

Quasi-Dense Reconstruction from Image Sequence

Maxime Lhuillier and Long Quan

Department of Computer Science
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong SAR

Abstract. This paper proposes a quasi-dense reconstruction from uncalibrated sequence. The main innovation is that all geometry is computed based on re-sampled quasi-dense correspondences rather than the standard sparse points of interest. It not only produces more accurate and robust reconstruction due to highly redundant and well spread input data, but also fills the gap of insufficiency of sparse reconstruction for visualization application. The computational engine is the quasi-dense 2-view and the quasi-dense 3-view algorithms developed in this paper. Experiments on real sequences demonstrate the superior performance of quasi-dense w.r.t. sparse reconstruction both in accuracy and robustness.

1 Introduction

3D reconstruction from uncalibrated sequences has been very active and successful in the past decade in computer vision. This is mainly due to the intrinsic formulation of geometric constraints in projective geometry and a better understanding of numerical and statistical properties of geometric estimation [19,42]. Many reconstruction algorithms based on point features have been published for short [5,13,16,6] or long sequences [38,36]. Almost all of these approaches have been based on sparse points of interests. More recent and complete systems based on these ideas are reported in [28,9,31,1,21] without any prior camera calibration or position information. Unfortunately, most modeling and visualization applications need dense or quasi-dense reconstructions rather than a sparse point clouds. Traditional dense stereo methods are limited to specific pre-calibrated camera geometries and closely spaced viewpoints [37,29,18,17]. Traditional dense stereo/motion analysis is not yet efficient and robust enough to be integrated into an on-line dense reconstruction to handle images captured by hand-held cameras. It should be noted that although the final results reported in [30] showed densely textured models, the method only applied the dense stereo reconstruction using an area-based algorithm after obtaining the geometry by a sparse method.

We propose to develop an intermediate approach to fill the gap between sparse and dense reconstruction methods for hand-held cameras. By quasi-dense reconstruction we mean that the geometry is directly computed on re-sampled points from quasi-dense pixel correspondences, rather than reconstructions of sparse points of interest. Quasi-dense correspondences are preferable to fully

dense ones owing to their greater robustness and efficiency with hand-held captured images. The most innovative part is that all geometry is computed with a simple and efficient quasi-dense correspondence algorithm proposed in [22,23,24] for image-based rendering applications. Quasi-dense correspondence has been integrated at the earliest stage from the building blocs of 2-view geometry and 3-view geometry up to the final sub-sequence merging. This not only gives object/scene reconstructions more suitable for visualization application, but also results in more accurate and robust estimation of camera and structure.

2 Review of Quasi-Dense Matching

The construction of quasi-dense matching map starts from matching some points of interest that have the highest “textureness” as seed points. This bootstraps a region growing algorithm to propagate the matches in its neighborhood from the most textured (therefore most reliable) pixels to less textured ones [22,24].

The algorithm can therefore be described in two steps: *Seed selection* and *Propagation*, which are illustrated in Figure 1.

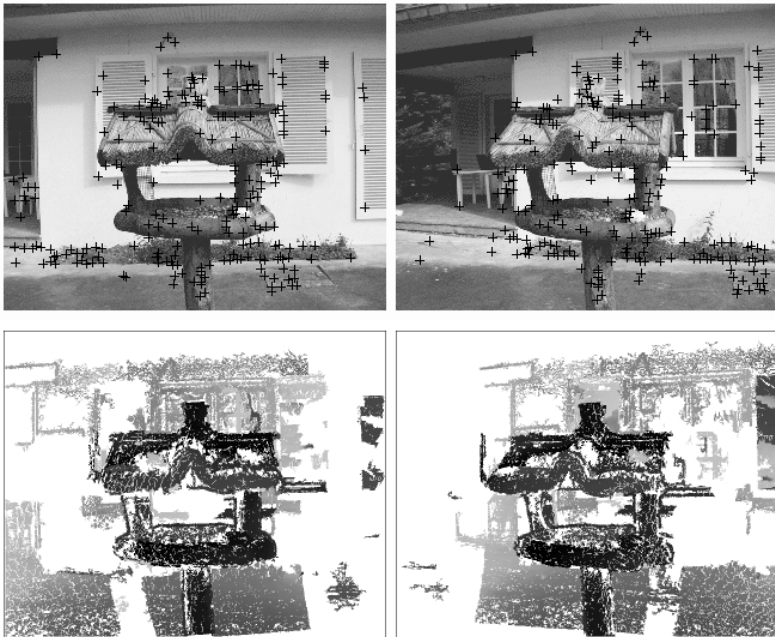


Fig. 1. Top: initial seed matches for two consecutive images of the Garden-cage sequence with big disparities (some seeds are bad mainly due to the shutter periodic textures). Bottom: the resulting propagation without the epipolar constraint.

Points of interest [25,12] are naturally good seed point candidates, as points of interest are by its very definition image points which have the highest tex-

tureness, i.e. the local maxima of the auto-correlation function of the signal. We first extract points of interest from two original images, then a ZNCC correlation method is used to match the points of interest across two images, followed by a cross validation for the pair of images. This gives the initial list of seed correspondences sorted by the correlation score.

At each step of the propagation, the match $(\mathbf{x}, \mathbf{x}')$ composed of two corresponding pixels \mathbf{x}, \mathbf{x}' with the best ZNCC score is removed from the current list of seed matches. The ZNCC is still used as it is more conservative than others such as sum of absolute or square differences in uniform regions, and is more tolerant in textured areas where noise might be important. Then we look for new potential matches $(\mathbf{u}, \mathbf{u}')$ in their immediate spatial neighborhood $\mathcal{N}(\mathbf{x}, \mathbf{x}')$. This neighborhood enforces a disparity gradient limit of 1 pixel in both image dimensions $\|(\mathbf{u}' - \mathbf{u}) - (\mathbf{x}' - \mathbf{x})\|_\infty \leq 1$ to deal with inaccurate or non available epipolar constraint. The matching uniqueness and the ending of the process are guaranteed by choosing only new matches $(\mathbf{u}, \mathbf{u}')$ that have not yet been selected.

The time complexity of this propagation algorithm is $O(n \log(n))$, only dependent of the number of final matches n , and the space complexity is linear in the image size. Both complexities are independent of disparity bound. Notice that 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 [43,39] for matching points of interest using correlation, but the crucial difference is that we need only to take the most reliable ones rather than trying to match a maximum of them. In some extreme cases, only one good seed match is sufficient to provoke an avalanche of the whole 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.

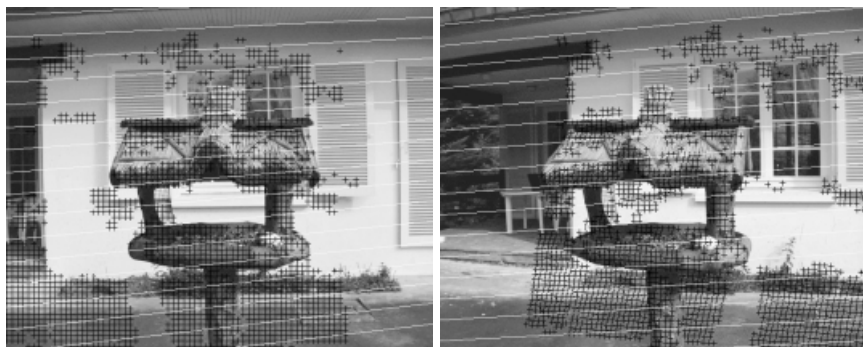


Fig. 2. The re-sampled matches from the propagation (described in Section 4) is represented as a set of matched black crosses in both images. These well spread matches are used to fit the fundamental matrix, shown as white epipolar lines.

3 Re-sampling

The matching map obtained from the propagation may still be corrupted and irregular. We assume that the scene surface is locally smooth enough to be approximated by small planar patches. Thus, the matching map can be regularized by locally fitting planar patches encoded by homographies. The first image is initially subdivided into small regular grid. For each square patch, we obtain all matched points of the square from the quasi-dense matching map. A plane homography should be tentatively fitted to these matched points of the square to look for potential planar patches. The RANdom SAMple Consensus (RANSAC) method [7] is used for robust estimation. In practice, the stability of the homography fitting decreases with the patch size. Our compromise between patch grid resolution and stability fitting is to fit a planar affine application (which counts only 6 d.o.f instead of 8 d.o.f of homography) in 8×8 -pixel squares.

The result is a list of matches shown by crosses in Figure 2, which is better spread in image space than the usual list of matched interest points shown at the top of Figure 1.

4 Estimating 2-View Geometry

The 2-view geometry of a rigid scene is entirely encoded by the fundamental matrix. The actual standard approach is to compute automatically fundamental matrix and correspondences from sparse points of interest [43,39] within a random sampling framework. There are also attempts of integrating the dense correspondence into the non-linear optimization of the fundamental matrix starting from an initial sparse solution by optimizing a global correlation score [11], but the algorithm is very slow in computation time (7-12 minutes vs. 20-40 seconds for the method to be proposed here for images of size 512×512 and similar processors) and fragile to handle occlusion for widely separated images as those in the image pair shown in Figure 2. In the context of our quasi-dense matching algorithm, we have two choices of integrating the geometry estimation into the match propagation algorithm for quasi-dense matching. The first is an epipolar constrained propagation which grows only those satisfying the epipolar constraint, while the second is an unconstrained one. The advantage of constrained propagation is that the bad propagation might be stopped earlier, but the domain of propagation might be reduced. Even more seriously, the geometry estimated with a robust method often tends to be locally fitted to a subset of images. We therefore prefer a strategy of an unconstrained propagation followed by a more robust constrained propagation as follows:

1. Detect points of interest in two images and compute the first correspondences by correlation and bidirectional consistency [10].
2. Run an unconstrained propagation.
3. Re-sample the obtained quasi-dense correspondences using a regular sampling grid in one image, and deduce the corresponding re-sampled points in the other image using the estimated local homographies.

4. Match the detected points of interest using the estimated local homographies, and add these matches in the list of re-sampled quasi-dense correspondences.
5. Estimate the fundamental matrix \mathbf{F} using a standard robust algorithm [39, 43,15] on re-sampled quasi-dense correspondences.
6. Run an epipolar constrained propagation by \mathbf{F} .
7. Again re-sample the quasi-dense matches from the constrained propagation and again add matched points of interest.
8. Re-estimate the fundamental matrix \mathbf{F} using the re-sampled quasi-dense correspondences.

5 Estimating 3-View Geometry

The 3-view geometry plays a central role for construction of longer sequences as 3 views is the maximum number of images which can be solved in closed-form, but also it is the minimum number of images which has sufficient geometric constraints to remove match ambiguity. The projective reconstruction from a minimum of 6 points in 3 views is therefore the basic computational engine for 3-view geometry both for robust assessment of correspondences using RANSAC and for optimal bundle adjustment of the final solution [33,34,40,15,35].

The quasi-dense 3-view algorithm can be summarized as follows.

1. Apply the previous quasi-dense 2-view algorithm to the pair i and $i - 1$ and the pair i and $i + 1$.
2. Merge the two re-sampled quasi-dense correspondences between the pair $i - 1$ and i and the pair i and $i + 1$ via the common i th frame as the set intersection to obtain an initial re-sampled quasi-dense correspondences of the image triplet.
3. Randomly draw 6 points to run RANSAC to remove match outliers using re-projection errors of points. For 6 randomly selected points, compute the canonical projective structure of these points and the camera matrices using the closed-form 6-point algorithm [33].
The other image points are reconstructed using the current camera matrices and re-projected back onto images to evaluate their consistency with the actual estimate.
4. Bundle adjust 3-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.

The general philosophy of exploiting strong 3-view geometry for long sequence reconstruction is the same as the previous methods [9,21,40], but it differs from [9] in the following aspects:

- We do not transfer point pairs for guided matching. The 3-view geometry only assesses inliers and outliers from the common re-sampled quasi-dense correspondences, it is therefore fast as the percentage of outliers is small.

- We do not use trifocal tensor parametrization of 3-view geometry as suggested in [9,1,15,35]. We use the P-matrix representation directly from the canonical projective structure of 6 points to reconstruct other points and evaluate their re-projection errors for assessing inliers/outliers of correspondences by RANSAC.
Tensor parametrization is hardly justified here as it gives a rather complicated over-parametrization of the 3-view geometry, more sophisticated numerical algorithms are necessary for its estimation. The transfer error tends to accept points which are large outliers to the re-projection error from the optimal estimate [9,8]. Tensor might be useful for guided matching [8], but is unnecessary in our case.
- We use the closed-form 6-point algorithm [33] rather than more recent methods proposed in [40,15,34] as the initial solution for robust search and optimization. It is direct and fast without any SVD computation compared with the algorithm [15] that we have also implemented and tested. The improvement provided by Schaffalitzky et al. [34] is necessary only when redundant data has to be handled.

6 Merging Pairs and Triplets into Sequences

From pairs and triplets to sequences, we essentially adapt the hierarchical merging strategy successfully used in [9,21] which is more efficient than an incremental merging strategy.

The general hierarchical N-view algorithm can be summarized as:

1. For each pair of consecutive images in the sequence, apply the quasi-dense 2-view algorithm described in Section 4.
2. For each triplet of consecutive images in the sequence, apply the quasi-dense 3-view algorithm described in Section 5.
3. Apply a hierarchical merging algorithm of sub-sequences. 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 merge consists of
 - a) Merging the two re-sampled quasi-dense correspondences between two sub-sequences using the 2 overlapping images.
 - b) Estimating the space homography between two common cameras using linear least squares.
 - c) Apply the space homography for all camera matrices and all points not common in the two sub-sequences.
 - d) Bundle adjust the sequence $[i..j]$ with all merged corresponding points.

In [9], several algorithms have been proposed to merge two triplets with 0, 1 or 2 overlapping views. The main advantage of imposing two-view overlapping is that camera matrices are sufficient for estimating the space homography to merge two reconstructions without any additional point correspondences between the two. It is also important to notice that both re-sampled quasi-dense points from 3-view geometry and sparse points of interest are contributing to the merging and optimization steps.

7 Optimal Euclidian Estimation of Reconstruction

The final step is to upgrade the projective reconstruction into a metric representation using self-calibration and optimal estimates of the metric representation.

- A linear solution [30] based on the parametrization of the dual of the absolute conic [41] is used for estimating constant but unknown focal lengths while assuming the other intrinsic camera parameters, such as principal point and aspect ration, are given. If the algorithm fails, we simply perform a one-dimensional exhaustive search of the focal lengths from a table of possible values.
- Transform the projective reconstruction by the estimated camera parameters to its metric representation. The metric reconstruction coordinate system is those of the camera in the middle of the entire sequence and the scale unit is the maximum distance between any pairs of camera positions.
- Re-parametrize each Euclidian camera by its 6 individual extrinsic parameters and one common intrinsic focal length. This natural parametrization allows us to treat all cameras equally when estimating uncertainties, but leaves the 7 d.o.f scaled Euclidian transformation gauge freedom [42,27,3]. Finally, apply an Euclidian bundle adjustment over all cameras and all quasi-dense points.
- A second Euclidian bundle adjustment by adding one radial distortion parameter for all cameras is carried out in the case where the non-linear distortions of cameras are non-negligibles, for instance, for image sequences captured by a very short focal length.

It is obvious that the sparse structure of the underlying numerical system as suggested in photogrammetry [2,26] and vision [14,42,15] has to be exploited for the implementations of both projective and Euclidian bundle adjustments as we are routinely handling at least 10 thousand 3D points. It is also natural to use reduced camera subsystem by eliminating the structure parameters.

8 Comparative Experiments

This section demonstrates the accuracy and robustness of the quasi-dense reconstruction method (QUASI) by comparing it with the standard sparse methods (SPARSE). We will use two sparse reconstruction algorithms based only on points of interest. The first consists of simply tracking all points of interest detected in each individual image. The second is a mixture of sparse and quasi-dense: it consists of assessing points of interest from individual images by geometry that is computed from quasi-dense algorithm, and to re-evaluate the complete geometry only from these matched interest points. In the following, it is meant by “SPARSE” the best result of these two methods.

To measure the reconstruction accuracy, we may consider the bundle adjustment as the maximum likelihood estimates of both camera and scene structure geometry, if we admit that the image points are normally distributed around

their true locations with an unknown standard deviation σ . This assumption is reasonable both from theoretical and practical point of view [20]. The confidence regions for a given probability can therefore be computed from the covariance matrix of the estimated parameters.

The covariance matrix is only defined up to the choice of gauge [42,27,3] 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 reprojection errors, d is the number of independent parameters of the minimization $d = 1 + 6c + 3p - 7$ (1 counts for the common focal length, c is the number of cameras, p is the number of reconstructed points and 7 is the gauge freedom choice). All results given here are gauge free: the covariance is computed without imposing gauge constraints, now 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 obtain the same conclusions for the comparisons between SPARSE and QUASI with a camera-centered gauge by fixing orientation and position of the middle camera (especially, the uncertainty σ_f is the same for all gauge choices since f is gauge invariant). Since the full covariance matrix is very big, only its diagonal blocs for cameras and points are computed using sparse pseudo-inversion method [3,15].



Fig. 3. A synthetic scene composed of two spline surfaces and a very distant plane with three textures mapped on it: random textured scene (left), indoor textured scene (middle) and outdoor textured scene (right).

We choose a 90% confidence ellipsoid for any 3D position vector: if \mathbf{C} is a 3×3 covariance sub-matrix of any camera position or point extracted from the full covariance matrix of all parameters, the confidence ellipsoid is therefore defined by $\Delta \mathbf{x}^T \mathbf{C}^{-1} \Delta \mathbf{x} \leq 6.25$, i.e. a 90% probability for a chi-square distribution with 3 degrees of freedom [32]. The maximum of semi-axes of 90% confidence ellipsoid is computed as the uncertainty bound for each 3D position. As the number of cameras is moderate, we only use the mean of all uncertainty bounds of camera positions $\bar{\mathbf{x}}_{c_i}$ to characterize the camera uncertainty. The number of points is however quite consequent, particularly for the QUASI method. To have a better characterization of their uncertainties, we compute 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 to assess the uncertainty of the reconstructed points. The uncertainty of the focal length f is given by the standard deviation σ_f .

8.1 Synthetic Examples

First we experiment on a synthetic scene with two spline surfaces and a very distant plane with three textures and uneven spread of points of interest: a well textured random scene, a low textured indoor scene and an outdoor scene of low texture, as illustrated in Figure 3. The camera moves around a vertical axis at the middle of the scene by 5 degrees 5 times, and the image size is 256×256 .

Table 1. Uncertainty measures for the synthetic scene. The right column are the accuracy of camera centers w.r.t ground truth

| | | | | | | | | | | |
|---------|------------|----------|-----|--------------------------|----------------|----------------------------|----------------------------|----------------------------|----------------|------------------------|
| Random | #3D points | σ | f | $\bar{\mathbf{x}}_{c_i}$ | \mathbf{x}_0 | $\mathbf{x}_{\frac{1}{4}}$ | $\mathbf{x}_{\frac{1}{2}}$ | $\mathbf{x}_{\frac{3}{4}}$ | \mathbf{x}_1 | $\ \mathbf{x}_{c_i}\ $ |
| QUASI | 2559 | .25 | 256 | 8.8e-4 | .011 | .031 | .056 | .96 | 1.6 | 9.5e-4 |
| SPARSE | 126 | .45 | 256 | 6.0e-3 | .065 | .11 | .15 | .19 | 4.7 | 3.1e-3 |
| Indoor | #3D points | σ | f | $\bar{\mathbf{x}}_{c_i}$ | \mathbf{x}_0 | $\mathbf{x}_{\frac{1}{4}}$ | $\mathbf{x}_{\frac{1}{2}}$ | $\mathbf{x}_{\frac{3}{4}}$ | \mathbf{x}_1 | $\ \mathbf{x}_{c_i}\ $ |
| QUASI | 1459 | .36 | 256 | 1.8e-3 | .022 | .046 | .076 | 0.14 | 5.0 | 1.6e-3 |
| SPARSE | 114 | .42 | 256 | 4.8e-3 | .055 | 0.61 | 0.69 | 0.11 | 1.6 | 3.8e-3 |
| Outdoor | #3D points | σ | f | $\bar{\mathbf{x}}_{c_i}$ | \mathbf{x}_0 | $\mathbf{x}_{\frac{1}{4}}$ | $\mathbf{x}_{\frac{1}{2}}$ | $\mathbf{x}_{\frac{3}{4}}$ | \mathbf{x}_1 | $\ \mathbf{x}_{c_i}\ $ |
| QUASI | 1547 | .34 | 256 | 1.6e-3 | .019 | .041 | .071 | 0.11 | 2.9 | 2.0e-3 |
| SPARSE | 66 | .46 | 256 | 7.3e-3 | .070 | 0.93 | 0.15 | 0.21 | 3.0 | 6.8e-3 |

The computed uncertainty measures are shown in Table 1. With 10 to 20 times more points, the QUASI uncertainties are usually 2 to 5 times smaller. As expected, the points on the textured and distant plane are very uncertain in comparison with the others. In this particular case of synthetic scene, all intrinsic parameters (including the known focal length f) are enforced by the final Euclidian bundle adjustment. Furthermore, the true camera motion is known: we compute the accuracy of the movement $\|\mathbf{x}_{c_i}\|$ as the mean of the Euclidian distance between the estimated and the true centers of cameras. The QUASI accuracy is usually 3 times better than the SPARSE one. These conclusions are the same if the focal length f is estimated in the final Euclidian bundle, with a better f accuracy for QUASI.

We have also experimented with other synthetic examples where the set of matched interest points is well spread in image space, and have found in these last cases that the accuracies are usually better for SPARSE than for QUASI, although uncertainties are usually better for QUASI than for SPARSE.

8.2 Real Examples

We also give detailed experimental results on three real sequences. The Corridor sequence (11 images at resolution 512×512) has a forward motion along the scene which does not provide strong geometry, but favors the SPARSE method as it is a low textured polyhedral scene, points of interest are abundant and well spread over the scene. The Lady sequence (20 images at 768×512) has a more favorable lateral motion in close-range. The Garden-cage sequence (34 images



Fig. 4. From left to right: Corridor (11 images at 512×512 resolution), Lady (20 images at 768×512), Garden-cage (34 images at 640×512) sequences.

at 640×512) are captured by a hand-held still camera (Olympus C2500L) with an irregular but complete inward walk around the object.

Table 2. 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.

| Corridor | #3D points | σ | f | σ_f | \bar{x}_{c_i} | \mathbf{x}_0 | $\mathbf{x}_{\frac{1}{4}}$ | $\mathbf{x}_{\frac{1}{2}}$ | $\mathbf{x}_{\frac{3}{4}}$ | \mathbf{x}_1 |
|----------|------------|----------|-----|------------|-----------------|----------------|----------------------------|----------------------------|----------------------------|----------------|
| QUASI | 16976 | 0.41 | 714 | 4.36 | $7.0e-4$ | .014 | .070 | .13 | .38 | 15700 |
| SPARSE | 427 | 0.52 | 761 | 17.3 | $1.7e-3$ | .016 | .056 | .12 | .32 | 106 |

Corridor. Table 2 shows the comparative uncertainty measures for the Corridor sequence. With almost 40 times redundancy, camera position (resp. focal length) uncertainties from QUASI are two times (resp. four times) smaller than those from SPARSE. However, the point uncertainties for SPARSE are slightly better than those of QUASI for the majority of points. As the camera direction and path is almost aligned with the scene points, the 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. Figure 5 shows the reconstruction results in which each 3D points is displayed as a small texture square around it, and illustrates a plane view of the 90% confidence ellipsoids.

Lady. For the Lady sequence, we show the results obtained from SPARSE and QUASI methods in Table 3 and Figure 6. The uncertainties for QUASI are smaller than for SPARSE, 6 times smaller for focal length and camera positions. We have noticed that the very small number of 3D points makes the SPARSE method fragile.

Garden-cage. The Garden-cage sequence is particularly difficult as it is consisting of a close-up bird cage and background houses and trees. The viewing field is therefore very profound. SPARSE methods failed because some triplets of consecutive images do not have sufficient matched interest points. The QUASI

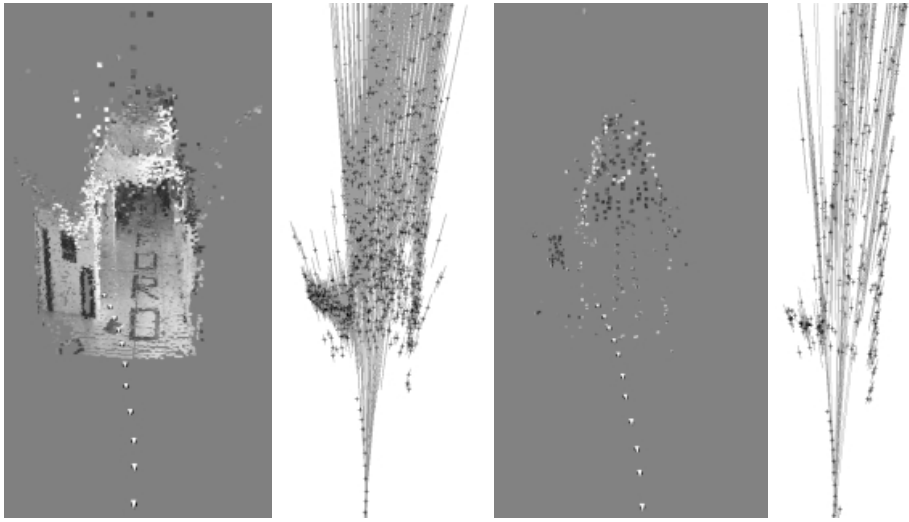


Fig. 5. QUASI (left) and SPARSE (right) reconstructions for Corridor and their 90% confidence ellipsoids viewed on a horizontal plane. Only 1 out of 10 ellipsoids for QUASI is displayed.

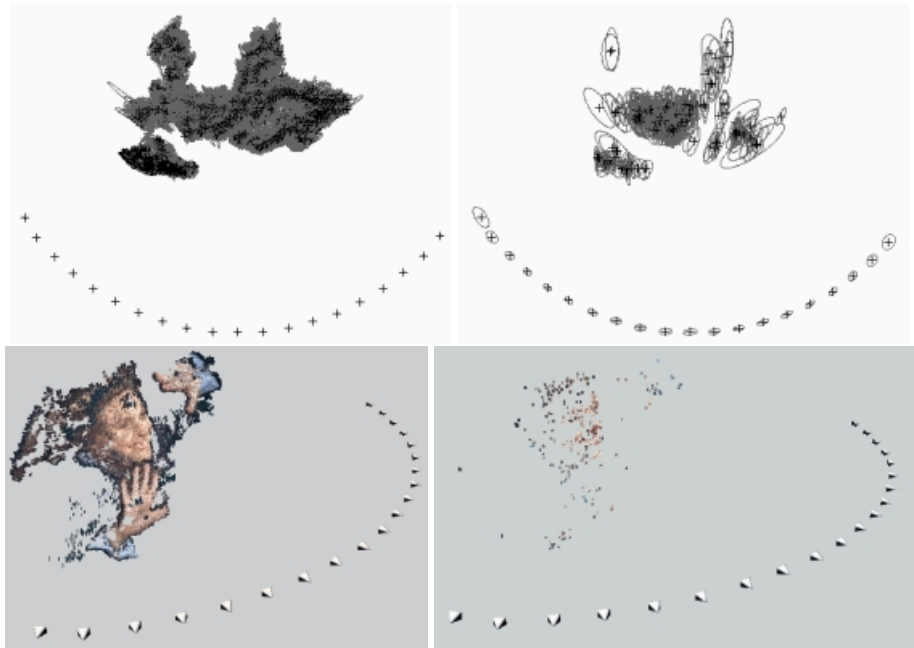


Fig. 6. QUASI (left) and SPARSE (right) reconstruction of the Lady sequence. The 90% ellipsoids for final Euclidian bundle adjustment are enlarged 4 times on the top.

Table 3. Uncertainties for the Lady sequence.

| Lady | #3D points | σ | f | σ_f | \bar{x}_{c_i} | \mathbf{x}_0 | $\mathbf{x}_{\frac{1}{4}}$ | $\mathbf{x}_{\frac{1}{2}}$ | $\mathbf{x}_{\frac{3}{4}}$ | \mathbf{x}_1 |
|--------|------------|----------|-----|------------|-----------------|----------------|----------------------------|----------------------------|----------------------------|----------------|
| QUASI | 26823 | .53 | 849 | 2.26 | 6.1e-4 | 1.2e-3 | 4.2e-3 | 5.4e-3 | 6.1e-3 | 1.9e-2 |
| SPARSE | 383 | .54 | 866 | 13.6 | 3.8e-3 | 5.2e-3 | 9.8e-3 | 1.1e-2 | 1.3e-2 | 2.6e-2 |

method gives the uncertainties in Table 4 and 90% ellipsoids in Figure 7. As the images have been captured with the smallest focal length available, the camera non-linear distortion is becoming not negligible. After a first round of Euclidian bundle adjustment, a second adjustment by adding one radial distortion parameter ρ for all cameras is carried out.

Let \mathbf{u}_0 , \mathbf{x} and \mathbf{u} be respectively the image center, undistorted and distorted points for an image, the first order radial distortion parameter ρ is defined as:

$$\mathbf{u} = \mathbf{u}_0 + (1 + \rho(r/2v_0)^2)(\mathbf{x} - \mathbf{u}_0),$$

where $r = \sqrt{\|\mathbf{x} - \mathbf{u}_0\|^2}$ and $(u_0, v_0)^\top = \mathbf{u}_0$.

We have estimated $\rho = -0.086$ with our method. This estimate is similar to that obtained with a very different method proposed in [4] for the same camera but different images: $\rho = -0.084$.

Computation times in minutes for the QUASI method are given in Table 5 for all sequences using a Pentium III 500 Mhz processor.

8.3 Robustness

The robustness of the methods can be measured by the success rate of reconstruction for a given sequence. The QUASI method is clearly more robust for all sequences we have tested: whenever a sequence is successful for SPARSE, it is equally for QUASI, while SPARSE fails for many other sequences (not shown in this paper) including the Garden-cage sequence in which QUASI succeeds. Furthermore in all our test, the SPARSE method was defined at the very beginning of this Section as the best result between a pure sparse and a mixed sparse-quasi methods, where the mixed one is sometimes the only one which succeed.

Table 4. Uncertainties for the Garden-cage sequence.

| Garden-cage | #3D points | σ | f | σ_f | ρ | σ_ρ | \bar{x}_{c_i} | \mathbf{x}_0 | $\mathbf{x}_{\frac{1}{4}}$ | $\mathbf{x}_{\frac{1}{2}}$ | $\mathbf{x}_{\frac{3}{4}}$ | \mathbf{x}_1 |
|-------------|------------|----------|-----|------------|--------|---------------|-----------------|----------------|----------------------------|----------------------------|----------------------------|----------------|
| QUASI | 50161 | .46 | 732 | 0.27 | -0.086 | 1.1e-4 | 3.8e-4 | 5.2e-4 | 1.9e-2 | 4.4e-2 | 0.12 | 2.6 |

9 Conclusion

In this paper, we have proposed a general quasi-dense 3D reconstruction from uncalibrated sequences. The main innovative idea is that all geometry is computed based on re-sampled quasi-dense correspondences rather than only standard sparse points of interest. Experiments demonstrate its superior performance

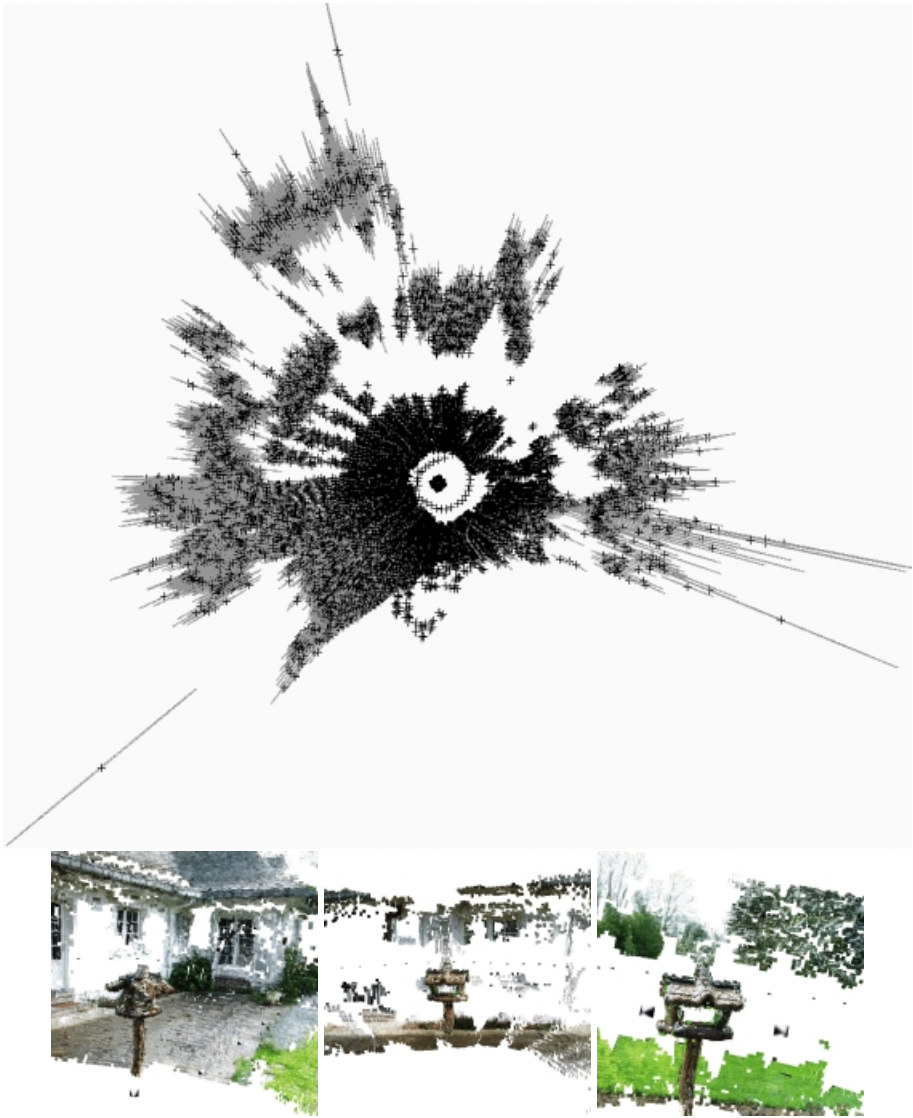


Fig. 7. Top view of the 90% confidence ellipsoids and 3 re-projected views of QUASI reconstruction. The small square shaped connected component at the center is the reconstructed bird cage while the visible crosses forming a circle are camera positions.

Table 5. Computation times (min.) for the QUASI method with a PIII 500 Mhz.

| | #cameras | #3D points | matching and 2-views | 3-views and merge |
|-------------|----------|------------|----------------------|-------------------|
| Corridor | 11 | 16976 | 6 | 11 |
| Lady | 20 | 26823 | 13 | 16 |
| Garden-cage | 34 | 50161 | 25 | 27 |

both in accuracy and robustness due to highly redundant and well spread input data. Quasi-dense reconstruction has also more visualization related application than sparse reconstruction. Future research directions include time reduction for longer sequences by intelligent decimation of reconstructed points in the hierarchical bundle, and all rendering related topics such as meshing and texture merging for a full 3D surface models.

Acknowledgments. We would like to thanks A. Zisserman for the “Corridor” sequence, D. Taylor for the “Lady” sequences, Jerome Blanc for the synthetic data, and Bill Triggs for discussions.

References

1. P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. *ECCV'96*.
2. D.C. Brown. The bundle adjustment – progress and prospects. *International Archive of Photogrammetry*, 21, 1976.
3. D.D. Morris, K. Kanatani and T. Kanade. Uncertainty modeling for optimal structure from motion. In *ICCV'99 Workshop Vision Algorithms : Theory and Practice*.
4. F. Devernay and O. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. *Proceedings of the SPIE Conference on Investigate and Trial Image Processing, San Diego, California, USA*, volume 2567.
5. O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? *ECCV'92*.
6. O. Faugeras and Q.T. Luong, *The Geometry of Multiple Images* The MIT Press, 2001.
7. 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.
8. A.W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3d model construction for turn-table sequences. *SMILE'98*.
9. A.W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. *ECCV'98*.
10. P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve discontinuities. *IJCAI'91*.
11. C. Gauclin and T. Papadopoulos. Fundamental matrix estimation driven by stereo-correlation. *ACCV'00*.
12. C. Harris and M. Stephens. A combined corner and edge detector. *Alvey Vision Conference*, 1988.
13. R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. *CVPR'92*.
14. R.I. Hartley. Euclidean Reconstruction from Uncalibrated Views. *DARPA-ESPRIT Workshop on Applications of Invariants in Computer Vision*, 1993.
15. R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.
16. A. Heyden, *Geometry and Algebra of Multiple Projective Transformations*. PhD. Thesis, Lund Institute of Technology, Sweden, December 1995.
17. M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. *ICCV'95*.
18. T. Kanade, P.J. Narayanan, and P.W. Rander. Virtualized reality: Concepts and early results. *Workshop on Representation of Visual Scenes*, June 1995.

19. K. Kanatani. *Statistical Optimisation for Geometric Computation: Theory and Pra*. Elsevier Science, 1996.
20. Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features ? *ICCV'01*.
21. S. Laveau. *Géométrie d'un système de N caméras. Théorie, estimation, et applications*. PhD Thesis, École Polytechnique, France, May 1996.
22. M. Lhuillier and L. Quan. Image interpolation by joint view triangulation. *CVPR'99*.
23. M. Lhuillier and L. Quan. Robust Dense Matching Using Local and Global Geometric Constraints. *ICPR'00*.
24. M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *IEEE Trans. on PAMI*, 2002.
25. B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI'81*.
26. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, 1980.
27. P. F. McLauchlan. Gauge independence in optimization algorithms for 3D vision. *Proceedings of the Vision Algorithms Workshop*, 2000.
28. D. Nister. Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. *ECCV'00*.
29. Y. Ohta and T. Kanade. Stereo by intra and inter-scanline search using dynamic programming. *IEEE Trans. on PAMI*, 7(2):139–154, 1985.
30. M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *ICCV'98*.
31. M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Metric 3d surface reconstruction from uncalibrated image sequences. *SMILE'98*.
32. 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, 1992.
33. L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. *IEEE Trans. on PAMI*, 17(1):34–46, January 1995.
34. F. Schaffalitzky, A. Zisserman, R. Hartley, and P.H.S. Torr. A six point solution for structure and motion. *ECCV'00*.
35. A. Sashua. Trilinearity in visual recognition by alignment. *ECCV'94*.
36. P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. *ECCV'96*.
37. H.S. Sawhney H. Tao and R. Kumar. A global matching framework for stereo computation. *ICCV'01*.
38. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *IJCV*, 9(2):137–154, November 1992.
39. P.H.S. Torr and D.W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *IJCV*, 24(3):271–300, 1997.
40. P.H.S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *IVC*, 5(15):591–605, 1997.
41. B. Triggs. Autocalibration and the Absolute Quadric. *CVPR'97*.
42. B. Triggs, P.F. McLauchlan, R.I. Hartley, and A. Fitzgibbon. Bundle ajustment — a modern synthesis. *Vision Algorithms: Theory and Practice*, 2000.
43. 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. *AI*, 78(1-2):87–119, 1994.