On Using Silhouettes for Camera Calibration

Edmond Boyer

MOVI - Gravir - INRIA Rhône-Alpes, Montbonnot, France
Edmond.Boyer@inrialpes.fr

Abstract. This paper addresses the problem of camera calibration using object silhouettes in image sequences. It is known that silhouettes encode information on camera parameters by the fact that their associated viewing cones should present a common intersection in space. In this paper, we investigate how to evaluate calibration parameters given a set of silhouettes, and how to optimize such parameters with silhouette cues only. The objective is to provide on-line tools for silhouette based modeling applications in multiple camera environments. Our contributions with respect to existing works in this field is first to establish the exact constraint that camera parameters should satisfy with respect to silhouettes, and second to derive from this constraint new practical criteria to evaluate and to optimize camera parameters. Results on both synthetic and real data illustrate the interest of the proposed framework.

1 Introduction

Camera calibration is a necessary preliminary step for most computer vision applications involving geometric measures. This includes 3D modeling, localization and navigation, am-ong other applications. Traditional solutions in computer vision are based on particular features that are extracted and matched, or identified, in images. This article studies solutions based on silhouettes which do not require any particular patterns nor matching or identification procedures. They represent therefore a convenient solution to evaluate and improve on-line a camera calibration, without the help of any specific patterns. The practical interest arises more specifically in multiple camera environments which are becoming common due, in part, to recent evolutions of camera acquisition materials. These environments require flexible solutions to estimate, and to frequently update, camera parameters, especially because often calibrations do not remain valid over time.

In a seminal work on motion from silhouettes, Rieger [1] used *fixed points* on silhouette boundaries to estimate the axis of rotation from 2 orthographic images. These fixed points correspond to epipolar tangencies, where epipolar planes are tangent to the observed objects' surface. Later on, these points were identified as *frontier points* in [2] since they go across the frontier of the visible region on a surface when the viewpoint is continuously changing. In the associated work, the constraint they give on camera motion was used to optimize essential matrices. In [3], this constraint was established as an extension of the traditional epipolar constraint, and thus was called the *generalized epipolar constraint*. Frontier points give constraints on camera motions, however they must first be localized on silhouette boundaries. This operation appears to be difficult:

P.J. Narayanan et al. (Eds.): ACCV 2006, LNCS 3851, pp. 1–10, 2006.

in [4] inflexions of the silhouette boundary are used to detect frontier points from which motion is derived, in [5] infinite 4D spaces are explored using random samples and in [6] contour signatures are used to find potential frontier points. All these approaches require frontier points to be identified on the silhouette contours prior to camera parameter estimation. However such frontier points can not be localized exactly without knowing epipoles. As a consequence, only approximated solutions are usually obtained by discrete sampling over a space of potential locations for frontier points or epipoles. We take a different strategy and bypass the frontier point localization by considering the problem globally over sets of silhouettes. The interest is to transform a computationally expensive discrete search into an exact, and much faster, optimization over a continuous space.

It is worth to mention also a particular class of shape-from-silhouette applications which use turntables and a single camera to compute 3D models. Such model acquisition systems have received noticeable attention from the vision community [7, 8, 9]. They are geometrically equivalent to a camera rotating in a plane around the scene. The specific constraints which result from this situation can be used to estimate all motion parameters. However, the associated solutions do not extend to general camera configurations as assumed in this paper.

Our approach is based first on the study of the constraint that both silhouettes and camera parameters must satisfy. We then derive two criteria: a quantitative smooth criterion in the form of a distance, and a qualitative discrete criterion, both being defined at any point inside a silhouette. This provides practical tools to qualitatively evaluate calibrations, and to quantitatively optimize their parameters. It appears to be particularly useful in multiple camera environments where calibrations often change, and for which fast on-line solutions are required.

This paper is organized as follows. Section 2 recalls background material. Section 3 precises constraints and respective properties of silhouettes, viewing cones and frontier points. Section 4 introduces the distance between viewing cones that is used as a geometric criterion. Section 5 introduces the qualitative criterion. Section 6 shows results on various data before concluding in section 7.

2 Definitions

Silhouette: Suppose that a scene, containing an arbitrary number objects, is observed by a set of pinhole cameras. Suppose also that projections of objects in the images are segmented and identified as foreground. $\mathcal{O}$ denotes then the set of observed objects and $\mathcal{I}_{\mathcal{O}}$ the corresponding binary foreground-background images. The foreground region of an image i consists of the union of objects' projections in that image and, hence, may be composed of several unconnected components with non-zero genus. Each connected component is called a *silhouette* and their union in image i is denoted $\mathcal{S}_i$.

Viewing Cone: Consider the set of viewing rays associated with image points belonging to a single silhouette in $\mathcal{S}_i$. The closure of this set defines a generalized cone in space, called *viewing cone*. The viewing cone's delimiting surface is tangent to the surface of the corresponding foreground object. In the same way that $\mathcal{S}_i$ is possibly

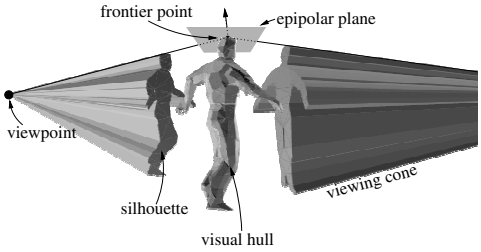

Fig. 1. A visual hull and 2 of its viewing cones

composed of unconnected components, the viewing cones of image i are possibly several distinct cones, one associated with each silhouette in $\mathcal{S}_i$. Their union is denoted $\mathcal{C}_i$. Note that individual objects are not distinguished here.

Visual Hull: The *visual hull* [10] is formally defined as the maximum surface consistent with all silhouettes in all images. Intuitively, it is the intersection of the viewing cones of all images (see figure 1). In practice, silhouettes are delimited by 2D polygonal curves, thus viewing cones are polyhedral cones and since a finite set of images are considered, visual hulls are polyhedrons. Assume that all objects are seen from all image viewpoints then:

$$\mathcal{VH}(\mathcal{I}_\mathcal{O}) = \bigcap_{i \in \mathcal{I}_\mathcal{O}} \mathcal{C}_i, \tag{1}$$

is the visual hull associated with the set $\mathcal{I}_\mathcal{O}$of foreground images and their viewing cones $\mathcal{C}_{i \in \mathcal{I}_\mathcal{O}}$. If all objects $\mathcal{O}$ do not project onto all images, then the reasoning that follows still applies to subset of objects and subsets of cameras which satisfy the common visibility constraint.

3 Geometric Consistency Constraint

In this section, the exact and optimal geometric consistency which applies with silhouettes is first established and its equivalence with more practical constraints is discussed.

3.1 Visual Hull Constraint

Calibration constraints are usually derived from geometric constraints reflecting geometric coherence. For instance, different image projections of the same feature should give rise to the same spatial location with true camera parameters. In the case of silhouettes, and under the assumption that no other image primitives are available, the only geometric coherence that applies comes from the fact that all viewing cones should correspond to the same objects with true camera parameters. Thus:

$$\mathcal{O} \subset \mathcal{VH}(\mathcal{I}_\mathcal{O}),$$

and consequently by projecting in any image i:

$$S_i \ \subset \ P_i(\mathcal{VH}(\mathcal{I}_\mathcal{O})), \ \forall i \in \mathcal{I}_\mathcal{O},$$

where $P_i()$ is the oriented projection[1] in image i. Thus, viewing cones should all intersect, and viewing rays belonging to viewing cones should all contribute to this intersection. The above expression is equivalent to:

$$\bigcup_{i \in \mathcal{I}_\mathcal{O}} [\mathcal{S}_i \; - \; P_i(\mathcal{VH}(\mathcal{I}_\mathcal{O}))] = \emptyset, \tag{2}$$

which says that the visual hull projection onto any image i should entirely cover the corresponding silhouette $\mathcal{S}_i$ in that image. This is the constraint that viewing cones should satisfy with true camera parameters. It encodes all the geometric consistency constraints that apply with silhouettes and, as such, is optimal. However this expression in its current form does not yield a practical cost function for camera parameters since all configurations leading to an empty visual hull are equally considered, thus making convergence over cost functions very uncertain in many situations. To overcome this difficulty, viewing cones can be considered pairwise as explained in the following section.

3.2 Pairwise Cone Tangency

We can easily derive from the general expression (2) the pairwise tangency constraint. Substituting the visual hull definition (1) in (2):

$$(2) \; \Leftrightarrow \; \bigcup_{i \in \mathcal{I}_\mathcal{O}} [\mathcal{S}_i \; - \; P_i(\bigcap_{j \in \mathcal{I}_\mathcal{O}} \mathcal{C}_j)] = \emptyset.$$

Since projection is a linear operation preserving incidence relations:

$$(2) \; \Rightarrow \; \bigcup_{i \in \mathcal{I}_\mathcal{O}} [\mathcal{S}_i \; - \; \bigcap_{j \in \mathcal{I}_\mathcal{O}} P_i(\mathcal{C}_j)] = \emptyset.$$

Note that, in the above expression, the exact equivalence with (2) is lost since projecting viewing cone individually introduces depth ambiguities and, hence, does not ensure a common intersection of all cones as in (2). By distributive laws:

$$(2) \; \Rightarrow \; \bigcup_{(i,j) \in \mathcal{I}_\mathcal{O} \times \mathcal{I}_\mathcal{O}} [\mathcal{S}_i \; - \; P_i(\mathcal{C}_j)] = \emptyset. \tag{3}$$

Expression (3) states that all viewing cones of a single scene should be pairwise tangent. By pairwise tangent, it is meant that all viewing rays from one cone intersect the other cone, and reciprocally. This can be seen as the extension of the epipolar constraint to silhouettes (see figure 2). Note that this constraint is always satisfied by concentric viewing cones, for which no frontier points exist. Note also that if (3) and (2) are not strictly equivalent, they are equivalent in most general situations.

[1] i.e. a projection such that there is a one-to-one mapping between rays from the projection center and image points.

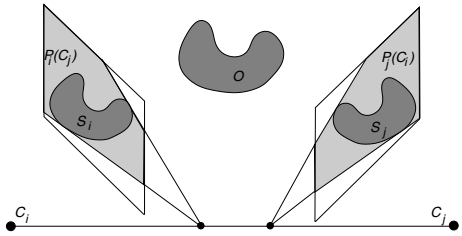

Fig. 2. Pairwise tangency constraint: silhouette $\mathcal{S}_i$ is a subset of the viewing cone projection $P_i(\mathcal{C}_j)$ in image i

3.3 Connection with Frontier Points

A number of approaches consider frontier points and the constraints they yield on camera configurations. Frontier points are particular points which are both on the objects' surface and the visual hull, which project onto silhouettes in 2 or more images, and where the epipolar plane is tangent to the surface (see figure 1). They satisfy therefore what is called the generalized epipolar constraint [3]. They allow hereby projective reconstruction when localized in images [5, 6]. The connection between the generalized epipolar constraint and the pairwise tangency constraint (3) is that the latter implies the former at particular frontier points. Intuitively, if two viewing cones are tangent then the generalized epipolar constraint is satisfied at extremal frontier points where viewing lines graze both viewing cones.

4 Quantitative Criterion

The pairwise tangency is a condition that viewing cones must satisfy to ensure that the same objects are inside all cones. In this section, we introduce a distance function that evaluates this condition.

4.1 Distances Between a Viewing Ray and a Viewing Cone

The distance function between a ray and a cone that we seek should preferably respect several conditions:

1. It should be expressed in a fixed metric with respect to the data, thus in the images since a 3D metric will change with camera parameters.
2. It should be a monotonic function of the respective locations of ray and cone.
3. It should be zero if the ray intersect the viewing cone. This intersection, while apparently easy to verify in the images, requires some care when epipolar geometry is used. Figure 3 depicts for instance a few situations where the epipolar line of a ray intersects the silhouette, though the ray does not intersect the viewing cone. These situations occur because no distinction is made between front and back of rays.
4. It should be finite in general so that situations in figure 3 can be differentiated.

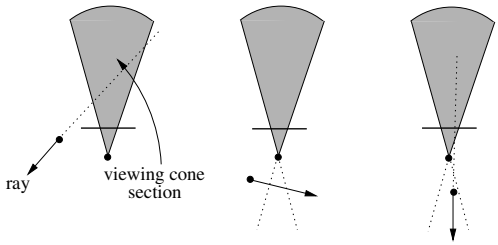

Fig. 3. A ray and the cross-section of the viewing cone in the corresponding epipolar plane. 3 of the situations where unoriented epipolar geometry will fail and detect intersections.

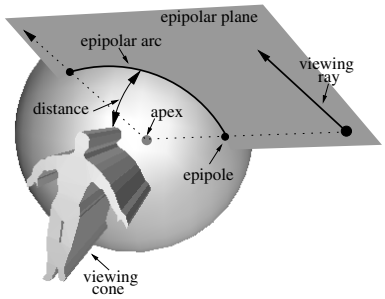

Fig. 4. The spherical image model: viewing rays project onto epipolars arcs on the sphere

In light of this, a fairly simple but efficient approach is to consider a spherical image model instead of a planar model (see figure 4), associated to an angular metric. The distance from a ray to a viewing cone is then the shortest path on the sphere from the viewing cone to the ray projection. This projection forms an epipolar circle-arc on the sphere delimited by the epipole and the intersection of the ray direction with the sphere. The ray projection is then always the shortest arc between these 2 points, which can coincide if the ray goes trough the viewing cone apex. Two different situations occur depending on the respective positions of the ray epipolar plane and the viewing cone:

1. The plane intersects the viewing cone apex only, as in figure 4. The point on the circle containing the epipolar arc and closest to the viewing cone must be determined. If such point is on the epipolar arc then the distance we seek is its distance to the viewing cone. Otherwise, it is the minimum of the distances between the arc boundary points and the viewing cone.
2. The plane goes through the viewing cone. The distance is zero in the case where the ray intersects the viewing cone section in the epipolar plane, and the shortest distance between the epipolar arc boundary points and the viewing cone section in the other case. This distance is easily computed using angles in the epipolar plane.

4.2 Distance Between 2 Viewing Cones

A distance function between a ray and a viewing cone has been defined in the previous section, this section discusses how to integrate it over a cone. The distance between

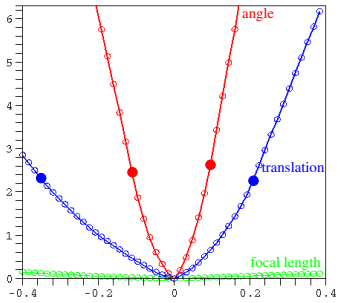

Fig. 5. The distance between 2 viewing cones as a function of: (green) one focal length which varies in the range $[f-0.4f, f+0.4f]$, with f the true value; (blue) one translation parameter to which is added from -0.4 to 0.4 of the camera-scene distance; (red) one Euler orientation angle which varies in the range $[\alpha-0.4\pi, \alpha+0.4\pi]$ with α the true value. The filled points denote the limit distances on curves above which the 2 cones do not intersect at all.

2 viewing cones is then simply defined by a double integration over the 2 concerned cones.

Recall that silhouettes and viewing cones are discrete in practice and thus defined by sets of contour points in the images and boundary rays in space. The simplest solution consists then in summing individual distances over boundary rays. Assume that r_i^k is the k^{th} ray on the boundary of viewing cone C_i, and $d(r_i^k, C_j) = d_{ij}^k$ is the distance between r_i^k and $\mathcal{C}_j$ as defined in the previous section. Then the distance D_{ij} between $\mathcal{C}_i$ and $\mathcal{C}_j$ is:

$$D_{ij} = \sum_k d_{ij}^k + \sum_l d_{ji}^l = d_{ij} + d_{ji}. \tag{4}$$

Remark that $D_{ij} = D_{ji}$ but $d_{ij} \neq d_{ji}$. The above expression is easy to compute once the distance function is established. It can be applied to all boundary viewing rays, however mainly rays on the convex hulls of silhouettes are concerned by the pairwise tangency constraint, we thus consider only them to improve computational efficiency. Figure 5 illustrates the distance D_{ij} between 2 viewing cones of a synthetic body model as a function of various parameters of one cone's camera. This graph demonstrates the smooth behavior of the distance around the true parameter values, even when the cones do not intersect at all.

5 Silhouette Calibration Ratio

Following the quantitative criterion, we introduce a simple qualitative criterion which evaluates how silhouettes contribute to the visual hull for a given calibration.

Recall that any viewing ray, from any viewing cone, should be intersected by all other image viewing cones, along an interval common to all cones. Let ω_r be an interval along ray r intersected by viewing cones, and let us call $\mathcal{N}(\omega_r)$ the number of image contributing (image for which a viewing cone intersects ω_r) inside that interval. Then

the sum over the rays r: $\sum_r \max_{\omega_r}(\mathcal{N}(\omega_r))$, should theoretically be equal to $m(n-1)$ if m rays and n images are considered. Now this criterion can be refined by considering each image contribution individually along a viewing ray. Let ω_r^i be an interval, along ray r, where image i contributes. Then the silhouette calibration ration C_r defined as:

$$C_r = \frac{1}{m(n-1)^2} \sum_r \sum_i \max_{\omega_r^i}(\mathcal{N}(\omega^i)), \tag{5}$$

should theoretically be equal to 1 since each image should have at least one contribution interval with $(n-1)$ image contributions. This qualitative criterion is very useful in practice because it reflects the combined quality of a set of silhouettes and of a set of camera parameters. Notice however that it can hardly be used for optimizations because of its discrete, and thus non-smooth, nature.

6 Experimental Results

The pairwise tangency presented in the previous section constraint camera parameters when a set of static silhouettes $\mathcal{I}_\mathcal{O}$ is known. For calibration, different sets $\mathcal{I}_\mathcal{O}$ should be considered. They can easily be obtained, from moving objects for instance, as in [5]. The distances between viewing cones are then minimized over the camera parameter space through a least square approach:

$$\hat{\theta}_{\mathcal{I}_\mathcal{O}} = \min_\theta \sum_{(i,j) \in \mathcal{I}_\mathcal{O} \times \mathcal{I}_\mathcal{O}} D_{ij}^2, \tag{6}$$

where θ is the set of camera parameters to be optimized. $\hat{\theta}_{\mathcal{I}_\mathcal{O}}$ is equivalent to a maximum likelihood estimate of the camera parameters under the assumption that viewing rays are statistically independent. The above quantitative sum can be minimized by standard non-linear methods such as Levenberg-Marquardt.

6.1 Synthetic Data

Synthetic sequences, composed of images with dimensions 300×300, were used to test the approach robustness. 7 cameras, with standard focal lengths, are viewing a running human body. All camera extrinsic parameters and one focal length per camera, assuming known or unit aspect ratios, are optimized. Different initial solutions are tested by adding various percentages of uniform noise to the exact camera parameters. For the focal lengths and the translation parameters, the noise amplitudes vary from 0% up to 40% of the exact parameter value; for the pose angle parameters, the noise amplitudes vary from 0% up to 40% of 2π. Figure 6 shows, on the left, the silhouette calibration ratios after optimization; and on the right, relative errors in the estimated camera parameters after optimization using 5 frames per cameras. These results first validate the silhouette calibration ratio as a global estimator for the quality of any calibration with respect to silhouette data. Second, they show that using only one frame per camera is intractable in most situations. However, they prove also that using several frames, calibration can be recovered with a good precision even far from the exact solution. Other

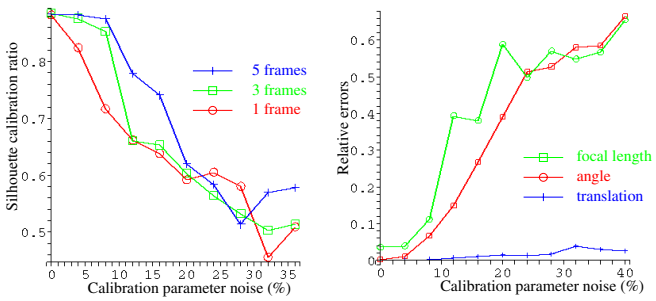

Fig. 6. Robustness to the initial calibration: right, the silhouette calibration ratio; left, the relative errors in the estimated camera parameters for the 5 frame case: errors relative to the true value for the focal length, errors relative to the distance camera-scene for the translation parameter and errors relative to π for the angle parameter

experiments, not presented due to lack of space, show that adding a reasonable amount of noise to silhouette vertices, typically a 1 pixel Gaussian Noise, only slightly changes these results.

6.2 Real Data

Our approach was also tested in a real environment with 6 firewire cameras viewing a moving person. A calibration obtained by optimizing an initial solution using known points is available and will be considered as the ground truth. In the following experi-

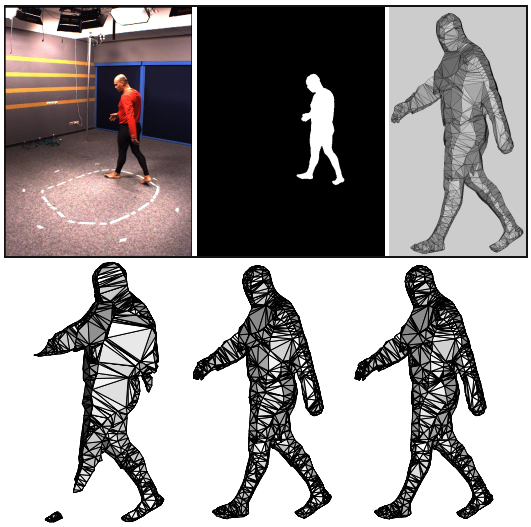

Fig. 7. Top, one of the original image, the corresponding silhouette and the visual hull model obtained with ground truth calibration. Bottom, 3 models which correspond to calibrations obtained with our method and using respectively 1, 3 and 5 frames per camera.

ments, we use the same initial solution for the calibration with viewing cones. As for the synthetic case, all camera extrinsic parameters and one focal length per camera are optimized. Figure 7 shows, on top, the input images and a visual hull model obtained using ground truth values for calibration. In the bottom, models obtained from the same silhouettes, but using our approach with respectively 1, 3 and 5 frames per camera. Apart from a scale difference, not shown and due to the fact that fixed dimensions were imposed for the ground truth solution, the 2 most-right models are very close to the ground truth one.

7 Conclusion

We have studied the problem of estimating camera parameters using silhouettes. It has been shown that, under little assumptions, all geometric constraints given by silhouettes are ensured by the pairwise tangency constraint. A second contribution of this paper is to provide a practical criterion based on the distance between 2 viewing cones. This criterion appears to be efficient in practice since it can handle a large variety of camera configurations, in particular when viewing cones are distant. It allows therefore multi-camera environments to be easily calibrated when an initial solution exists. The criterion can also be minimized using efficient and fast non-linear approach. The approach is therefore also aimed at real time estimation of camera motions with moving objects.

References

1. Rieger, J.: Three-Dimensional Motion from Fixed Points of a Deforming Profile Curve. Optics Letters **11** (1986) 123–125
2. Cipolla, R., strm, K., Giblin, P.: Motion from the Frontier of Curved Surfaces. In: Proceedings of 5th International Conference on Computer Vision, Boston (USA). (1995) 269–275
3. Åström, K., Cipolla, R., Giblin, P.: Generalised Epipolar Constraints. In: Proceedings of Fourth European Conference on Computer Vision, Cambridge, (England). (1996) 97–108 Lecture Notes in Computer Science, volume 1065.
4. Joshi, T., Ahuja, N., Ponce, J.: Structure and Motion Estimation from Dynamic Silhouettes under Perspective Projection. In: Proceedings of 5th International Conference on Computer Vision, Boston (USA). (1995) 290–295
5. Sinha, S., Pollefeys, M., McMillan, L.: Camera Network Calibration from Dynamic Silhouettes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington, (USA). (2004)
6. Furukawa, Y., Sethi, A., Ponce, J., Kriegman, D.J.: Structure from Motion for Smooth Textureless Objects. In: Proceedings of the 8th European Conference on Computer Vision, Prague, (Czech Republic). (2004)
7. Fitzgibbon, A., Cross, G., Zisserman, A.: Automatic 3d model construction for turn-table sequences. In: Proceedings of SMILE Workshop on Structure from Multiple Images in Large Scale Environments. Volume 1506 of Lecture Notes in Computer Science. (1998) 154–170
8. Mendoça, P., Wong, K.Y., Cipolla, R.: Epipolar Geometry from Profiles under Circular Motion. IEEE Transactions on PAMI **23** (2001) 604–616
9. Jiang, G., Quan, L., Tsui, H.: Circular Motion Geometry Using Minimal Data. IEEE Transactions on PAMI **26** (2004) 721–731
10. Laurentini, A.: The Visual Hull Concept for Silhouette-Based Image Understanding. IEEE Transactions on PAMI **16** (1994) 150–162