsistent re-projections, mainly due to the difference between the object and the background colors. Thus, we use the following evolution equation

$$\frac{\partial u}{\partial t} = \nabla w \nabla u + min(w, w_{max})||\nabla u||_2(c + H),$$

where c is an inflating constant introduced and used in segmentation works [31, 5]. Note that the term $min(w, w_{max})||\nabla u||_2 c$ is negligible in areas where $w \approx 0$, i.e., in the close neighborhood of the reconstructed points. This is a much desired outcome. We choose

22

- $\epsilon = 0$ in the immediate neighborhood of reconstructed points $d(\mathbf{x}, \mathcal{P}) < 2\Delta x$ in the unit cube $[0, 1]^3$ and $\epsilon = 1$ elsewhere;

- $c = -5$ and $w_0 = 0.1$ with $u < 0$ inside the current surface $u = 0$.

BR3D+2D   sequentially applies the BR3D and BR3D2D methods. Fifty iterations are used with BR3D2D.

BR3DS   is a mixed method combining both 3D points and the silhouette information using a weighting function, $w(\mathbf{x}) = min(d(\mathbf{x}, \mathcal{P}), \epsilon + d(\mathbf{x}, \mathcal{S}))$. We choose $\epsilon = 2\Delta x$ to favor the 3D points, $\mathcal{P}$, over the visual hull, $\mathcal{S}$, in the immediate neighborhood of the reconstructed points. This method should be initialized by BR3D. Otherwise, the evolving surface may never reach the concave parts of the object.

BR3D+S   sequentially applies the BR3D and BR3DS methods. Fifty iterations are used with BR3DS. First, the surface is only attracted by 3D points including those of the object concavities. Second, the surface does not move in the immediate ϵ-neighborhoods of 3D points, but it moves toward the visual hull in the areas closer to the visual hull than the 3D points.

Freeze plane   To avoid the convergence of the dynamic surface to the empty surface, a freeze plane is often introduced to stop/freeze the surface evolution in one of the two delimited half spaces. The freeze plane is manually placed to fi ll in the biggest gap, often on the bottom or on the back of the object if the sequence is not complete.

5.2.3   Results, comparisons, and discussions

The reconstructed surfaces and experiments summary are shown in Figures 14 and 15 on many image sequences taken by a hand-held still digital camera, except for the Lady 1 sequence taken with a special device. Each row includes three images of the given sequence, which are followed by the reconstructed stereo points, and a Gouraud-shaded and a textured-mapped view of the surface, both from the same viewpoint.

BR3D vs. BR3D+2D   Combining 2D image information using BR3D2D can signifi cantly improve the fi nal reconstruction results as using only a 3D distance function may fail when there are no suffi - cient reconstructed points on some parts of the surface. This is illustrated in Fig. 9.

3D distance vs. image correlation   Using only image correlation as suggested in [9, 16] makes convergence very diffi cult for low-textured objects. Here we take a reasonably textured object, the
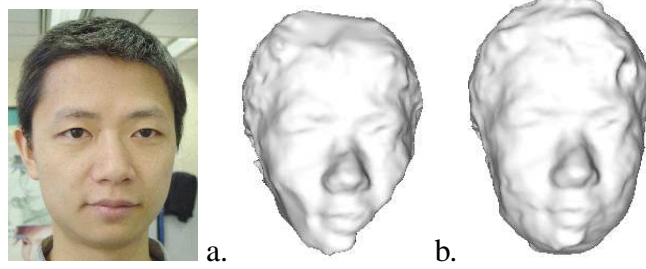
Figure 9: Surface geometry obtained by BR3D in (a) and BR3D+2D in (b). There are many missing 3D points in the low-textured cheeks (cf. Figure 14), so BR3D using only 3D information gives poor results while BR3D+2D gives good results by adding 2D information.
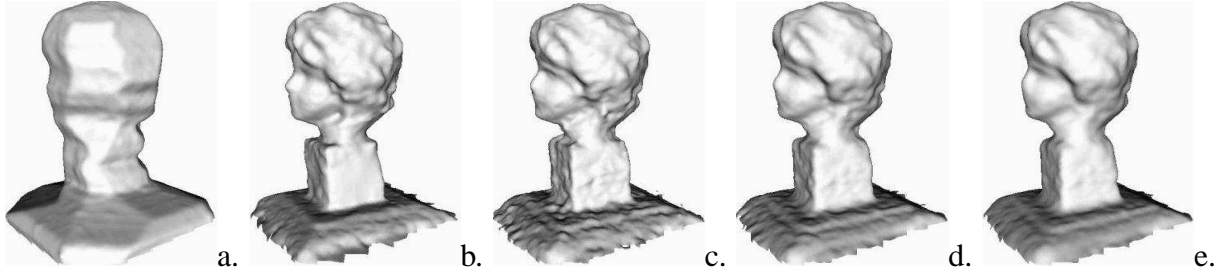

Figure 10: Surface computed using (a) initialization, (b) BR3D method, (c) BR3D+BR2D with $w_0 = 0.1$, (d) BR3D+BR2D with $w_0 = 0.5$, and (e) BR3D+BR2D with $w_0 = 1$.

bust, to test the BR2D method and compare it with the others. The surface initialization is shown on the left of Figure 10 and is obtained with the method described in Section 5.2.1. Figure 11 shows the results by the BR2D method with $w = 0.1(1 - \rho)$ and $\rho \in [-1, 1]$ for 400 and 1000 iterations with $w_0 = 0.1$ (left), $w_0 = 0.5$ (middle), and $w_0 = 1$ (right), using a 9×9 ZNCC-window. The lower bound $w_0 = 0.1$ gives a noisy surface (see the pyramid part). The upper bound $w_0 = 1$ gives a too smoothed surface (see the flat nose). The intermediate bound gives a compromise between the two. The original correlation [9, 16] with full regularization is even smoother than the upper bound $w_0 = 1$ case. Also, the convergence is extremely slow. It is still not done around the intersection of the concave part between the cube and the pyramid after 800 iterations. We have also found that the original correlation method is actually slower than BR2D, since its time step, Δt_{max}, is 380 smaller. We experimented with the two-step method, BR3D+BR2D. The results are shown in Figure 10. The results are similar to the previous case and not very satisfactory. However, this method is more efficient: the 100 steps of BR3D-iterations take only about 5 min. on a P4 2.4 GHz (including initialization), compared with the 20 (resp. 50) min. for 400 (resp. 1000) BR2D steps. Figure 12 shows the difference between the BR3D and BR3D+BR2D methods, with the best previous bound $w_0 = 0.5$ and only 50 iterations for BR2D. Still, the nose is too smooth for BR3D+2D and the chin is also degraded.

24

Isotropic vs. anisotropic smooth Using Laplacian/ isotropic $\triangle u$ instead of the curvature/anisotropic smooth $||\nabla u||_2 H = ||\nabla u||_2 div \frac{\nabla u}{||\nabla u||_2}$ leads to faster evolution, as the level-set function update, $||\nabla u||_2 = 1$, is done twice as frequently for the anisotropic smoothing than for the isotropic smoothing, which has a smaller discretization neighborhood. It is also important to observe that no apparent difference occurs between these two different smooths in the final surface geometry, as shown in Figure 12 (same conclusion with $w_0 = 0.5$ and $w_0 = 1$). This suggests that the benefit of using curvature-based smoothing is negligible for our context.

With vs. without silhouette Figure 13 shows results obtained by BR3D, BR3D+2D, BR3D+S and the pure silhouette method S for the Man 3 sequence. Using only 3D points by BR3D misses the low-textured cheeks, and using only the visual hull by S misses many important concavities on the surface, like in the areas of the ears and nose. Combining the two gives excellent final results.

Adding silhouette information improves the pure 3D results; both automatic and interactive extraction of silhouettes from unknown backgrounds have been used for different cases. Note that silhouette information is only optional in our approach and that the majority of our results presented here do not use it.
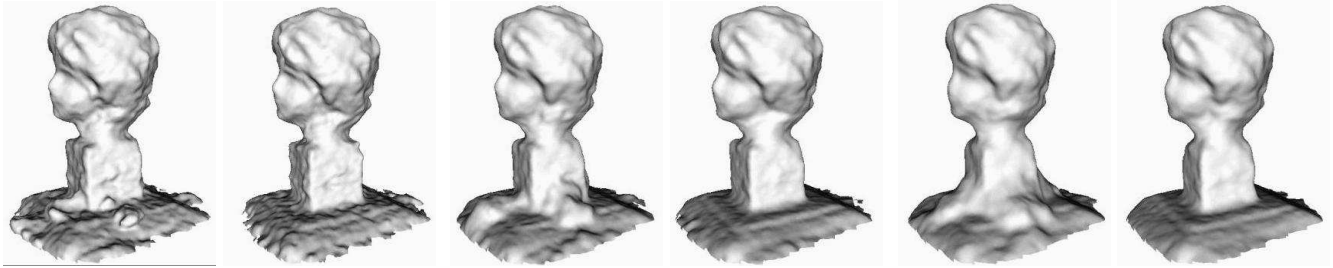

Figure 11: Surfaces obtained with the BR2D method after 400 and 1000 iterations with $w_0 = 0.1$ (the first two), $w_0 = 0.5$ (the middle two), and $w_0 = 1$ (the last two).
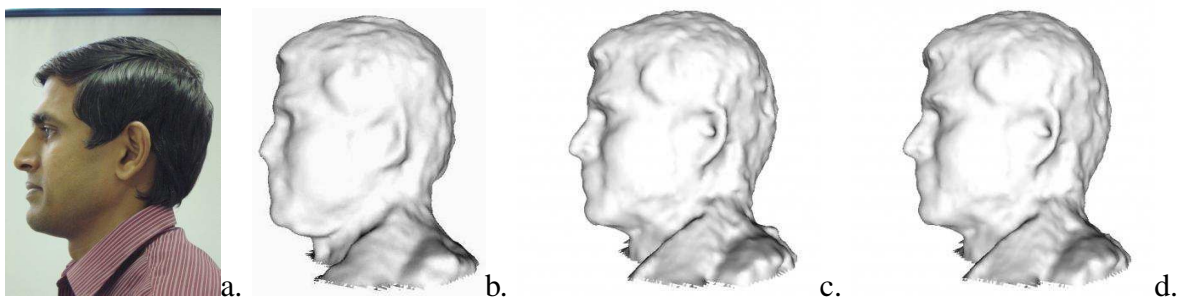

Figure 12: Surfaces computed using different smoothing methods. (a) One original image. (b) BR3D+BR2D with $w_0 = 0.5$. (c) Curvature-based smoothing BR3D. (d) Laplacian-based smoothing BR3D.
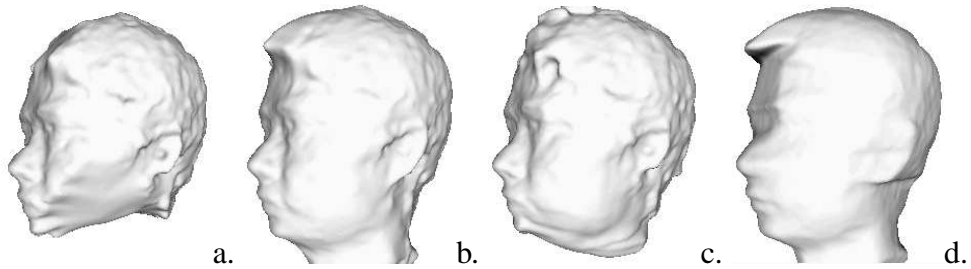
Figure 13: Surfaces computed using (a) BR3D method with only quasi-dense 3D points, (b) BR3D+S method with a combination of the quasi-dense 3D points and the silhouettes, (c) BR3D+2D method with a combination of the quasi-dense 3D points and the image photo-consistency, and (d) S method with only the silhouettes.

6 Conclusion

This paper describes a quasi-dense approach to practical surface model acquisition. In addition to presenting a complete system of 3D modeling from raw images captured from hand-held cameras, the main contributions of this paper are threefold: first, the introduction of new point features as the re-sampled points from the quasi-dense disparity map to densify the feature points to overcome the sparseness of the points of interest; second, an automatic quasi-dense geometry computation from uncalibrated images. For efficient evaluation of the reconstruction accuracy, we developed a fast gauge-free estimation algorithm. This quasi-dense based approach gives more robust and more accurate reconstruction results. It also works for largely separated images and requires fewer images than the standard sparse approach. It produces a high density of points that can be used for direct surface reconstruction. Third, new surface reconstruction algorithms integrating both 3D data points and 2D images. This is possible because of a unified functional based on a minimal surface formulation. We believe that the new functionals have far less fewer minima than those derived from 2D data alone and that this will result in more stable and more efficient algorithms. For the efficient evolution of surfaces, we also propose a bounded regularization method based on level-set methods. Its stability is also proved.

The methods have been intensively tested on many real sequences and very convincing results have been shown, including fully textured face models with hair. This is a significant practical advance as no other active or passive system that we are aware of can deliver full-head models with such a simple setup. However, the main limitation of our system is due to the choice of the surface evolution approach, which assumes a closed and smooth surface. The surface reconstruction module is not designed for outdoor or polyhedric objects.
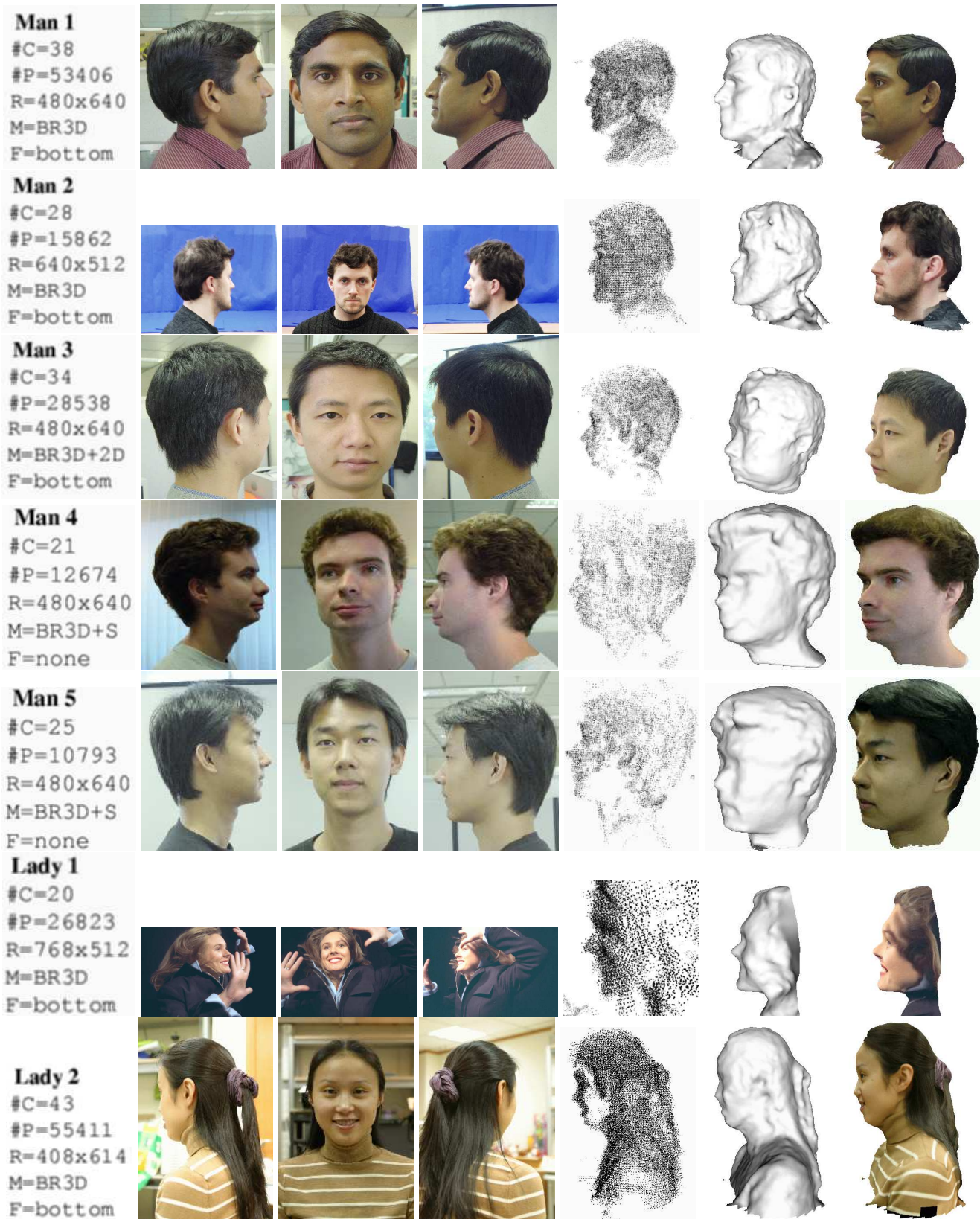
Figure 14: Each row illustrates one example of reconstruction by showing the details of the experiment, three frames of the sequence, the reconstructed quasi-dense 3D points, Gouraud-shaded surface geometry, and the textured-mapped surface geometry. In the details of the experiment, $\#C$ is the number of cameras, $\#P$ the number of points; R the image resolution, M the surface reconstruction method, and F the location of the freeze plane. Running times are about 5 and 3 minutes for BR3D(S) and BR3D2D with a P4 2.4GHz. Some of the data are available from the authors' websites.
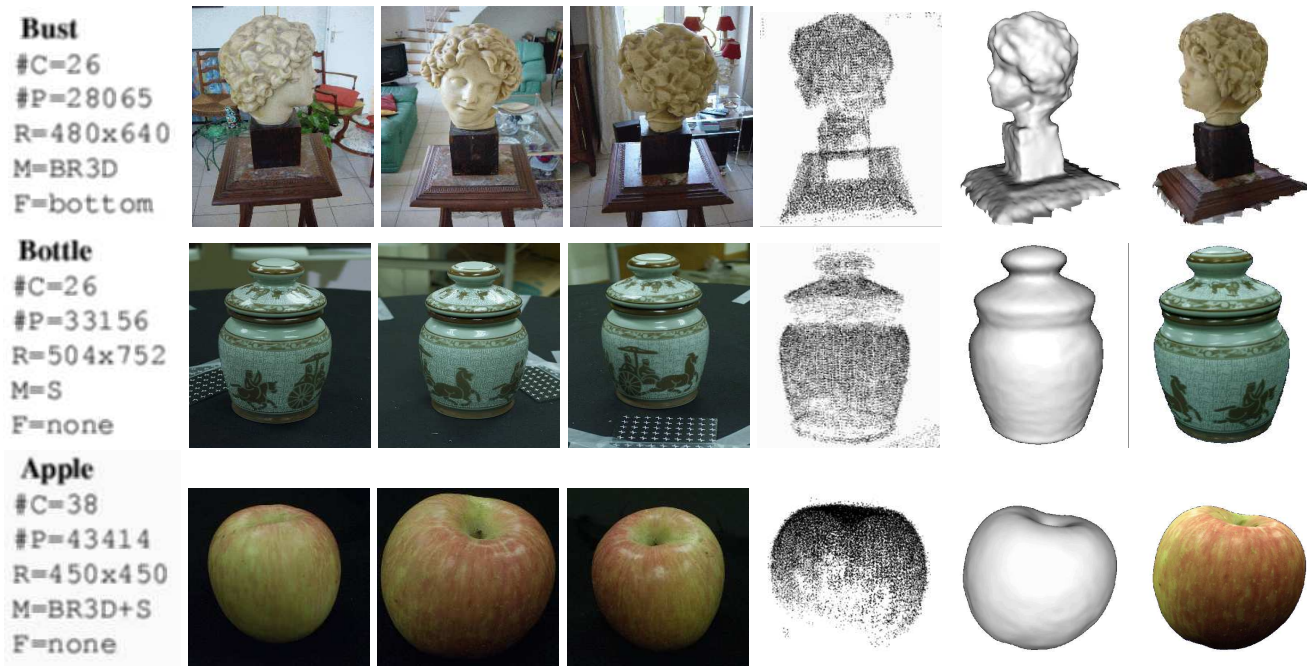
Bust
#C=26
#P=28065
R=480x640
M=BR3D
F=bottom

Bottle
#C=26
#P=33156
R=504x752
M=S
F=none

Apple
#C=38
#P=43414
R=450x450
M=BR3D+S
F=none

Figure 15: Same comments as in Figure 14.

Appendix

A   Optimal quasi-dense geometry estimation

A.1   Parametrization of the bundle optimization

The bundle adjustment as a global optimization procedure for either projective or Euclidean reconstruction is now a standard both in photogrammetry and computer vision, although the implementation strategy for the cost function, the parametrization and the optimization may vary a lot in exploiting the particular problem structure [58]. The cost function is always taken to be the quadratic of the total re-projection error with temporary outliers having infinite cost. The optimization procedure is the standard Levenberg-Marquardt method with a Gauss-Newton step [43]. The global parametrization is a gauge-free over-parametrization [19, 33]. The state vector is partitioned into camera and structure parameters. In the projective case, a camera is parametrized by 11 of its 12 homogeneous entries while locally normalizing its largest absolute value to 1. In the Euclidean case, each camera is parametrized by its six extrinsic parameters (camera center, t_x, t_y, t_z, and the three local Euler angles, $\delta x, \delta y, \delta z$). Also, all cameras are parametrized by their common focal length f and one optional common radial distortion ρ. For both projective and Euclidean cases, each 3D point is parametrized by three of their four homogeneous entries while normalizing its largest absolute value to one. In our quasi-dense reconstruction case, the points (typically more than twenty thousand) overwhelmingly outnumber the cameras. It is obvious that camera reduction by eliminating the structure parameters

28

suggested in photogrammetry [4, 48] and computer vision [19, 58, 20, 10] is a natural good choice for the implementation of both projective and Euclidean bundle adjustments.

A.2   Estimating a projective and Euclidean quasi-dense geometry

Projective quasi-dense three-views   The quasi-dense correspondences computed from a two-view geometry still contain outliers as the two-view geometry only defines a one-to-many geometric relationship. The three-view geometry [45, 10, 20] plays the most important role as it is the minimum number of images that have sufficient geometric constraints to resolve correspondence ambiguity.

Given the quasi-dense correspondences of all pairs, we proceed as follows to obtain the quasi-dense correspondences of all triplets of images.

1. Merge the two quasi-dense correspondences between the pair $i-1$ and i and the pair i and $i+1$ via the common ith frame as the set intersection to obtain initial quasi-dense correspondences of the image triplet.

2. Randomly draw six points to run RANSAC on the whole set of points. This further removes match outliers using re-projection errors of points. For six randomly selected points, compute the canonical projective structure of theses points and the camera matrices [45]. The other image points are reconstructed using the current camera matrices and reprojected back onto images to evaluate their consistency with the actual estimate.

3. Optimize the three-view geometry with all inliers of triplet correspondences by minimizing the re-projection errors of all image points by fixing one of the initial camera matrices.

Projective quasi-dense N-views   ¿From the quasi-dense correspondences of all pairs and triplets, we merge them into the whole sequence to create a consistent N-view quasi-dense geometry encoded by the projective structure. We essentially adapt the hierarchical merging strategy successfully used in [12], which is more efficient than an incremental merging strategy.

Given the quasi-dense correspondences of all pairs and triplets of images, the projective geometry of a longer sequence $[i..j]$ is obtained by merging two shorter sequences $[i..k+1]$ and $[k..j]$ with two overlapping frames, k and $k+1$, where k is the median of the index range $[i..j]$. The merger consists of

1. Merging the two quasi-dense correspondences between two sub-sequences using the two overlapping images.

2. Estimating the space homography between two common cameras using linear least squares.

3. Applying the space homography for one of the two sub-sequences to bring both of them into the same projective basis.

4. Optimizing the sequence $[i..j]$ with all merged corresponding points.

After this procedure, we optimize the whole sequence once more by enforcing the closure of the sequence if the sequence is closed [12] i.e., the next to the last frame of the sequence should be the first one. This reduces the accumulation errors caused by the sequential ordering of the frames. The final result of this step is a set of 3D points projectively consistent with all camera projection matrices on a common space projective basis.

Euclidean quasi-dense geometry   ¿From the quasi-dense correspondences encoded as a consistent projective structure of points and cameras, we use an auto-calibration method and an optimization method [4, 58] to compute the Euclidean coordinates of the set of quasi-dense points and all camera poses and intrinsic parameters as follows:

1. Initialize the Euclidean structure of the quasi-dense geometry by auto-calibration. We assume a constant unknown focal length as we rarely change the focal length for capturing the same object. Then, a one-dimensional exhaustive search of the focal lengths from a table of possible values is implemented for optimization of the focal length. The initial value for the focal length might be either computed by some linear auto-calibration methods [57, 40, 38] or obtained from the digital camera.

2. Transform the projective reconstruction by the estimated camera parameters to its metric representation.

3. Re-parametrize each Euclidean camera by its six individual extrinsic parameters and one common intrinsic focal length. This natural parametrization allows us to treat all cameras equally when estimating uncertainties, but it leaves the seven d.o.f scaled Euclidean transformation as the gauge freedom [58]. Finally, apply an Euclidean bundle adjustment over all cameras and all quasi-dense correspondence points.

4. Run a second Euclidean bundle adjustment by adding one radial distortion parameter for all cameras in the case where the non-linear distortions of cameras are non-negligible, for instance, for image sequences captured by a very short focal length. In practice, for an object of the size of a human face, this is unnecessary.

30

B Stability Proof of the Bounded Regularization Method

Assuming that the stability condition is the same for curvature-based and Laplacian-based regulariza-tions, the stability, $||u^{n+1}||_\infty \le ||u^n||_\infty$, is achieved if $\Delta t \le \Delta t_{max}$ with

$$\Delta t_{max} = \frac{\Delta x^2}{6w_{max} + ||\Delta x(|d^{0x}w| + |d^{0y}w| + |d^{0z}w|)||_\infty}.$$

This can be proved as follows. Let's assume that $w \ge 0, ||w||_\infty \ge w_{max} \ge 0$. Take $\Delta x = \Delta y = \Delta z$ as the space step, and Δt the time step. Also denote the centered (resp. forward and backward) differences for the x-axis at the 3D grid point ijk as d_{ijk}^{0x} (resp. $d_{ijk}^{+x}, d_{ijk}^{-x}$). Further, let $a_x = \frac{\Delta t}{\Delta x}d_{ijk}^{0x}w$ and $b = min(w_{ijk}, w_{max})\frac{\Delta t}{(\Delta x)^2}$ with similar notations, a_y, a_z.

We may choose the simplest up-wind discretization scheme for $\nabla w \nabla u$ and the centered dis-cretization scheme for $\triangle u$. Then, the equation $\frac{\partial u}{\partial t} = min(w, w_{max})\triangle u + \nabla w \nabla u$ is discretized by $u_{ijk}^{n+1} = u_{ijk}^n + \Delta t(l_{ijk}^n + c_{ijk}^{xn} + c_{ijk}^{yn} + c_{ijk}^{zn})$ such that

$$l_{ijk}^n\Delta t = b(u_{i+1jk}^n + u_{i-1jk}^n + u_{ij+1k}^n + u_{ij-1k}^n + u_{ijk+1}^n + u_{ijk-1}^n - 6u_{ijk}^n)$$

and $c_{ijk}^{xn} = d_{ijk}^{0x}w((d_{ijk}^{0x}w < 0)?d_{ijk}^{-x}u : d_{ijk}^{+x}u)$ with similar notations for c_{ijk}^{yn} and c_{ijk}^{zn}.

Now, we only need to prove for the case $0 \le a_x, a_y \le 0, 0 \le a_z$, as other cases can be performed in a similar manner. In this case, we have $u_{ijk}^{n+1} = u_{ijk}^n + a_x(u_{i+1jk}^n - u_{ijk}^n) + a_y(u_{ijk}^n - u_{ij-1k}^n) + a_z(u_{ijk+1}^n - u_{ijk}^n) + b(u_{i+1jk}^n + u_{i-1jk}^n + u_{ij+1k}^n + u_{ij-1k}^n + u_{ijk+1}^n + u_{ijk-1}^n - 6u_{ijk}^n)$. We notice that u_{ijk}^{n+1} is a weighted sum of $u_{i'j'k'}^n$ such that the sum of the weights is 1. It is also easy to check that all these weights are in $[0, 1]$ iff $|a_x| + |a_y| + |a_z| + 6b \le 1$. This condition is satisfied if $\Delta t \le \Delta t_{max}$. In this case, u_{ijk}^{n+1} is in the convex hull of $u_{i'j'k'}^n$. Thus, we have proved that $\forall ijk, |u_{ijk}^{n+1}| \le ||u^n||_\infty$.

References

[1] 2d3 Ltd. Boujou, 2000. http://www.2d3.com.

[2] P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In B. Bux-ton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 683–695. Springer-Verlag, April 1996.

[3] A.F. Bobick and S.S. Intille. Large Occlusion Stereo. *International Journal of Computer Vision*, 33(3):181–200, 1999.

[4] D.C. Brown. The bundle adjustment – progress and prospects. *International Archive of Photogrammetry*, 21, 1976. Update of "Evolution, Application and Potential of the Bundle Method of Photogrammetric Triangulation", ISP Symposium Commission III, Stuttgart, 1974.

[5] V. Caselles, R. Kimmel. Minimal surfaces based object segmentation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(4):394–398, 1997.

[6] V. Caselles, R. Kimmel and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[7] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of* SIGGRAPH*, New Orleans, LA*, pages 303–312, 1996.

[8] F. Devernay and O.D. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *Proceedings of the* SPIE *Conference on Investigate and Trial Image Processing, San Diego, California, USA*, volume 2567. SPIE - Society of Photo-Optical Instrumentation Engineers, July 1995.

[9] O. Faugeras and R. Keriven. Complete dense stereovision using level set methods. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, pages 379–393, 1998.

[10] O. Faugeras and Q.T. Luong. *The Geometry of Multiple Images*. The MIT Press, 2001.

[11] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381 – 395, June 1981.

[12] A.W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *European Conference on Computer Vision*, pages 311–326, june 1998.

[13] P. Fua. Parametric models are versatile: The case of model based optimization. In *ISPRS WG III/2 Joint Workshop, Stockholm, Sweden*, September 1995.

[14] P. Fua. From multiple stereo views to multiple 3d surfaces. *International Journal of Computer Vision*, 24(1):19–35, 1997.

[15] S. Gibson, J. Cook, T. Howard, R. Hubbold and D. Oram. Accurate camera calibration for off-line, video-based augmented reality. In *IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 37-46, 2002.

[16] J. Gomes and O. Faugeras. Reconciling Distance Functions and Level Sets. *Journal of Visual Communication and Image Representation*, 11:209-223, 2000.

[17] B. Guenter, C. Crimm, D. Wood and H. Malvar. Making faces. In SIGGRAPH *1998*, pages 55–66, 1998.

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

[19] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the* DARPA−ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 187–202, October 1993.

[20] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.

[21] A. Heyden. *Geometry and Algebra of Multiple Projective Transformations*. PhD thesis, Lund Institute of Technology, 1995.

[22] H. Hoppe, T. Derose, T. Duchamp, J. McDonalt, and W. Stuetzle. Surface reconstruction from unorganized points. *Computer Graphics*, 26:71–77, 1992.

[23] S. Kichenassamy, A. Kumar, P. Olver and A. Tannenbaum. Gradient fbws and geometric active contour models. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, June 1995.

[24] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, 2002.

[25] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, volume 1, pages 307–314, 1999.

[26] S. Laveau. *Géométrie d'un système de N caméras. Théorie, estimation, et applications*. PhD thesis, École Polytechnique, May 1996.

[27] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1140–1146, 2002.

[28] M. Lhuillier and L. Quan. Quasi-dense reconstruction from image sequence. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 2, pages 125–139, 2002.

[29] M. Lhuillier and L. Quan. Surface reconstruction by integrating 3d and 2d data of multiple views. In *Proceedings of the 9th International Conference on Computer Vision, Nice, France*, 2003.

[30] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981.

[31] R. Malladi, J.A. Sethian and B.C. Vemuri. Shape modeling with front propagation: A level set approach. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.

[32] J. Matas, O. Chum, M. Urban and T. Padjdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *Proceedings of the 13th British Machine Vision Conference*, University of Cardiff, September 2002.

[33] P. F. McLauchlan. Gauge independence in optimization algorithms for 3D vision. In *Proceedings of the Vision Algorithms Workshop, Dublin, Ireland*, 2000.

[34] D.D. Morris. *Gauge Freedoms and Uncertainty Modeling for 3D Computer Vision*. PhD thesis, Carnegie Mellon University - The Robotics Institute, March 2001.

[35] D.D. Morris, K. Kanatani and T. Kanade. Uncertainty modeling for optimal structure from motion. In *Vision Algorithms : Theory and Practice*, Workshop held during ICCV'99 in Kerkyra (Corfu), Greece, September 1999.

[36] P.J. Narayanan, P.W. Rander, and T. Kanade. Constructing virtual worlds using dense stereo. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, pages 3–10, 1998.

[37] D. Nister. Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. In *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland*, pages 649–663, 2000.

[38] D. Nister. *Automatic Dense Reconstruction from Uncalibrated Video Sequences*. PhD thesis, Ericsson and University of Stockholms, March 2001.

[39] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[40] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 90–95, January 1998.

[41] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Metric 3d surface reconstruction from uncalibrated image sequences. In R. Koch and L. Van Gool, editors, *European Workshop, SMILE*, pages 139–154. Springer-Verlag, 1998.

[42] M. Pollefeys, F. Verbiest, and L. Van Gool. Surviving dominant planes in uncalibrated structure and motion recovery. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 2350 of *Lecture Notes in Computer Science*, Springer-Verlag, May 2002.

[43] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.

[44] P. Pritchett and A. Zisserman. Wide Baseline Stereo Matching. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 754–760, 1998.

[45] L. Quan. Invariants of six points and points projective reconstruction from three uncalibrated images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(1):34–46, 1995.

[46] S.M. Seitz and C.R. Dyer. Photo-realistic scene reconstruction by voxel coloring. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1067–1073. IEEE Computer Society Press, June 1997.

[47] J.A. Sethian. *Level Set Methods and Fast Marching Methods*. The CUP Press, 1999.

[48] C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[49] C. Strecha, T. Tuytelaars and L. Van Gool. Dense matching of Multiple Wide-Baseline Views. *International Conference on Computer Vision*, pages 1194-1201, 2003.

[50] R. Szeliski. Rapid octree construction from image sequences. *Computer Vision, Graphics and Image Processing*, 58(1):23–32, July 1993.

[51] R. Szeliski and R. Weiss. Robust shape recovery from occluding contours using a linear smoother. Technical report, Digital Equipment Corporation, Cambridge Research Lab, December 1993.

[52] C.K. Tang and G. Medioni. Curvature-augmented tensor voting for shape inference from noisy 3d data. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 24(6):858–864, 2002.

[53] B. Tell and S. Carlsson. Wide baseline point matching using affine invariants computed from intensity profiles. In *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland*, pages 814–828, 2000.

[54] P.H.S. Torr and D.W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[55] P.H.S. Torr and A. Zisserman. Robust Parameterization and Computation of the Trifocal Tensor. *Image and Vision Computing*, 15:591–607, 1997.

[56] P.H.S. Torr and A. Criminisi. Dense Stereo using Pivoted Dynamic Programming. In *Proceedings of the British Machine Vision Conference*, 414-423, 2002.

[57] B. Triggs. Auto-calibration and the Absolute Quadric. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 609–614, June 1997.

[58] B. Triggs, P.F. McLauchlan, R.I. Hartley, and A. Fitzgibbon. Bundle adjustment — a modern synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag, 2000.

[59] T. Tuytelaars and L. Van Gool. Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions. In *Proceedings of the 11th British Machine Vision Conference*, Bristol, September 2000.

[60] R. Whitaker. A level-set approach to 3d reconstruction from range data. *International Journal of Computer Vision*, 29(3):203–231, 1998.

[61] Z. Zhang and Y. Shan. A progressive scheme for stereo matching. *SMILE'2, ECCV'00 Workshop*.

[62] Z. Zhang, Z. Liu, D. Adler and M.F. Cohen. Robust and rapid generation of animated faces from video images: a model-based modeling approach. Technical Report MSR-TR-2001-101, Microsoft Research, 2001.

[63] 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. *Artificial Intelligence*, 78:87–119, 1995.

[64] H.K. Zhao, S. Osher, B. Merrigan and M. Kang. Implicit and non-parametric shape reconstruction from unorganized points using variational level set method. *Computer Vision and Image Understanding*, 80:295–319, 2000.

[65] H.K. Zhao, S. Osher and R. Fedkiw. Fast surface reconstruction using the level set method. In *IEEE Workshop on Variational and Level Set Methods in Computer Vision*, July 2001.