3D Curve Sketch: Flexible Curve-Based Stereo Reconstruction and Calibration

Ricardo Fabbri and Benjamin Kimia
Brown University
Division of Engineering
{rfabbri,kimia}@lems.brown.edu

Abstract

Interest point-based multiview 3D reconstruction and calibration methods have been very successful in select applications but are not applicable when an abundance of feature points are not available. They also lead to an unorganized point cloud reconstruction where the geometry of the scene is not explicit. The multiview stereo methods on the other hand yield dense surface geometry but require a highly controlled or calibrated setting. We propose and develop a novel framework for 3D reconstruction and calibration based on image curve content, whose output is a 3D curve sketch*, an unorganized set of 3D curve fragments. This approach, which is meant to augment the previous approaches, results in a reconstruction of* geometric curve structure *which can serve as a scaffold on which surface patches can be potentially reconstructed. It is intented for the setting where a number of images are available with coarsely calibrated cameras. The approach operates in two stages. A reliable partial 3D curve sketch is first reconstructed and this is used to refine the cameras to yield a more complete 3D curve sketch in a second stage. A key advantage of this approach is the ability to integrate information across a large number of views. The results have been evaluated on a few datasets.*

1. Introduction

The 3D reconstruction of scenes from images taken from multiple cameras and the calibration of these cameras are fundamental problems in Computer Vision. The state-of-the-art approaches to these problems either find correspondences among *isolated interest points* which give a 3D point cloud reconstruction of the scene, or are *intensity-based multiview stereo methods*, which give detailed mesh reconstructions of simple objects but use controlled acquisition. These approaches have been successful for select domains of application, resulting in autocalibration and useful 3D reconstructions. However, their requirements/assumptions are not applicable in general, motivating a novel approach based on image curve content to complement existing approaches.

The application of **interest-point-based methods** have been successful in scenes with texture-rich images, such as
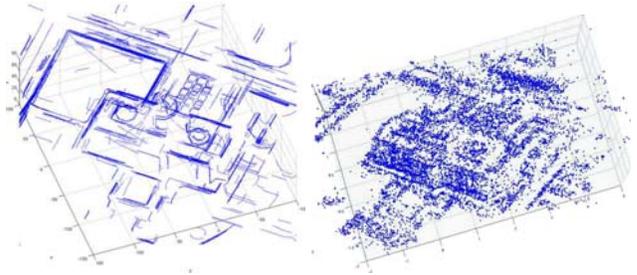

Figure 1: The 3D reconstuction as a 3D curve sketch (left) gives explicit geometry in contrast to the 3D cloud of point reconstruction. See supplementary data for 3D results.

in Phototourism [21, 1]. Despite their success, these methods are not applicable in general settings. First, they assume an abundance of interest points per independently moving object, possibly only with a fair degree of texture in projected images. Such a rich texture is not always available in scenes with homogeneous regions, *e.g.*, some man-made environments, or when objects project to a small number of pixels, relative to the basis of computation of interest points. In these cases, there may be sufficient image curve geometry, however, to reconstruct the scene and recover the cameras.

Second, the stability of interest points is significantly reduced as the baseline exceeds $30°$ [20], so these methods are fundamentally restricted to a limited range of baselines. In contrast, certain curve features, such as sharp ridges on a building, persist over a much greater range of views. Third, reconstruction using interest points results in an unorganized cloud of 3D points where the geometric structure of the underlying curves and surfaces is not explicit. This is not much of a problem if the focus is on calibrating the viewpoints as in Phototourism [1]. However, when the 3D object geometry is required, such as in modeling for architecture, archaeology, entertainemnt, object recognition, and robotic manipulation, it would be useful to augment the output with an explicit geometric structure such as a 3D curve sketch.

The second category of 3D reconstruction techniques, **multi-view stereo methods**, produce detailed 3D reconstructions of objects imaged under controlled conditions by a large number of precisely calibrated cameras [9, 12, 15, 11]

(see [26] for a review). However, they cannot handle general scenes, since most approaches are specific to a single object or to a specific type of object, such as buildings. In addition, they often require accurate camera calibration, operate under controlled acquisition, and are often initialized by the visual hull of the object or a bounded 3D voxel volume.

The goal of this paper is to augment current multiview reconstruction and calibration technology by developing a generally applicable framework based on **image curves** when a large number of views are available, *e.g.*, a video sequence. Image curve fragments are attractive because they have good localization, they have greater invariance than interest points to changes in illumination, are stable over a greater range of baselines, and are typically denser than interest points. Furthermore, the reflectance or ridge curves provide boundary condition for surface reconstruction, while occluding contour variations across views indicate surface properties [5]. The notion that image curves contain much of the image information can be supported by a recent study [6] which found that generally subjects are able to judge 3D surface normals of an object depicted by a line-drawing almost as accurately as for objects depicted by a shaded image.

However, two fundamental obstacles must be surmounted before image curve fragments can be used in 3D reconstruction and camera calibration. First, given a pair of image curve fragments, each from a different view, there is an ambiguity in intra-curve correspondence, *i.e.*, given a point on curve one, it is not clear which point on curve two it corresponds to, Figure 2(a). This is not the case when matching two isolated feature points. Second, the linking of edges into curve fragments is not stable across views, even for slight baseline differences, Figure 2(b), resulting in multiple transitions as the view changes [16].
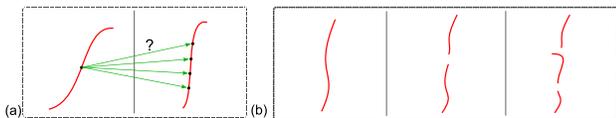

Figure 2: Challenges in using curve fragments in multiview geometry: (a) point correspondence ambiguity along the curve, (b) instabilities with slight changes in viewpoint.

Previous curve-based methods have circumvented problems in various ways. First, the use of general but *closed curves* resolves the curve transition issue [22, 3], but requires the successful segmentation of images which is unlikely. Second, the use of *algebraic curves* resolves the transition issue with low-order models based on extrinsic polynomial equations such as conics [8, 17], quadrics [23, 29], and higher order algebraic curves [18]. However, algebraic curve geometry is restrictive, limiting their application to general scenery. Third, *silhouettes and visual hulls* have been successfully used in several aproaches [15, 14, 10, 19, 4, 7], but require a *highly controlled acquisition* for their extraction, and work only for one object per scene. Fourth, *straight lines*

are a simple, compact and distinctive geometric representation to be used in generating model hypotheses and detecting objects for man-made environments. However, this approach severely restricts the application domain by not allowing for curved structures. Finally, current multiview methods based on general curves require *highly accurate calibration*, which is not easily available in general.

Overview of our approach: Figure 3 illustrates our approach. We assume a large number of views are available, $N \geq 6$, *e.g.*, from video or from multiple cameras monitoring a scene. The cameras have been *coarsely* calibrated, typically with $2 - 5$ pixel error. The goal is to produce a dense collection of unorganized 3D curve fragments, or the **3D curve sketch**, which reflect the underlying geometry arising from a combination of 3D surface geometry and viewing/illumination arrangements, *i.e.*, occluding contours, reflectance contours (albedo discontinuity), shadow curves, shade curves, ridges, *etc*. We also use this reconstruction to refine the cameras so that a better and denser 3D curve sketch can be obtained, and so that a better distinction can be made between occluding contours and view-stationary contours such as reflectance curves and ridges.

Our approach is divided into two stages. In the first stage, where cameras are calibrated coarsely, the goal is to reconstruct a partial, but reliable **core 3D curve sketch** to be used in the second stage for refining calibration and for obtaining a more complete 3D curve sketch. We form a 3D curve fragment hypothesis by pairing two long curve fragments, each from a different view, with sufficient epipolar overlap. We call the two views from which a curve pair hypothesis is formed as the **hypothesis views**. These views rotate in the reconstruction process among available view pairs. Each curve pair hypothesis is reprojected onto a set of other views, the **confirmation views**, and rated for consistency with the image and curve differential geometry. Our approach can be thought of as an automated version of the curve-based CAD system from [28]. Those hypotheses with enough evidence in confirmation views are then reconstructed to form the initial core 3D curve sketch, as discussed in Section 2.

The core 3D curve sketch enables a curve-based measurement of calibration accuracy, by summing up the distances between reprojected curves and supporting image curves. This allows for a refinement of cameras through bundle adjustment, resulting in subpixel calibration errors. The refined cameras in turn allow for additional sensitivity so that smaller curve fragment pair hypotheses can be confirmed or discarded. This is discussed in Section 3. Section 4 discusses the details of our implementation and Section 5 evaluates the approach on several datasets. Figure 1 highlights the results.

2. Curve-based Multiview Stereo

Our approach is based on reasoning with image curve fragments obtained from a large number of views. We denote
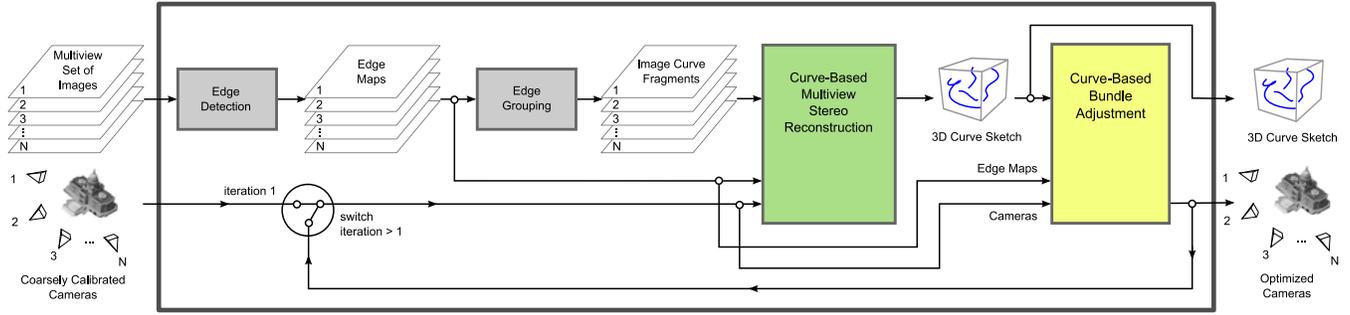
Figure 3: A schematic overview of our approach: The input to our system is *(i)* a set of N images taken of the scene where N is at least 6 but typically larger, and *(ii)* a set of N coarsely calibrated cameras, one for each view: the system pairs a curve fragment from one view with one from another view (hypothesis views) to form a hypothesis which is verified or discarded based on the edge map of other views (confirmation views). Once a set of reliable curve fragment hypotheses are formed, curve-based bundle adjustment is used to refine the cameras and the process is repeated. The output of our system is a set of 3D curve fragments (the 3D curve sketch) and refined cameras.

image curve fragments by γ_i^v, where v indexes into views, $v = 1, \ldots, N$, and i enumerates the image curves within frame v, $i = 1, \ldots, M_v$. The goal is to produce a set of 3D curve fragments $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_K\}$, supported by sufficient evidence from the image curve fragments. A putatively corresponding set of image curve fragments $\{\gamma_{i_1}^1, \ldots, \gamma_{i_n}^n\}$ from n views can arise from a single 3D curve Γ_k, *e.g.*, from sharp ridges, reflectance curves, shadows and shades, which we refer to as *view-stationary curves*. Alternatively, a set of image curve fragments can be apparent contours arising from occluding contours on a surface, which we refer to as view-nonstationary curves. In practice, except for *low-curvature surfaces*, the set of occluding contours cluster in 3D, approximating a single 3D curve. In the first stage of our reconstruction where the calibration errors are high, view-stationary and some view-nonstationary curves are considered together, but they are differentiated in the second stage up to the accuracy of the refined calibration.

Hypothesis formation. The search for a potentially corresponding set of curve fragments $w_k = \{\gamma_{i_1}^1, \gamma_{i_2}^2, \ldots \gamma_{i_n}^n\}$ can be computationally overwhelming due to the combinatorics. However, observe that a set of corresponding curves arising from a single 3D curve Γ_k can be delineated by *any* pair in the set. Thus, the 3D curve fragment reconstruction can be reduced to forming corresponding pairs of curves and removing redundancy among pairs that agree. Since it is not known in advance which pair of curve fragments correspond, we must consider all pairs which can potentially arise from a single 3D curve, *i.e.*, those with common epipolar overlap.

Definition 1. Two curve fragments form a **curve fragment pair hypothesis** if a common portion of each arise from a common 3D curve fragment, Figure 4

In the first stage, only sufficiently long curve fragments (typically $l > \tau_l = 40$ pixels for HD video) are considered, since they are reliable indicators of image structure. This is relaxed for the subsequent stages when a more precise calibration is available. The goal in the initial stage is to produce
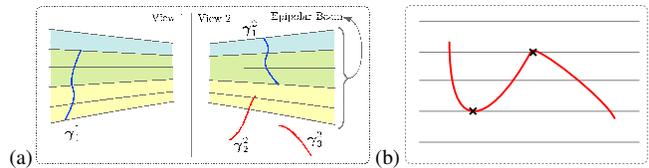

(a)   (b)

Figure 4: (a) The epipolar set of lines spanned by points along the curve fragment γ_1^1 in view 1 delineates the region where potentially corresponding curves can be in view 2, ruling out many curves, *e.g.*, γ_3^2, which fall outside this region, and selecting those with minimal epipolar overlap ($\tau_e = 5$ pixels), *e.g.*, γ_1^2 but not γ_2^2. (b) We avoid multiple intersection with epipolar lines by breaking curve fragments at epipolar tangencies.

a core set of reliable curves Γ_k which can be used to bootstrap for a better calibration. Since long curves typically undergo a wide range of transitions with view changes, we only insist on a sufficient epipolar overlap ($\tau_e = 5$ pixels). Multiple curve intersections are handled by breaking all curve fragments at epipolar tangencies, Figure 4(b), and keeping the ones with length greater than τ_l. The result is a set of curve fragment pair hypotheses $\mathcal{W} = \{w_k, k = 1, \ldots, K\}$ collected from pairs of hypothesis views, Figure 7.

Hypothesis test and confirmation: Observe that due to potential partial occlusion, a reprojection of Γ_k cannot be expected to be confirmed on each view. Rather, confirmation by a few views, even by one view if the evidence is strong enough, should be sufficient. However, failing a presence in all other views implies that the curve fragment pair hypothesis is invalid.

Specifically, without loss of generality, renumber the hypothesis views as views 1 and 2, and renumber the curve fragment pair hypothesis under consideration as $w_k = (\gamma_1^1, \gamma_1^2)$ only, for the purpose of this discussion. The curve fragment pair corresponds to a 3D curve fragment Γ_k which projects to an image curve fragment $\gamma^{k,v}$ in each view v. We could potentially seek confirmation for $\gamma^{k,v}$ by matching it to the curve fragments in view v $\{\gamma_1^v, \gamma_2^v, \ldots, \gamma_{n_v}^v\}$. How-

ever, this would expose us to the frailty of the edge grouping process, since the matching curve fragment may exist in the form of several subfragments, *e.g.*, Figure 2(b). On the other hand, while relying on edges has the advantage of localized support without errors from grouping transitions, use of edges location to clutter is not sufficiently selective in its support of curve hypotheses.

We select a middle-ground using curvelets [27], local groupings of a few edges (say 5-10). The curve transition of Figure 2(b) affect these curvelets as well, but only very few in many are affected. In addition, the grouping process provides sufficient geometric context to make their support meaningful. We represent these local groupings as a pointwise map endowed with tangent and curvature.[1] In this paper, we do not use curvature, but we plan to use it in the future and expect the addition of even a rough form of curvature will be very helpful. Thus, for this paper the curvelets are reduced to edge location with tangent attributes.
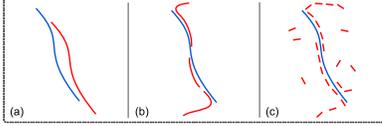

Figure 5: The reprojection of a curve fragment pair (γ_1^1, γ_1^2) onto a third view (blue) can enjoy the support of an existing curve fragment (red) as in (a), but with non-negligible likelihood the transition of edge grouping process can result in broken curve segments, missing edges, and false groupings (b). This motivates the use of a small neighborhood of an edge – not curve fragments – in the confirmation process.

Formally, define a supporting edge element, Figure 5, as

Definition 2. *(Supporting edge)* An edge at point p_0 with tangent $\boldsymbol{t}_0$ of an edge map M supports an an image curve point $\boldsymbol{\gamma}(s)$ with tangent $\boldsymbol{t}(s)$ when:

$$
\begin{cases}
\mathbf{p}_0 = \underset{\mathbf{p} \in M}{\operatorname{argmin}}\, d(\mathbf{p}, \boldsymbol{\gamma}(s)) & (1) \\
d(\mathbf{p}_0, \boldsymbol{\gamma}(s)) < \tau_d & (2) \\
\measuredangle(\boldsymbol{t}_0, \boldsymbol{t}(s)) < \tau_\theta, & (3)
\end{cases}
$$

where $d(\mathbf{p}, \boldsymbol{\gamma}(s))$ is the Euclidean distance between a point $\mathbf{p}$ and the curve point $\boldsymbol{\gamma}(s)$.

The use of tangent orientation is significant since a reprojected curve on clutter would otherwise receive high support. This is confirmed by our experiments since performance drops when τ_θ is increased. Let the number of supporting edges of a curve point $\boldsymbol{\gamma}(s)$ in view v be denoted $\phi(\boldsymbol{\gamma}(s))$. Then, the total support S a hypothesis w_k has from

---

[1]Robert and Faugeras [24] (and later Schmid and Zisserman [25] using the Trifocal tensor) also use differential geometry in a trinocular setting, but these systems require extremely well-calibrated cameras, are restricted to 3 views, and employ additional heuristic constraints.

view v is

$$
S_v(w_k) = \int_{s=0}^{L_k^v} \phi(\boldsymbol{\gamma}^{k,v}(s))ds, \tag{4}
$$

where L_k^v is the length of $\boldsymbol{\gamma}^{k,v}$. A view that supports a curve above τ_v is considered a *supporting view*. We solicit evidence from more than one confirmation view for each curve fragment pair hypotheses (γ_1^1, γ_1^2),

$$
\mathcal{S}(w_k) = \sum_{v=3}^{N} [S_v(w_k) > \tau_v]\, S_v(w_k), \tag{5}
$$

by summing support over all supporting views.

The set of curve fragment pair hypotheses $\mathcal{W}$ is now thresholded for extent of support τ_t. However, observe that an image curve fragment γ_1^1 may pair with more than one curve fragment in view 2, due to epipolar ambiguity, *e.g.*, (γ_1^1, γ_1^2) and (γ_1^1, γ_2^2) may be both possible hypotheses, exceeding a threshold of support, but both cannot simultaneously be valid. Thus, we use a greedy assignment to resolve conflicts: the curve fragment pair hypotheses are rank-ordered and iteratively the highest-ranking hypothesis removes all conflicting lower-ranking hypotheses.

It should be emphasized that while this process may lead to numerous false negatives in forming curve fragment pair hypotheses (especially in the first stage when the threshold of support is higher for reliability *i.e.*, a curve fragment in one view may break up differently in a second view), this is not a major concern because the set of corresponding curve fragments $\{\gamma_{i_1}^1, \gamma_{i_2}^2, \ldots, \gamma_{i_n}^n\}$ need only be represented by one curve fragment pair. Thus, if a veridical curve fragment pair is not selected in hypothesis views 1 and 2, it is likely to be selected in another pair of hypotheses. In fact, the confirmed curve fragment pair hypotheses induces a grouping on the remaining views. This notion of interacting 3D curve reconstruction and edge linking is actually a significant departure from purely bottom-up edge linking approaches, but its development is beyond the scope of this paper.

While ideally the set of curve fragment pair hypotheses $\mathcal{W}$ should be formed from all pairs of hypothesis views and tested on all the remaining ones, this is not practical when the number of views is in tens or hundreds, as in a video sequence, thus motivating a view selection strategy. On one hand, reliability of reprojection onto a third view is directly related to how close the two views are: reprojections from two views with a small baseline are not stable. On the other hand, pairs of views with a large baseline may not have many image curves with a common source due to occlusion. Our experiments indicate that a baseline of $b_{min} = 40°$ leads to stability comparable to calibration accuracy. Beyond this value of baseline, the number of hypothesis views is only constrained by the extent of computations that can be afforded. We select pairs of consecutive views $b = 40°$ apart as hypothesis views. An identical strategy holds for selecting confirmation views. The only difference is that in the

confirmation process visibility is more of an issue than reconstruction stability. Thus, the selection follows a smaller baseline $b_c = 1/3b$, in the vicinity of the hypothesis views but excluding the selected hypothesis views.

3. Camera Calibration Refinement

The set of validated curve fragment pair hypotheses $\overline{\mathcal{W}}$

$$\overline{\mathcal{W}} = \{w_k = (\boldsymbol{\gamma}_{i_1}^{v_1}, \boldsymbol{\gamma}_{i_2}^{v_2}), k = 1, \ldots, K\}, \qquad (6)$$

together with camera parameters for all views $\mathcal{P} = (P_1, P_2, \ldots, P_N)$, reconstruct the 3D curve sketch $\{\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \ldots, \boldsymbol{\Gamma}_K\}$. Assuming the set of curve pair correspondences $\overline{\mathcal{W}}$ is correct, the curve-based reprojection error, *i.e.*, the degree the reprojected $\boldsymbol{\Gamma}_k$, $\boldsymbol{\gamma}^{k,v}$ agrees with its supporting edges in the edge map M_v defines the quality of the calibration. Formally, define the reprojection error of a 3D curve fragment $\boldsymbol{\Gamma}_k$ arising from views v_1 and v_2 as

$$f_{k,\mathcal{M}}(\mathcal{P}) = \sum_{\substack{v=1 \\ v \notin \{v_1, v_2\}}}^{N} \int_0^{L_u} d(\boldsymbol{\gamma}^{k,v}, M_v) ds, \qquad (7)$$

where $d(\boldsymbol{\gamma}_0^{k,v}(s), M_v)$ is the distance between the point $\boldsymbol{\gamma}_0^{u,v}(s)$ and the line containing the closest oriented edge point in M_v. Note that the closest edge point must satisfy the distance and orientation thresholds, τ_d and τ_θ, as in equations (2) and (3) respectively; oriented distance takes into account both distance and orientation and is defined as total reprojection error from the reliable, core 3D curve sketch as

$$f_{\overline{\mathcal{W}},\mathcal{M}}(\mathcal{P}) = \sum_{k=1}^{K} f_{k,\mathcal{M}}(\mathcal{P}). \qquad (8)$$

We then seek cameras that minimize the error

$$\hat{\mathcal{P}} = \operatorname*{argmin}_{\mathcal{P}} f_{\overline{\mathcal{W}}}(\mathcal{P}), \qquad (9)$$

The independence of curve projections from each other and the limited visibility of a curve across views implies a sparsity that allows for a successful implementation of the optimization using Levenberg-Marquardt as used in the traditional bundle adjustment. The total calibration error is reduced to subpixel accuracy, as illustrated in Figure 6.
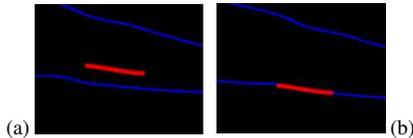

Figure 6: Example of a projected curve $\boldsymbol{\gamma}^{k,v}$(red) on edge map M_v(blue) before calibration refinement (a) and after (b).

4. Implementation

Edge and curve fragment detection: The integration of curve content over many views when clutter is present relies heavily on the orientation of curve tangents in addition to curve position. Thus, the edge detector used must yield very reliable orientation estimages. We use the subpixel edge map produced by a third-order operator [27], which explicitly addresses the orientation reliability, and their symbolic edge linker which addresses the grouping of edges into curve fragments. Examples are shown in Figure 7.

Initial calibration. Initial calibration is obtained using traditional automatic structure from motion techniques [21, 1], or if that fails, by manually specifying corresponding points and solving for the cameras. The initial calibration is typically imprecise (1-6 pixels reprojection error), which we call "coarse" calibration. We assume intrinsic parameters to be precisely known, with the imprecision residing on the extrinsic parameters. We are also in the process of developing a curve-based autocalibration system to augment point-based autocalibration using the ideas presented here.

Epipolar line intersection. The breakup of curve fragments at points of epipolar tangency implies that each segment can be parametrized by the epipolar angle. This allows for a rapid test of whether an epipolar line intersects a curve by simply looking up the two curve boundary epipolar angles. Furthermore, the intersection of an epipolar line and a given curve is reduced to simply checking which discrete parameter interval of the curve the epipolar line belongs to. The removal of points near epipolar tangency by insisting that all curve tangents have an angle of τ_α or more ($\tau_\alpha = 10°$) also allows for a stable intersection point computation.

Curve reprojection. The 3D curve fragment $\boldsymbol{\Gamma}_k$ is reprojected onto view v. We can also use an image to image transfer (*e.g.* the Trifocal transfer [13]), but we followed the simpler implementation.

The distance $d(\boldsymbol{\gamma}^{k,v}, M_v)$ uses the distance transform of edge maps, which is precomputed to find the closest edges to each point of the curve, but then a more accurate subpixel distance is recomputed for these edges.

5. Experimental Results

The Capitol Building Sequence consists of 500 frames covering a $90°$ helicopter fly-by from the Rhode Island State Capitol, using a High-Definition camera (1280×720). Intrinsic parameters were obtained from Matlab Calibration toolbox. Extrinsic parameters were obtained by manually marking/tracking 30 corresponding points on every view they appear in. This data is challenging due to a mix of curved and linear structures, a complex background, and structures of various sizes. **The Capitol Building High Sequence** consists of 256 frames coming from a $270°$ helicopter fly-by at higher altitude. Data objects of interest appear in lower resolution than in the regular Capitol sequence, leading to higher reconstruction errors.

The Downtown Sequence consists of 173 frames covering a $360°$ helicopter fly-by over downtown Providence, using a High-Def. camera (1280×720). The sequence was au-
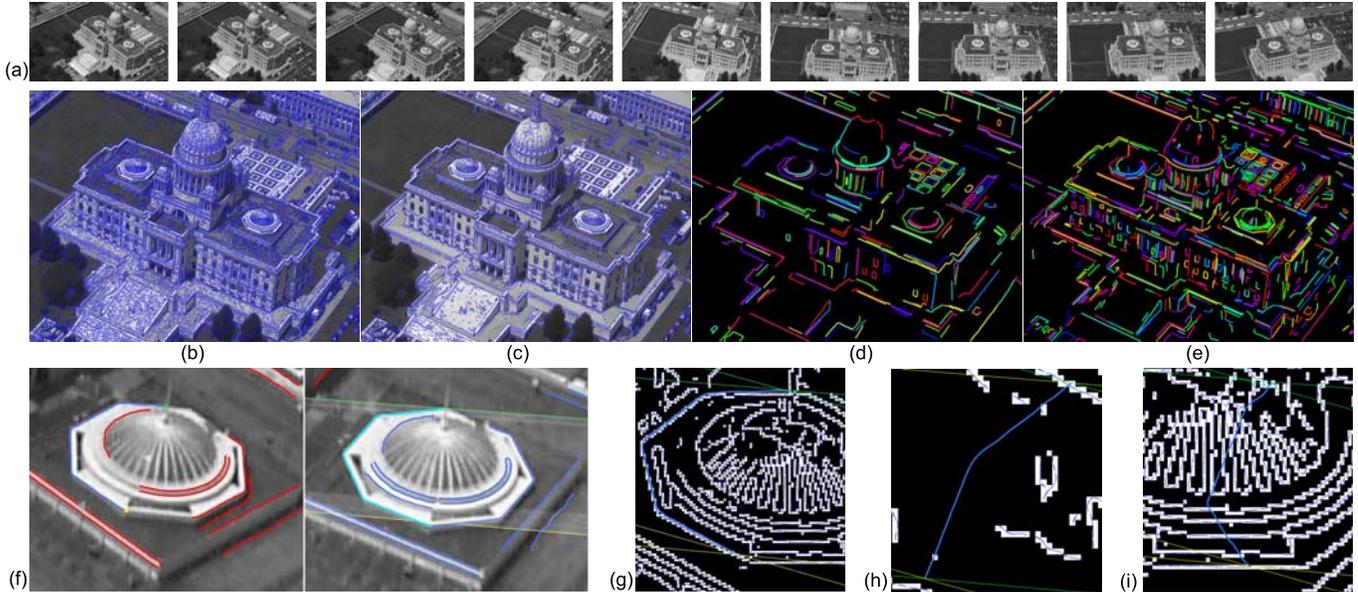
Figure 7: (a) A sample of views from the "Capitol Sequence"; (b) detected subpixel edgels at low gradient threshold; (c) detected subpixel edgels at a higher gradient threshold used in the confirmation views; (d) curve fragments with length > 40 pixels used in Stage I; (e) fragments with length > 20 pixels; (f) a selected curve is highlighted in view 1 and the curves in view 2 with sufficient epipolar overlap are marked in blue. In the next three images the reprojected hypothesis curve pair is shown in cyan, confirmed in (g) but rejected in (h) and (i). The example in (i) shows the need for orientation in the presence of clutter!

tomatically calibrated using Bundler [1], a SIFT-based structure from motion program. This data contains correlated and repeated structures, *e.g.*, a large number of long straight lines, easily confused with other lines due a to lack shape.

The Dinossaur Sequence from [26] was selected because it is a standard multiview-stereo dataset used by the computer vision community, even though it is not representative of the types of general scenery for which our approach can be most useful. The Cameras sample 363 views at 640×480 on a hemisphere around the object. The data is low resolution and contains a large number of occluding contours.

Ground truth for the 3D surfaces is only available for the simplest scenes (*e.g.* Dino). It is virtually impossible to get comprehensive ground-truth for the challenging real-world datasets tested in this paper. Even if such a ground truth were available, the evaluation of the system as a whole would confound the performance of the proposed approach with that of the edge detector and linker. Rather, we focus instead on evaluating the ability of the system to find the correct correspondence between curve fragments in two views, by manually recording the ground-truth correspondence between two views of each dataset.

A GUI tool was developed for this purpose to insert, delete, and inspect correpondences between two views. The views selected for this were such that they had a baseline near the minimum required for accurate 3D reconstruction for the given dataset. Curve fragments were detected on each hypothesis view and broken at epipolar tangencies. The human operator clicks on each curve fragment in view 1, and the GUI tool highlights the candidate curves in image 2 having epipolar overlap.

The operator is instructed to mark as ground truth correspondents *all* the curve fragments which come from the same underlying physical 3D structure, *i.e.*, having subsets coming from the same 3D positions. If there is any ambiguity in this selection, the operator is allowed to use every information possible to try to disambiguate the match: she is allowed to hypothesize a match in view 2 and the software will display reprojections for that match in all other confirmation views; she can also overlay the curve fragments onto the original images. The user can then gather evidence until the hypothesis is clearly decided. If all fails, the user can request to reconstruct the hypothesis and examine it in 3D; she can also reconstruct all the unambigous groundtruth marked so far and examine how well the reconstructed 3D curve hypothesis fits in the context of the already reconstructed ground truth. This kind of reasoning leaves little doubt as to what the corresponding curves should be. Ambiguity can still remain, such as a few very closeby and similarly shaped curves in a low-spatial frequency section of the image. In such cases, the user is allowed to mark two alternatives as equally good matches, but these form a small percentage of the total data. Positive matches are the pairs (γ_i^1, γ_j^2) marked by the human in the ground truth. Conversely, negative matches are all the pairs not marked in the ground truth.

Results. Figure 8 shows Precision/Recall curves for each dataset comparing the core 3D curve sketch to the ground truth. All datasets get 100% precision at $1/3$ recall, which is 40 curves for the Downtown, 65 curves for the Capitol se-

quence, and 30 curves for the Dino sequence. This is already plenty enough for calibration (much more than the usual 40 corresponding points needed for stable estimation [13]). Note that these numbers of core reliable curves are much higher if we pick the right operating point *per* dataset.

The role of differential geometry, or orientation, in this paper, is explored by elarging the orientation threshold to $90°$ so that it no longer plays a role. The plots on the right of Figure 8 show a significant degradation as the role of orientation is reduced. We expect a similar role for curvature.

Figure 9 shows the quality of the core 3D curve sketch which is obtained by integrating a large number of hypothesis views. This is best viewed in 3D (see supplentary data) to appreciate the 3D structure presented by the 3D curve sketch. Observe the potential of this curve sketch as a scaffold on which surface patches can be constructed.

6. Conclusion

We have presented a novel framework for multiview reconstruction and calibration refinement based on image curve content. The approach augments existing interest-point based and stereo approaches in providing explicit curve geometry as well in extending applications where the assumption of these methods fail but image curve content is present. A key capability is integration across many views, *e.g.*, as in Google's Street View. The present paper is expected to form the initial building block in a broader effort to use image evidence of the explicit geometry of curves and surfaces and reconstruct these by integrating information across many views. The 3D curve sketch presented here, when enriched by interpolating across epipolar gaps, will then be the initial scaffold on which surfaces may be constructed. The effort is also underway for curve-based auto calibration based on the ideas presented here.

Acknowledgements

References

[1] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. In *ICCV '09*.

[2] K. Astrom, R. Cipolla, and P. Giblin. Generalised epipolar constraints. *IJCV*, 33(1):51–72, 1999.

[3] R. Berthilsson, K. Åström, and A. Heyden. Reconstruction of general curves, using factorization and bundle adjustment. *IJCV*, 41(3):171–182, 2001.

[4] A. Bottino and A. Laurentini. The visual hull of smooth curved objects. *PAMI*, 26(12):1622–1632, 2004.

[5] R. Cipolla and P. Giblin. *Visual Motion of Curves and Surfaces*. Cambridge University Press, 1999.

[6] F. Cole, K. Sanik, D. DeCarlo, A. Finkelstein, T. Funkhouser, S. Rusinkiewicz, and M. Singh. How well do line drawings depict shape? In *SIGGRAPH*, Aug. 2009.

[7] D. Crispell, D. Lanman, P. Sibley, Y. Zhao, and G. Taubin. Shape from depth discontinuities. *LNCS*, 5416, 2009.

[8] S. De Ma. Conics-based stereo, motion estimation, and pose determination. *IJCV*, 10(1):7–25, 1993.

[9] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. In *CVPR'07*.

[10] Y. Furukawa and J. Ponce. Carved visual hulls for image-based modeling. *IJCV*, 81(1):53–67, 2009.

[11] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. Seitz. Multi-view stereo for community photo collections. In *ICCV '07*.

[12] M. Habbecke and L. Kobbelt. A surface-growing approach to multi-view stereo reconstruction. In *CVPR'07*.

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

[14] C. Hernandez, F. Schmitt, and R. Cipolla. Silhouette coherence for camera calibration under circular motion. *PAMI*, 29(2):343–349. 2007

[15] C. Hernández Esteban and F. Schmitt. Silhouette and stereo fusion for 3D object modeling. *CVIU*, 96(3):367–392. 2004

[16] V. Jain, B. B. Kimia, and J. L. Mundy. Segregation of moving objects using elastic matching. *CVIU*, 108:230–242. 2007

[17] F. Kahl and A. Heyden. Using conic correspondence in two images to estimate the epipolar geometry. In *ICCV'98*, page 761.

[18] J. Y. Kaminski and A. Shashua. Multiple view geometry of general algebraic curves. *IJCV*, 56(3):195–219. 2004

[19] 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. *PAMI*. 2007.

[20] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3d objects. *IJCV*, 73(3):263–284, 2007.

[21] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *IJCV*, 59(3):207–232, 2004.

[22] J. Porrill and S. Pollard. Curve matching and stereo calibration. *IVC*, 9(1):45–50, 1991.

[23] L. Reyes and E. Bayro Corrochano. The projective reconstruction of points, lines, quadrics, plane conics and degenerate quadrics using uncalibrated cameras. *IVC*, 23(8):693–706. 2005.

[24] L. Robert and O. D. Faugeras. Curve-based stereo: figural continuity and curvature. In *CVPR'91*.

[25] C. Schmid and A. Zisserman. The geometry and matching of lines and curves over multiple views. *IJCV*, 40(3):199–233, 2000.

[26] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR'06*.

[27] A. Tamrakar and B. B. Kimia. No grouping left behind: From edges to curve fragments. In *ICCV '07*.

[28] H. Wu and Y. Yu. Photogrammetric reconstruction of free-form objects with curvilinear structures. *The Visual Computer*, 21(4):203–216, 2005.
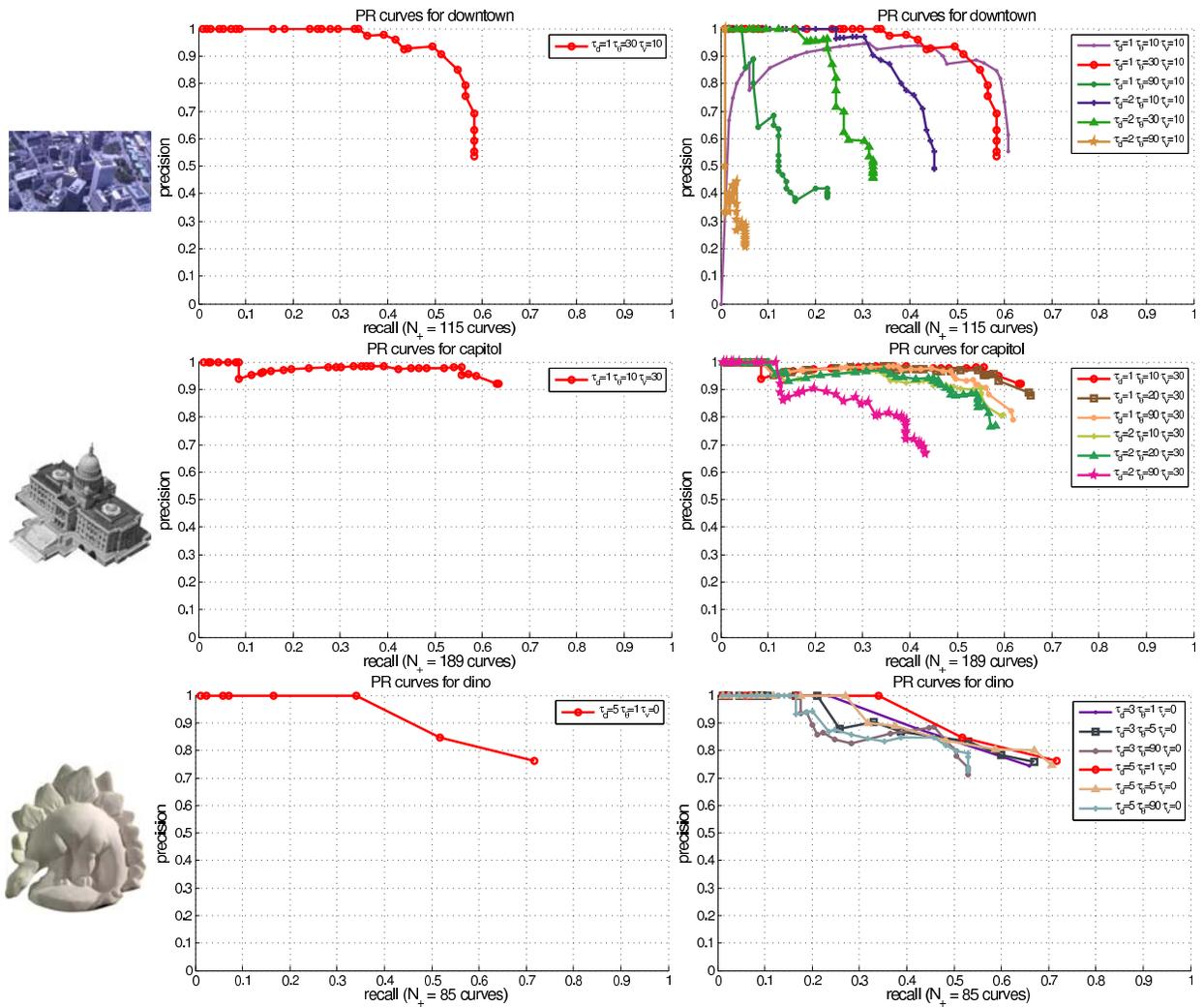
Figure 8: Precision/Recall curves for the core 3D curve sketch correspondence, obtained by varying the total support score τ_t. **Left:** the optimal evaluation curve shown in red; **Right:** varying the orientation threshold τ_θ to demonstrate the usefulness of the differential geometry constraint. Notice how eliminating orientation ($\tau_\theta = 90°$) degrades performance. These experiments used disambiguation threshold $\tau_r = 1.5$, length $\tau_l = 40px$ for the Downtown and Capitol sequences (hi-res), and $\tau_l = 20px$ for the Middlebury Dino sequence (low-res).
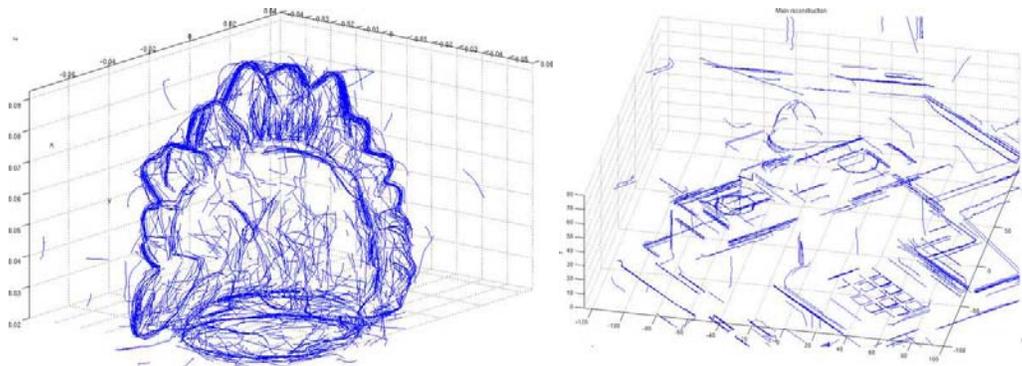

Figure 9: The 3D curve sketches for the Dino and Capitol sequences. See supplementary data for 3D results.

[29] H. Zhang, K. Wong, and G. Zhang. Camera calibration from images of spheres. *PAMI*, 29(3):499–502, 2007.