

Sampling-based Multiview Reconstruction without Correspondences for 3D Edges

Damien Teney
University of Liège, Belgium
Damien.Teney@ulg.ac.be

Justus Piater
University of Innsbruck, Austria
Justus.Piater@uibk.ac.at

Abstract—This paper introduces a novel method for feature-based 3D reconstruction using multiple calibrated 2D views. We use a probabilistic formulation of the problem in the 3D, reconstructed space that allows using features that cannot be matched one-to-one, or which cannot be precisely located, such as points along edges. The reconstructed scene, modelled as a probability distribution in the 3D space, is defined as the intersection of all reconstructions compatible with each available view. We introduce a method based on importance sampling to retrieve individual samples from that distribution, as well as an iterative method to identify contiguous regions of high density. This allows the reconstruction of continuous 3D curves compatible with all the given input views, without establishing specific correspondences and without relying on connectivity in the input images, while accounting for uncertainty in the input observations, due e.g. to noisy images and poorly calibrated cameras. The technical formulation is attractive in its flexibility and genericity. The implemented system, evaluated on several very different publicly-available datasets, shows results competitive with existing methods, effectively dealing with arbitrary numbers of views, wide baselines and imprecise camera calibrations.

I. INTRODUCTION AND RELATED WORK

The problem of 3D scene reconstruction using multiple 2D images from different viewpoints is fundamental in computer vision. The variety of applications, from robotic interaction to phototourism or reverse engineering, has led to the development of numerous methods over the years. These can be broadly classified into two categories: (i) intensity-based multiview stereo methods, which produce *dense* surface reconstructions, and (ii) feature-based methods, which recover *sparse* 3D models of geometric features. Although many of these methods have proven successful in select fields of application, their typical requirements and limitations in operating conditions motivated the development a novel, feature-based method, particularly suited to the use of hard-to-match features. This method, which we successfully applied to the particular problem of 3D curve reconstruction, will be introduced after reviewing related literature.

Methods of the first category mentioned above typically aim at producing detailed 3D reconstructions of objects, enforcing photometric consistency and surface continuity constraints to recover a dense shape description. However,

those methods can typically only operate in precisely controlled settings, usually only with Lambertian surfaces, and with large numbers of precisely calibrated cameras. Those typical requirements for controlled acquisition conditions often prove impractical for general applications (see [1] for a review). While dense reconstructions can offer visually striking results, there are many applications where sparse reconstructions are sufficient, as argued below.

Methods of the second category aim at reconstructing sparse 3D models, made up of isolated geometric features, such as points or edges. Such methods are particularly interesting as they provide more expressive and efficient representations than dense surfaces, typically at a fraction of the computational cost. The classical methods rely on the detection of interest points in the individual 2D views, and then use their local appearance (e.g. using SIFT descriptors [2]) to propose likely matches between observations from different views. The geometric consistency between pairs or triples of points can then be enforced using the well-known epipolar or trifocal constraints [3], effectively leading to the reconstruction of a 3D point cloud compatible with the observations. The first limitations of this approach are obviously those of the extraction and matching of image features, which works best on texture-rich images, but can perform poorly on scenes with mostly homogeneous surfaces or little detail [4]. Moreover, the matching of local appearance descriptors is made harder as the baseline between the considered viewpoints increases [5], practically limiting this approach to the consideration of close pairs of views at a time.

Other methods of the second category make use of image curves, or edges, extracted in the available 2D views [4], [6]–[9]. Reconstructions made up of edge segments convey more geometric information than point clouds [6] and offer greater invariance to changes in illumination and viewpoint. Edge-based reconstructions have moreover proved directly useful for practical applications like pose estimation [10], [11], or the prediction of grasping points of objects [12]. The classical approach, described above, of matching observations between different views (now lines or curves) is however a non-trivial problem [6], exacerbated by the variability in the extraction of said edges from the 2D images. Li *et al.* [4]

reviewed various schemes, e.g. using extended projective geometry [13] or differential geometry [4], or restricting the problem to closed curves [14]. Common drawbacks are strong requirements for precisely calibrated camera [4], [9], [13] and limitations to pairs or triples of views at a time [13]. In [15], Kaess *et al.* focuses on the subproblem of fitting parametric curves to contours identified in several images, using a Monte Carlo-type search as we do. They do not however consider the reconstruction of entire scenes with several objects and the inevitable uncertainty in the input observations. Kahl *et al.* [7] present an approach that also avoids establishing correspondences between views, but delivers results only on simple scenes, reconstructing only small numbers of short curve fragments. We present results on arguably more challenging datasets and in much more varied conditions (see Section IV).

Multiview reconstruction is part of the larger problem of simultaneous localization and mapping (SLAM). In contrast to SLAM, this paper assumes calibrated views and does not make use of core assumptions made by most SLAM methods, most importantly the abundance of input views and feature tracking across views. Some SLAM methods are nevertheless relevant to the current discussion. Klein *et al.* [16] use edges as image features and show how complementary they are to interest points. They focus on the localization problem, and do not deliver convincing results for reconstruction of said edges. Civera *et al.* [17] propose, as we will do, an alternative probabilistic formulation to the classical Gaussian measurement uncertainty, but also focus on localization. [18] goes beyond precisely localizable features by tracking surface patches under photometric constraints to provide a dense reconstruction, but is based on frame-to-frame tracking.

The method proposed in this paper aims at reconstructing a sparse 3D model of geometric features. The key principle is the definition, using each available 2D view, of a probability density in the 3D reconstructed space, which is *compatible* with the view considered. This distribution thus encompasses all backprojected 3D features that could have produced the considered image. Considering all available views, the intersection, or product, of those distributions is then proposed as the distribution of 3D features of the reconstructed scene. We present in Section II an efficient algorithm for obtaining individual samples from that distribution, effectively yielding a set of 3D features (edge fragments in our implementation) describing the reconstructed scene. A second algorithm is proposed that iteratively identifies contiguous regions of high density in the 3D space, which links such samples together, forming continuous 3D curves.

The strength of the proposed approach is to handle non-precisely localizable features, which cannot be matched one-to-one, or which present uncertainty in some dimension of the observation (like a point along an edge). The resulting curve reconstruction method therefore does not rely on

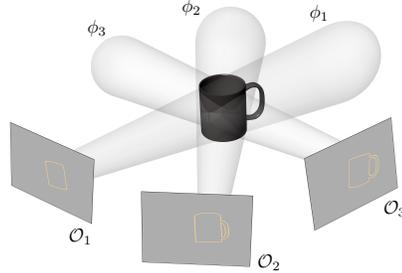


Figure 1. The proposed method uses the observations of each input view, \mathcal{O}_k , to define probability distributions ϕ_k in the 3D, reconstructed space; the reconstructed model lies at the intersection of those distributions.

connectivity in input images, effectively accounting for the variability in the extraction of edges from the images. Other reconstruction methods have been designed to handle uncertainty in the input data, often by relaxing the matching and geometry constraints. For example, Fabbri *et al.* [6] implemented a two stage process, where an initial robust reconstruction is used to optimize the calibration of the cameras, to then obtain a finer reconstruction in the second stage. That approach, which can be traced back to the classical RANSAC algorithm, proved robust, but, in addition to being arguably computationally inefficient, lacks the genericity and flexibility of the formulation presented below. Note finally that similar probabilistic models of objects and image observations have been used in the past [10], [11], and this work can be seen as their extension to the problem of 3D reconstruction.

We must finally remark that reconstruction without correspondences is not new. A basic formulation of the problem was presented in [19]. In [20], Dellaert *et al.* used expectation-maximization to recover the structure of a scene, handling however only precisely localized features, and only presented results on toy examples under several unrealistic assumptions. More recently, [21] showed how to recover the camera transformation between pairs of views using the radon transform, but without considering the 3D structure of the scene at all.

II. PROBABILISTIC RECONSTRUCTION FROM 2D VIEWS

We now present the proposed method, first in a general formulation, then applied to the use of edge segments. Those features correspond, in the input images, to points extracted along lines of maximum gradient, and characterized by their position and orientation on the image plane (see Section III-A). In the reconstructed model, they correspond to oriented 3D points, that we typically represent by short, fixed length, 3D line segments (see Fig. 4b for example); they can be connected together to form continuous curves (e.g. 4c).

A. Probability distributions from image observations

The key idea of the method is to define, from each available 2D view, a probability density over the reconstructed 3D space, which is *compatible* with the observations in that view (1). In other words, it describes the distribution of backprojected 3D features that could have produced the considered image, given the uncertainty present in that image, and in the available estimation of the camera parameters. Formally, each view $k \in [1, N]$ is described by a set of image features, or *observations*

$$\mathcal{O}_k = \{y_i\}_{i \in [1, M_k]}, \quad (1)$$

where $y_i \in \mathbb{R}^2 \times \mathcal{A}$ are the image features, characterized by their position in the image, and some descriptor in an appearance space \mathcal{A} . In the case of edge segments, which have an orientation but no direction, the appearance descriptor is an element on the semicircle (i.e. an angle in $[0, \pi]$), and $\mathcal{A} = S_1^+$. Considering instead more classical interest points, described by their position in the image and their local appearance, the space \mathcal{A} would then contain normalized texture descriptors. The 3D, reconstructed model, is to be defined on a corresponding space $\mathbb{R}^3 \times \mathcal{A}'$. With edge segments, then characterized by a 3D orientation, we have $\mathcal{A}' = S_2^+$.

We will now define, for a view k , a probability distribution ϕ_k on the reconstructed space, using kernel density estimation (KDE). Each element y_i of the considered view is associated with an element of the reconstructed 3D space, $y'_i \in \mathbb{R}^3 \times \mathcal{A}'$. This element can simply be obtained by setting a normalized value for the extra dimensions; e.g., the depth and 3D orientation of our edge segments can be fixed to lie on the image plane in the 3D world (see Fig. 2). This now allows us, using KDE, to define the distribution ϕ_k by its probability density function

$$\phi_k(x) = \frac{1}{M_k} \sum_{i=1}^{M_k} K_i(y'_i, x), \quad (2)$$

where K_i are kernel functions on $\mathbb{R}^3 \times \mathcal{A}'$. Intuitively, one kernel $K_i(y'_i, x)$ models the distribution of all reconstructed features that could have produced the observation y_i . The details, which will depend on the type of features used, are straightforward in the case of edge segments. Looking at the position only, it represents a constant probability density along the backprojected ray (see Fig. 2). Formally, we measure the distance between a given y'_i and x by 3 scalars:

- i. d_1 , the closest distance in position between x and the line defined by y'_i (the backprojected ray),
- ii. d_2 , the depth of x , relative to the camera center,
- iii. d_3 , the difference in orientation between x , and the plane corresponding to the backprojection of the orientation of y'_i .

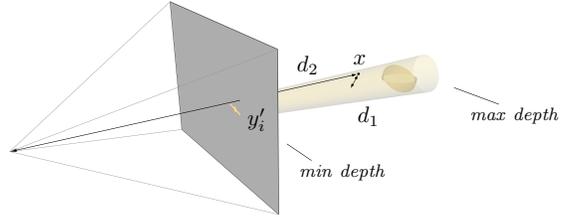


Figure 2. Illustration of an observation y'_i (an oriented point on the image plane) and its associated kernel $K_i(y'_i, \cdot)$ in the reconstructed, 3D space, both for the position and the orientation (surfaces of equidensity in transparent orange). The kernel, evaluated at a point x , uses the distances d_1 and d_2 , resp. to the axis of the backprojected cone and to the camera center (see text for details); d_3 is not represented.

We then define our kernel function $K_i(y'_i, x)$ as the product of 3 independent kernels that make use of those distance measures: a Gaussian kernel on (d_1/d_2) (inducing a conical surface of equidensity for the position, see Fig. 2), a box kernel on d_2 , and a von Mises-Fisher kernel on d_3 (which is a Gaussian-like distribution on orientations [22]). Note that the effect of the box kernel on the depth only corresponds to fixing a hard threshold on the distance to the reconstruction. Indeed, the only assumption that can generally be made here is that a reconstructed point must lie in front of the camera, and within a realistic depth range.

The geometric meaning of our definition of a kernel is quite intuitive, and is illustrated in Fig. 2. For example, the surfaces of equidensity for the position in the 3D space correspond to truncated cones, extending along the camera’s projection rays. The selection of the bandwidth of the kernels is discussed in Section III-B.

The definition of the kernels could be extended to other types of image features, or to include edge curvature for example. We propose another minor extension that takes into account the uncertainty along the orientation of an edge, thereby “flattening” the cone of Fig. 2. For this purpose, the distance d_1 is separated in 2 components d'_1 and d''_1 , respectively aligned and orthogonal to the orientation of the edge; they are then simply evaluated as (d'_1/d_2) and (d''_1/d_2) in Gaussian kernels of respectively large and small variance, thus allowing more slack along the orientation of the edge (see specific results in Section IV-A).

Finally, as a side note, let us remark that defining a probability distribution over the *reconstructed* space, as we did, differs from the classical formulation of the problem, where the reconstructed model is compared, once reprojected in the image space, against the 2D input observations. We will remark that, under certain parameterizations, the two approaches can be rendered equivalent. Our formulation was however chosen in this presentation, as it offers a more intuitive formulation of the sampling-based reconstruction methods that we will propose below.

B. 3D Reconstruction of individual points

The probability distributions ϕ_k we have defined make each use of one single view. We now combine them to produce another distribution ψ in the reconstructed space that is globally consistent with all available views. It is given by its probability density function

$$\psi(x) = \frac{1}{C} \prod_{k=1}^N (\phi_k(x) + \varepsilon), \quad (3)$$

where C is a normalization constant, and ε is a fudge constant, small relative to the scale of $\phi_k(x)$. This definition practically uses the intersection of the ϕ_k , relaxed by the constant ε . This allows observations that appear in some but not *all* input images to produce a nonzero density region in the reconstructed space. This proves necessary in practice, to handle e.g. self-occlusions and missing observations.

Equation (3) gives a formal definition of the 3D reconstruction of the scene. The main goal however is to obtain an explicit and practical representation of this model. Sampling directly from ψ is generally not feasible, but we propose an approximate method based on importance sampling (see for example [10], [23]). Importance sampling (IS) allows one to sample a target distribution $p(x)$, assuming one can evaluate $p(x) = \bar{p}(x)/Z$ up to some normalization constant Z , by using samples x^ℓ from a *proposal distribution* p' , ideally similar to p . IS accounts for the difference between the target and proposal distributions by assigning to each sample x^ℓ a weight given by

$$w^\ell = \bar{p}(x^\ell) / p'(x^\ell). \quad (4)$$

The collection of weighted samples $\{(x^\ell, w^\ell)\}_{\ell=1}^L$ is then, under mild assumptions, asymptotically consistent with the target distribution. This procedure is obviously most efficient as the proposal distribution is close to the target distribution. In practice, the collection of weighted samples is then generally resampled, to a smaller set of $L' (< L)$ *unweighted* samples.

The proposal function used here is given by

$$\psi'(x) = \frac{1}{C'} \sum_{\substack{(k_1, k_2) \\ \in \text{pairs}(1, N)}} \phi'_{k_1}(x) \phi'_{k_2}(x), \quad (5)$$

where C' is a normalization constant, and $\text{pairs}(1, N)$ denotes the list of all unique pairs of indices between 1 and N . Each density function ϕ'_k is a variation of the ϕ_k defined above, in which the kernels used are all box functions. Intuitively, ψ' simply corresponds to all the intersections of pairs of views. Sampling from $\psi'(x)$ is easily done by choosing two arbitrary views k_1 and k_2 , and triangulating two random observations y_1 and y_2 from each, the kernels of which intersect at least by a small amount (i.e. the 3D projections of which intersect each other within a small threshold). The bandwidth of the box kernels of ϕ' will

be chosen so that they extend up to a reasonable cutoff threshold of the exact kernels of ϕ . This ensures that the proposal distribution ψ' will generate samples in all of the most interesting regions of the target distribution ψ . The weights assigned to the proposal samples of ψ' are then simply computed using (4). They can then be resampled to obtain a set of non-weighted points.

C. 3D Reconstruction of continuous curves

The method presented above reconstructs individual points as samples from a probability distribution in the 3D space. Some interesting parts of the scene may however correspond to regions of lower density (e.g. due to missing observations in one or several views), but which can however still be identified as local maxima. Moreover, in the particular case of curve reconstruction, one wants to reconstruct continuous curves, and not individual points. Those two objectives can be met through the iterative procedure described below, which uses the individual samples as starting points for a stochastic exploration of the reconstructed space.

For each reconstructed curve, the procedure starts with a sample $x_0 \in \mathbb{R}^3 \times S_2^+$. It then iterates, searching at each step for a point x_{i+1} along a ridge of locally maximum probability density. Formally, local proposals are generated from a point $x = (p, \theta)$ of position $p \in \mathbb{R}^3$ and orientation $\theta \in S_2^+$ (a unit 3-vector), as a set of L samples:

$$\text{proposals}(x) = \{(p + \Theta_\kappa(\theta) * \Gamma_{(\alpha, \beta)}, \Theta_\kappa(\theta))_j\}_{j \in [1, L]}, \quad (6)$$

where Γ is a gamma distribution that generates the distance in position to a proposal, and Θ is a Von Mises-Fisher distribution used to randomize the orientation. This uses the assumption that the next point of the curve is most likely in the direction of the current point. The parameters κ, α, β define how “spread out” the proposals are from an exactly straight line. The likelihood of each proposal is evaluated (Eq. 3), and the best one is selected as the new point x_{i+1} of the curve. The procedure is repeated, unless the likelihood of all L proposals fall below a threshold, indicating the probable end of the curve. That threshold is fixed beforehand as fraction of the mean density of a batch of samples of the whole scene. The procedure is comparable to the classical Canny algorithm, which, likewise, follows ridges of local optima until falling below a predefined threshold. Note that the use of a purely random walk scheme for selecting the neighbours in our method — as opposed to estimating local derivatives of a likelihood function (as could be done using differential geometry) — is motivated by the genericity of the procedure, which we plan to apply to other types of image features as future work. Finally, as the scene is being reconstructed, we “prune” ψ , removing the kernels that have significant overlap with the curves already reconstructed. This helps reconstructing parts of the scene of low probability density, initially masked out by regions of

higher density, and also avoids reconstructing several times the same portions of a scene.

III. IMPLEMENTATION

A. Edge detection in input images

The image features we use are oriented 2D points, identified along the edges in the images. We selected the method of [24], which is a simple method based on image gradients that extracts the orientation of the edges significantly better than the traditional method, which simply uses the direction orthogonal to the gradient. That method was chosen instead of more sophisticated ones which take texture or global segmentation into account, as they can be extremely slow and are thus not an option for many applications of 3D reconstruction. This also ensured a fair comparison with other published methods which used basic gradient-based edges as well.

B. Choice of parameters of the reconstruction

The kernels associated with the observations are parametrized by their bandwidth in position and orientation. This size should reflect the estimated uncertainty in the input data, and can be set according to a small fraction to the estimated scale of the scene. Our experiments showed however that the method was not particularly sensitive to the choice of those parameters. For example, in the experiments (with both small and large camera calibration errors) of Section IV-D, with 640×480 pixel images, the size of the kernels was set to allow a corresponding maximum deviation in the images of about 12 pixels and 20° .

The parameters used for local proposals (Eq. 6) are also to be set relatively to the scale of the scene. For example, the scene of Section IV-D, measuring about 1000 mm in diameter, used local proposals corresponding to a spacing of 5 mm and a deviation in orientation of 15° on average, with $L = 50$.

Finally, the running time of the iterative procedure grows linearly with the number of reconstructed points. The cost associated with the reconstruction of a point mostly corresponds to the evaluation of ψ (Equation 3) for the proposals. One evaluation involves the processing of every kernel of every input view, and is thus $O(N\bar{M})$, where N is the number of views, and \bar{M} the average number of observations per view. We currently use this basic implementation. However, a cleverer implementation could efficiently preselect the few kernels likely to be relevant to the evaluation of a given point, using an ordered data structure. Since the influence of a kernel in its distribution drops below insignificant values past some distance, one could, in this way, restrict the evaluation of the kernels to a small fraction of them.

IV. EXPERIMENTAL RESULTS

The proposed method was evaluated on 4 very different datasets. It is notoriously hard to produce ground truth

reconstructions for evaluating feature-based methods, due to the ambiguous selection of the features to reconstruct. Datasets for benchmarking dense reconstruction methods have been produced; however, the ground truth model is not necessarily made public [1], and the selection of actual edges from continuous surfaces [25] or 2.5D models makes it hard to design a meaningful quantitative evaluation of a method like ours. Competing methods for curve reconstruction faced a similar situation, which explains why no extensive qualitative evaluations were published. [6] made an exception, but they only evaluate their ability to match correct curve fragments between views, using a set of manually labelled ground truth correspondences — which was unfortunately not made public.

Practically, our prototype software was implemented in Matlab. Running times of such an implementation (especially of an iterative method) have little meaning, as the only switch to a compiled language offers potentially enormous room for improvement. Bearing this in mind, we report, as a base point, that a reconstruction as shown in Fig. 4 or Fig. 6a currently takes about 2 to 5 minutes on a standard laptop without multithreading. Most recent competing methods do not discuss the issue of efficiency; Fabbri *et al.* [6] report running times in the order of minutes on scenes like the dinosaur (see below). Let us note moreover that most parts of our algorithm are straightforward to parallelize.

A. Synthetic toy example

The lack of datasets with proper ground truth motivated the use of a synthetic toy example, in order to evaluate and demonstrate basic properties of the proposed method. The scene, pictured in Fig. 3a, contains curves of various lengths and shapes. Their exact 3D shape is used to directly generate the 2D edge maps used as input to the reconstruction method. This bypasses the stage of edge extraction from 2D images, focusing this evaluation on the reconstruction process alone. To simulate realistic conditions and missing observations, random parts of the curves are masked when generating those edge maps. The scene itself measures about 500 mm in diameter; we use 7 views from different viewpoints around the scene, at a distance of approximately 900 mm.

We compare reconstructions and ground truth using the accuracy/completeness metrics proposed in [1]. To obtain accuracy, we measure the Euclidean distance from each reconstructed point to the closest ground truth curve. The accuracy is then defined as the distance so that 90% of the points fall below that threshold. To obtain the completeness, we consider a number of points sampled uniformly along the ground truth curves, and count the ratio of them that have a part of the reconstruction within a reasonable distance (15° in orientation, and $5/8$ mm in position for scenes without/with noise). The exact choice of those thresholds is not relevant here, since we use them to compare different

methods and not to obtain absolute performance values. We report accuracy/completeness scores for 4 different reconstruction methods: (i) a baseline method where we perform random triangulations (Eq. 5), and keep a fixed number (1000) of points with a probability density (computed as in Eq. 3, but without orientation) above a threshold; this corresponds approximately to the basic approach where one simply imposes a maximum 2D distance between the re-projected reconstruction and the input observations; (ii) our sampling method (Section II-B) used to recover the same number (1000) of points; (iii) our iterative method for curve reconstruction (Section II-C); (iv) the same method accounting for uncertainty specifically along the orientation of edges (Section II-A).

Each method is run with different lower thresholds on the probability density of reconstructed points, setting the tradeoff to be made between accuracy and completeness. We report results in Fig. 3b, with and without noise on camera calibrations (in the form of added Gaussian perturbations of $\sigma = 4$ mm on the camera positions). We also plot, in Fig. 3c, the accuracy and the local probability density (Eq. 3) of a number of random samples (Eq. 5). This allows verifying that there is indeed a correlation between the probability density obtained through our definition, and the actual correctness of a reconstructed point. Reconstructions showing good accuracy can however sometimes correspond to low probability densities, which explains why our sampling method alone cannot recover the entire scenes, as opposed to the iterative method. Moreover, we also verify that the correlation between accuracy and probability density still holds when adding noise (as above) to the camera calibrations.

B. Dinosaur

The “dinosaur” dataset is standard for the evaluation of dense reconstruction methods [1]; we use the version made of 16 views from a circle around the object. We show in Fig. 4b a reconstruction made of individual points, obtained using our sampling method. These samples are drawn mostly in the regions of high probability density of the reconstructed space, with more samples in the regions the most precisely defined, e.g. along the crest on the back of the animal (such sharp edges correspond to well-defined edges in the 2D images). In Fig. IV-Bc, we show a reconstruction of continuous curves of the same scene; those curves are correctly identified along ridges of local maxima of the probability density function, yielding a high quality reconstruction of the object. Those results, directly comparable with those presented in [6], show a clear advantage, particularly in the level of noise in the reconstruction.

C. High-resolution building

Strecha *et al.* [25] produced a dataset of high resolution pictures of buildings for evaluating dense reconstruction

methods. We chose to evaluate our method on one of those scenes (“Herz-Jesu-P8”) as it represents a very different type of input data than our other evaluations. The images are of high resolution, but the nature of the scene (very textured surfaces and lots of fine details) renders the extraction of stable edges from the 2D images a difficult problem already. The reconstructed 3D model (Fig. 5a-b) exhibits missing parts, which are a direct consequence of this problem (corresponding to missing observations in the input data). The 8 viewpoints span only a small arc, roughly in the same plane, leaving a great deal of uncertainty in the *depth* dimension, in particular for the edges parallel to that plane. This can be observed when viewing the reconstructed model from the top (Fig. 5b), as some supposedly straight edges meander in this dimension. The same curves however, when re-projected on an input image, always closely match the input images (Fig. 5c). [25] uses a particular distance measure to evaluate dense reconstruction methods, requiring non-public information (calibration uncertainty), which prevented direct performance comparisons.

D. Office desk

We finally consider an indoor scene, containing typical household items with little texture (see Fig. 6a), shot from 12 different viewpoints around them. This represents the type of scenes that motivated our approach, in the context of robotic applications, where a robot would take the pictures using an arm-mounted camera. The extrinsic calibration of the camera would thus be known with a precision corresponding to the accuracy of the robotic arm. In this evaluation however, and for purely practical reasons, we used a checkerboard pattern in the scene with standard calibration software. We obtained visually excellent reconstructions (Fig. 6). A challenging part of the scene is the checkerboard, as it contains many lines close together, in both similar and different orientations. We then intentionally added noise to the positions of the 12 cameras, to verify the influence on the reconstruction (see Fig. 6b-e for details). The highest tested level of additional noise, drawn from a Gaussian distribution of $\sigma = 8$ mm, introduces corresponding translation errors as large as 10 pixels on the 640×480 pixel images. Experiments show that a reconstruction is still possible; some regions of the reconstructed space now receive a probability density lower than the allowed threshold, explaining missing parts in the reconstruction. Those results are representative of the performance we obtained on many experiments of similar nature.

V. CONCLUSIONS AND FUTURE WORK

We presented a novel method for feature-based 3D reconstruction from multiple calibrated views. We introduced a probabilistic formulation that admits hard-to-match features particularly suited to edge segments. The reconstructed scene is modelled as a probability density in the 3D space,

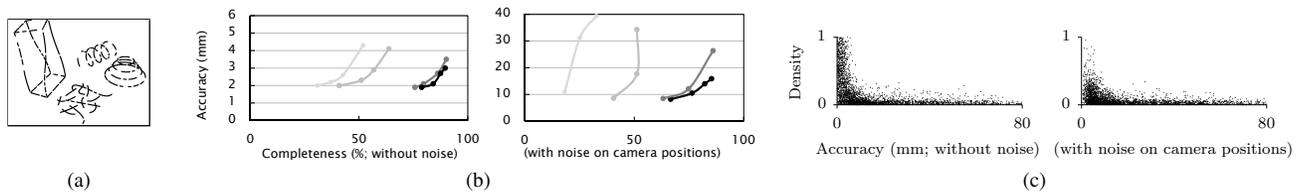


Figure 3. Evaluation on synthetic scene: (a) example input edge map, note missing observations; (b) accuracy/completeness scores for (light to dark) random sampling based on position only, our sampling method using orientation, our iterative method with conical kernels, then with “flattened” kernels (accounting for uncertainty along the edge orientations); (c) density/accuracy of random samples (density evaluated up to a multiplicative constant).

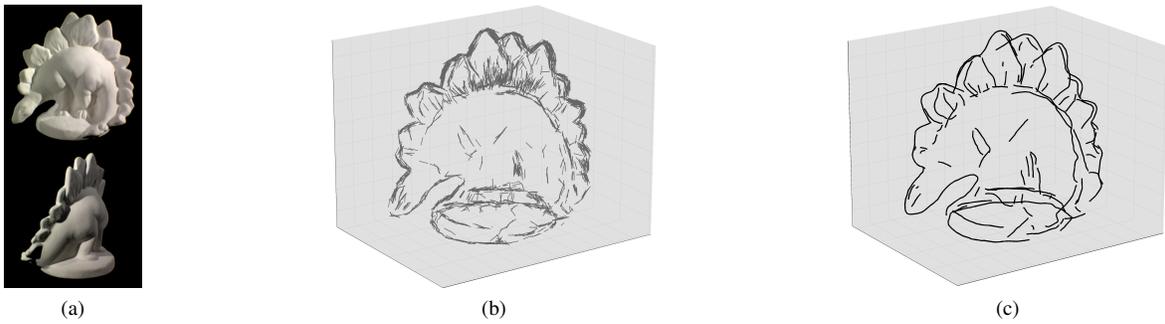


Figure 4. (a) Example images of the dinosaur dataset; (b) individual reconstructed 3D points obtained through our sampling method; (c) reconstruction of continuous curves.

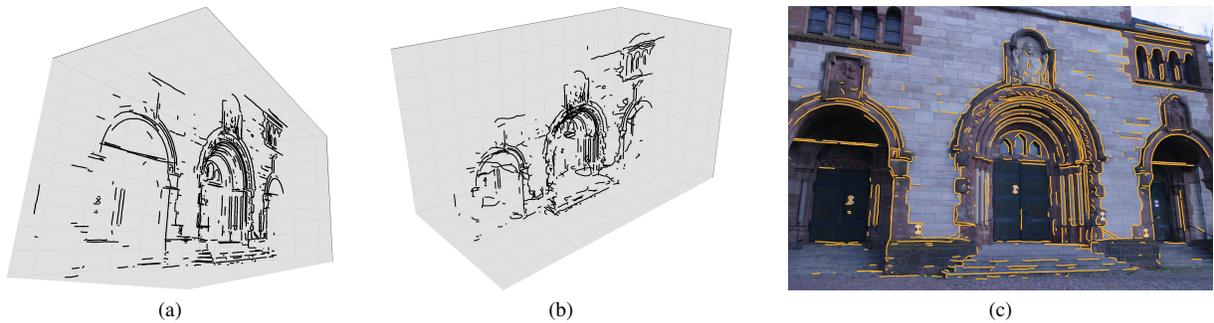


Figure 5. (a) Reconstruction of the building dataset, missing parts are mostly due to missing observations, difficult to extract from the input images; (b) other view of the reconstruction, showing the imprecisions in depth, as the input viewpoints span only a small arc in front of the building; (c) reconstructed edges, reprojected on an input image, match however closely.

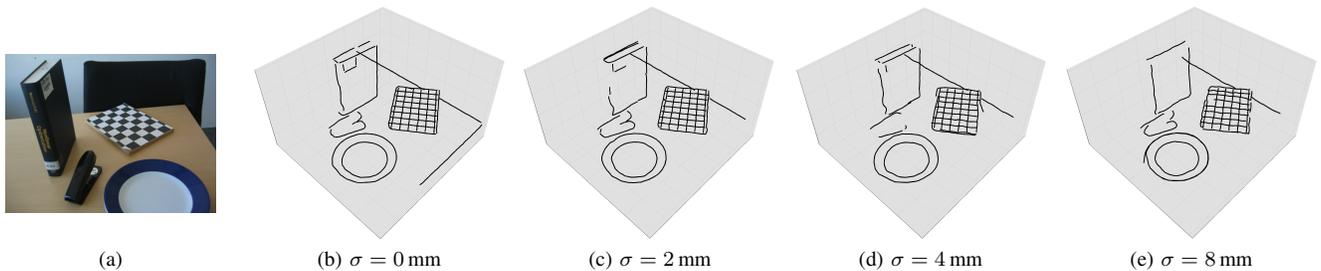


Figure 6. Reconstruction of scene with error in camera calibration; one input image (a); 3D reconstructions (rendered from a novel viewpoint) with original estimated camera calibration (b) and with added perturbation on camera position from Gaussian noise of variance σ (c-e); significant levels of error still allow reconstruction, at the price of some imprecisions (plate, checker board) and missing edges (book, lower edge of the table).

from which we can draw individual samples. Those are then used as starting points to reconstruct continuous 3D curves. The effectiveness of the approach was demonstrated on existing and new datasets, and showed competitive results with an existing method, while exhibiting more technical flexibility and genericity in its formulation. An important direction for future work is the evaluation of this method on features other than edges.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 (Specific Programme Cooperation, Theme 3, Information and Communication Technologies) under grant agreement no. 270273, Xperience. Damien Teney is supported by a research fellowship of the Belgian National Fund for Scientific Research.

REFERENCES

- [1] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2006, pp. 519–528. [1](#), [5](#), [6](#)
- [2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. [1](#)
- [3] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004. [1](#)
- [4] G. Li, Y. Gene, and S. Zucker, "Multi-view edge-based stereo by incorporating spatial coherence," in *3-D Digital Imaging and Modeling, 2007, Sixth International Conference on*, 2007, pp. 341–348. [1](#), [2](#)
- [5] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *International Journal of Computer Vision*, vol. 73, pp. 263–284, 2007. [1](#)
- [6] R. Fabbri and B. Kimia, "3D curve sketch: Flexible curve-based stereo reconstruction and calibration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1538–1545. [1](#), [2](#), [5](#), [6](#)
- [7] F. Kahl and J. August, "Multiview reconstruction of space curves," in *IEEE International Conference on Computer Vision (ICCV)*, 2003, pp. 1017–1024 vol.2. [1](#), [2](#)
- [8] S. Liu, K. Kang, J.-P. Tarel, and D. B. Cooper, "Free-form object reconstruction from silhouettes, occluding edges and texture edges: A unified and robust operator based on duality," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 1, pp. 131–146, 2008. [1](#)
- [9] K. Potsch and A. Pinz, "3D geometric shape modeling by '3D contour cloud' reconstruction from stereo videos," in *Computer Vision Winter Workshop 2001*, Mitterberg, Austria, 2001. [1](#), [2](#)
- [10] R. Detry, N. Pugeault, and J. Piater, "A probabilistic framework for 3D visual object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1790–1803, 2009. [1](#), [2](#), [4](#)
- [11] D. Teney and J. Piater, "Probabilistic Object Models for Pose Estimation in 2D Images," in *DAGM*, ser. LNCS, vol. 6835/2011. Heidelberg: Springer, 2011, pp. 336–345. [1](#), [2](#)
- [12] M. Popović, D. Kraft, L. Bodenhagen, E. Başeski, N. Pugeault, D. Kragic, T. Asfour, and N. Krüger, "A strategy for grasping unknown objects based on co-planarity and colour information," *Robotics and Autonomous Systems*, 2010. [1](#)
- [13] C. Schmid and A. Zisserman, "The geometry and matching of lines and curves over multiple views," *International Journal of Computer Vision*, vol. 40, no. 3, pp. 199–233, 2000. [2](#)
- [14] R. Berthilsson, K. Astrom, and A. Heyden, "Reconstruction of general curves, using factorization and bundle adjustment," *International Journal of Computer Vision*, vol. 41, pp. 171–182, 2001. [2](#)
- [15] M. Kaess, R. Zboinski, and F. Dellaert, "Mcmc-based multi-view reconstruction of piecewise smooth subdivision curves with a variable number of control points," in *European Conference on Computer Vision (ECCV)*. Springer, 2004, pp. 329–341. [2](#)
- [16] G. Klein and D. Murray, "Improving the agility of keyframe-based slam," in *European Conference on Computer Vision (ECCV)*, 2008. [2](#)
- [17] J. Civera, A. J. Davison, and J. M. M. Montiel, "Unified inverse depth parametrization for monocular slam," in *Robotics: Science and Systems*, 2006. [2](#)
- [18] R. Newcombe, S. Lovegrove, and A. Davison, "DTAM: Dense Tracking and Mapping in Real-Time," in *IEEE International Conference on Computer Vision (ICCV)*, 2011. [2](#)
- [19] Y. qing Cheng, R. T. Collins, A. R. Hanson, and E. M. Riseman, "Triangulation without correspondences," in *DARPA Image Understanding Workshop*, 1994, pp. 993–1000. [2](#)
- [20] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun, "Structure from motion without correspondence," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000. [2](#)
- [21] A. Makadia, C. Geyer, and K. Daniilidis, "Correspondence-free structure from motion," *International Journal of Computer Vision*, vol. 75, no. 3, pp. 311–327, 2007. [2](#)
- [22] R. A. Fisher, "Dispersion on a sphere," in *Proc. Roy. Soc. London Ser. A.*, 1953. [3](#)
- [23] E. B. Sudderth, "Graphical models for visual object recognition and tracking," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006. [4](#)
- [24] A. Tamrakar and B. B. Kimia, "No grouping left behind: From edges to curve fragments," in *IEEE International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8. [5](#)

- [25] C. Strecha, W. von Hansen, L. J. V. Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 5, 6