Camera Calibration Using Silhouettes

Edmond Boyer

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Camera Calibration Using Silhouettes

Edmond Boyer

N° 5559

Avril 2005

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--5559--FR+ENG

Camera Calibration Using Silhouettes

Edmond Boyer

Thème COG — Systèmes cognitifs
Projets Movi

Rapport de recherche n° 5559 — Avril 2005 —18 pages

Abstract: This report addresses the problem of estimating camera parameters from images where object silhouettes only are known. Several modeling applications make use of silhouettes, and while calibration methods are well known when considering points or lines matched along image sequences, the problem appears to be more difficult when considering silhouettes. However, such primitives encode also information on camera parameters by the fact that their associated viewing cones should present a common intersection in space. In this paper, we investigate the problem both on the theoretical and practical viewpoint. In particular, we clarify why, and how, a set of image silhouettes of the same scene give constraints on camera parameters; these constraints justifying a calibration approach. The main contributions of this paper with respect to existing works is first to establish the optimal criterion that camera parameters should satisfy with respect to silhouettes, and second to provide a practical approach based on this criterion. Results on both synthetic and real data are shown to give insights into the method potential for camera calibration.

Key-words: Camera Calibration, silhouettes, multi-camera environments.

Unité de recherche INRIA Rhône-Alpes

Calibrage de caméras à partir de silhouettes

Résumé : Ce document traite du calibrage de caméras à partir de silhouettes images. De nombreuses applications, en vision par ordinateur, utilisent les silhouettes pour construire des modèles tri-dimensionels, mais très peu le font pour calibrer des caméras. Le problème reste en effet difficile, et peu de solutions fiables existent. Pourtant, ces primitives fournissent de l'information sur les paramètres des caméras. Leurs cônes de vues doivent en effet présenter une intersection commune dans l'espace. Dans ce document, le problème est traité d'un point vue théorique aussi bien que pratique. En particulier, nous clarifions pourquoi, et comment, un ensemble de silhouettes d'une même scêne fournissent des contraintes sur les paramètres des caméras; cela justifiant une méthode de calibrage. Les principales contributions sont, dans un premier temps, d'établir un critère optimal que les paramètres des caméras doivent respecter en lien avec un ensemble de silhouettes puis, dans un deuxième temps, de proposer une approche pratique de calibrage basée sur ce critère. L'intérêt de cette approche est demontré par des résultats expérimentaux obtenus avec des données synthétiques et réèlles.

Mots-clés : Calibrage de caméras, silhouettes, environnement multi-caméras.

1 Introduction

Camera calibration is a necessary preliminary step for most computer vision applications involving geometric measures. This includes 3D modeling, localization and navigation, among other applications. Traditional solutions in computer vision are based on particular features that are extracted and matched, or identified, in images. This articles 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 [Rieger86] 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 [Cipolla95] 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 [Åström96], 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 be first localized on silhouette boundaries. This operation appears to be difficult and in [Joshi95], inflexions of the silhouette boundary are first used to detect frontier points, both features being used afterward to estimate motion. Interestingly, an attempt is made in [Bottino03] to bypass calibration and to estimate shapes directly from orthographic silhouettes with unknown viewpoints, however no practical solution is given.

All the above approaches assume static scenes, and thus consider only a few silhouettes as input data. In contrast, it was recently proposed in [Sinha04] to consider moving silhouette sequences in order to calibrate sets of static cameras. In this work, the generalized epipolar constraint is used to detect frontier points. A projective to Euclidean reconstruction of these points is then plugged into a bundle adjustment to optimize camera parameters. The clear interest of this approach is to provide a solution for frontier point localization. However, its complexity remains high since numerous infinite 4D spaces must be explored using random samples. Nevertheless, our approach is complementary to it.

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 [Fitzgibbon98, Mendoça01, Jiang04]. 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.

Camera calibration approaches usually proceed in two steps: an initial solution is first found for camera parameters, and this solution is then optimized through geometric constraints. In this paper, we assume that an initial solution is given, an old calibration or a guess from the user for

instance, and we focus on the second step in the particular case of silhouette features. We first identify constraints that image silhouettes provide on camera configurations. From this, an optimal calibration criterion is derived. This criterion does not require frontier point localization as most of the afore mentioned approaches. Our contribution with respect to existing works is twofold: first, geometric constraints concerning silhouettes are presented in a unified way and their relations are explicited; second we propose a practical scenario to evaluate, and improve when necessary, any calibration using extracted silhouettes in video sequences.

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 explains the implementation setup and shows results on various data before concluding. Note that only the main results are stated in this paper and the details of the proofs are left for a future paper.

2 Definitions

In this section, we give definitions for the different geometric entities that are manipulated throughout the paper.

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 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 viewing cones or silhouettes of the same image are not distinguished here, this because the principles presented in this paper apply with one or several objects in the scene.

Visual Hull The *visual hull* [Baumgart75, Laurentini94] 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). It can also be seen as the closure of all points in space which project inside all silhouettes. 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:

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

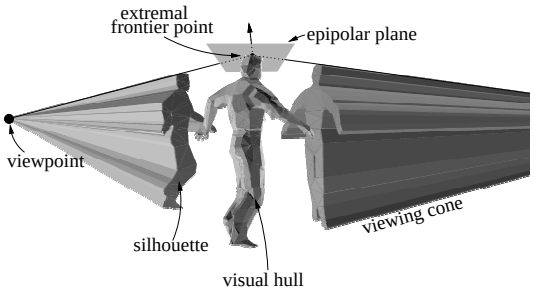

Figure 1: A visual hull and 2 of its viewing cones.

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 visibility constraint.

Frontier points are particular points which are both on the objects' surface and the visual hull, and where viewing cones intersect. At such points, the epipolar plane is tangent to the surface (see figure 1). Their interest for calibration comes from the fact that they represent 3D scene points which project onto at least 2 silhouette boundaries. In the following, we will distinguish subsets of frontier points, **extremal frontier points**, which belong to the convex hull boundaries of viewing cones. A non-extremal frontier points is then a frontier point which is possibly on a convex region of the viewing cone but not on the convex hull boundary of this viewing cone.

3 Geometric consistency constraint

In this section, formal considerations are made on constraints that a set of static silhouettes gives on camera configurations, through viewing cones and visual hulls. The exact and optimal geometric consistency which applies with silhouettes is first established and its equivalence with more practical constraints is then shown.

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

From the general expression (2), we can easily derive and explicit 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.

3.3 Connection with frontier points

A number of approaches consider frontier points and the constraints they yield on camera configurations. Frontier points are surface points which project onto silhouettes in 2 or more images, and

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

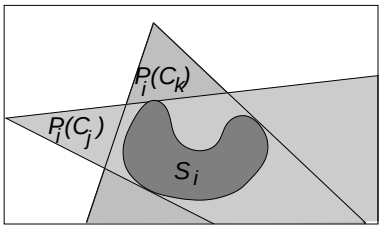

Figure 2: Pairwise tangency constraint: silhouette $\mathcal{S}_i$ is a subset of any viewing cone projection in image i.

where the epipolar plane is tangent to the surface. They satisfy therefore what is called the generalized epipolar constraint [Åström96]. They allow hereby projective reconstruction when localized in images [Sinha04]. The main limitation of frontier points comes from their localizations in images, which appear to be difficult and time consuming since their positions and numbers change with epipolar geometry. We study here how they relate to expression (3), the interest being that this expression handle them implicitly and, hence, do not require any localization.

Property 1 *Pairwise tangency constraints on viewing cones of objects implies generalized epipolar constraints on extremal frontier points of these objects.*

This is a consequence of the fact that if the epipolar constraint is not satisfied at an extremal frontier point, then one viewing cone is necessarily *above* the other at that point, breaking the tangency constraint. For non-extremal frontier points, the epipolar constraint can not be checked in general, except when the visibility from more than one viewpoint can be ensured. This eliminates most non-extremal frontier points, in particular hyperbolic frontier points which are hidden by T-junctions. Particular situations where the visibility can be ensured for non-extremal frontier points occur when camera centers project outside a silhouette but inside its convex hull. These situations are however not frequent, and even in that case the epipolar constraint and the pairwise tangency constraint both hold. The reciprocal is also true and the generalized epipolar constraint implies the pairwise tangency constraint. It can be seen as a consequence of the epipolar map properties. This map associates epipolar correspondents on silhouette boundary contours, or rays between viewing cone boundaries. It is known to be smooth between frontier points except at contour singularities [Åström96, Boyer97] such as cusps and T-junctions. Hence:

Property 2 *Pairwise tangency constraints on viewing cones of objects and generalized epipolar constraints on frontier points of these objects are equivalent.*

On the practical viewpoint, this means that pairwise tangency constraints on viewing cones can be used as a matching constraint instead of generalized epipolar constraints, and without the need to detect frontier points.

3.4 Optimality of the Pairwise tangency constraint

An important remaining question is the equivalence of the pairwise tangency constraint (3) with the optimal visual hull constraint (2). Such an equivalence is not necessarily evident since the pairwise tangency constraint ensures that all viewing cones contribute along any silhouette viewing ray, but not that the resulting contribution segments intersect, which is the case with constraint (2). To verify this equivalence, let us first distinguish the two following situations:

1. A single convex object is observed. Under the assumption that viewing cones which are pairwise tangent have a common intersection, and since in that particular case there is a single contribution per viewing cone on a silhouette viewing ray, all these contributions necessarily intersect. Consequently expressions (3) and (2) are equivalent.

2. Several, and possibly non-convex objects, are observed. In that case, the above reasoning applies to the convex hull of the scene which projects onto the convex hulls of the silhouettes in the images. Now to ensure that along any viewing rays there is a common contribution from all viewing cones, we not only need to assume, as before, that all convex viewing cones intersect, but also that the pairwise tangency constraint on the scene's convex hull is sufficient to ensure incidence relations inside the convex hull. This is the case if extremal frontier points are in sufficient number on the scene's convex hull to allow for a projective reconstruction.

The number of extremal frontier points is in general sufficient if at least 3 cameras and a reasonably complex scene are considered. To this purpose, notice that a frontier point that projects onto n images gives $2n - 3$ constraints on the camera configuration, which corresponds in general to 1 constraint unless more than 2 viewpoints are coplanar with the frontier point. Thus, expressions (2) and (3) are, in general, equivalent under the assumption that viewing cones which are pairwise tangent have a common intersection. Conditions under which this assumption holds are identified in the following.

Single intersection of viewing cones
As mentioned earlier, pairwise tangency is a two by two constraint and, as such, does not imply that all viewing cones, or their convex hulls, intersect at a common location and, hence, give rise to a visual hull. An intuitive example of that is shown in figure 3 with viewing cones reduced to single rays for simplicity.

Remark that situations such as figure 3-(b) requires viewing rays to be coplanar. In a similar way, extending to viewing cones with more than one ray shows that ambiguous situations appear when extremal frontier points and viewpoints are coplanar:

Property 3 3 *or more viewing cones which are pairwise tangent intersect if viewpoints and extremal frontier points are not all coplanar.*

Ambiguous situations appear therefore if both viewpoints and extremal frontier points are coplanar. In practice, this situation occur only with orthographic cameras having coplanar optical axis. Thus, in the general case and with perspective cameras, pairwise tangency will define a visual hull. It should however be noticed that situations with coplanar perspective cameras having long focal lengths can be sensitive, and such situations can appear with turntable acquisition systems. Finally:

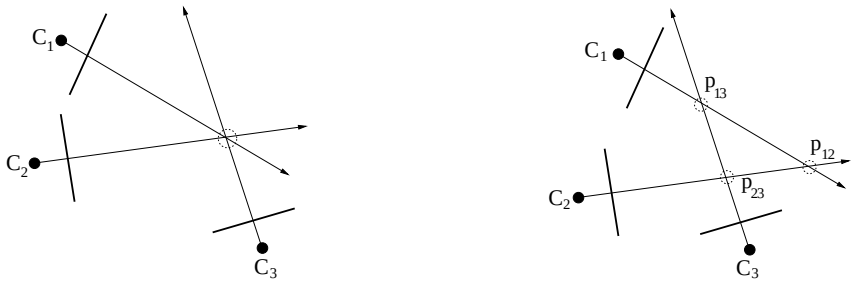

Figure 3: Example where single ray viewing cones are pairwise tangent in both situations (a) and (b), but do not define a visual hull in (b).

Property 4 *The pairwise tangency constraint is equivalent to the visual hull constraint if viewpoints and extremal frontier points are not coplanar and if extremal frontier points are in sufficient number.*

Consequently, the pairwise tangency constraint is optimal with respect to the given data -silhouettes- and under the mentioned assumptions.

4 Viewing cone distance

It was shown previously that pairwise tangency is an optimal condition that viewing cones must satisfy to ensure that the same objects are inside all cones. This condition can be used to constraint camera parameters when seeking for calibration. This can be achieved in a classical way, through a ML-least squares estimation for instance, if a distance function between viewing cones is defined. This is the purpose of this section which proposes first a distance function between a viewing ray and a viewing cone, and second explicits how to integrate this function over cones to form a viewing cone distance.

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 4 depicts

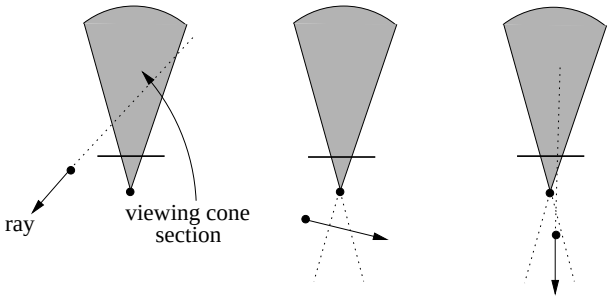

Figure 4: A ray and the cross-section of the viewing cone in the corresponding epipolar plane. 3 of the situations where unoriented epipolar geometry will fail and detect intersections.

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 4 can be differentiated.

In light of this, a fairly simple but efficient approach is to use a spherical image model as shown in figure 5, 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:

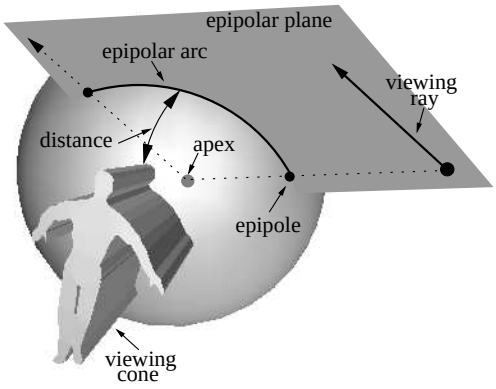

Figure 5: The spherical image model: viewing rays project onto epipolars arcs on the sphere.

1. The plane intersects the viewing cone apex only, as in figure 5. 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.

A few remarks are in order:

- As mentioned earlier, distances should be expressed in a fixed metric such as image pixels. This is not the case if angles between rays are considered on a unit sphere since this sphere varies with camera intrinsic parameters. This can be partially compensated by considering the image space as an ellipsoid with focal lengths as axis lengths.

- The above reasoning could also be implemented using conventional projective geometry, not however that a proper algebra can be established with the help of oriented projective geometry [Stolfi91], which preserves orientation.

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 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 6 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.

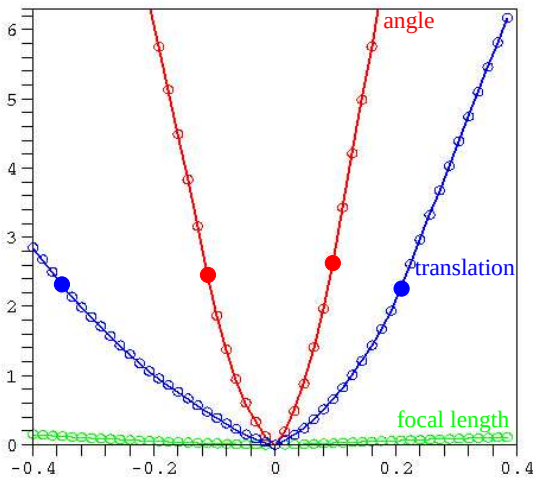

Figure 6: The distance between 2 viewing cones as a function of: (green) one focal length which varies in the range $[f - 0.4f, f + 0.4f]$, with f the true value; (blue) one translation parameter to which is added from -0.4 to 0.4 of the camera-scene distance; (red) one Euler orientation angle which 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.

5 Implementation and 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 [Sinha04]. The distances between viewing cones are then minimized over the camera parameter space through a least square approach:

$$\hat{\theta}_{\mathcal{I}_\mathcal{O}} = \min_{\theta} \sum_{\mathcal{I}_\mathcal{O}} \sum_{\mathcal{C}_i, \mathcal{C}_j, r_i^k} (d_{ij}^k)^2, \tag{5}$$

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. As mentioned before, only viewing rays on the silhouette convex hulls are considered. The above quantitative sum can be minimized by standard non-linear methods such as Levenberg-Marquardt. On the other hand, it does not give qualitative insights on the estimated parameters. To this purpose, we propose a simple qualitative criterion which evaluates how well silhouettes contribute to the visual hull for a given calibration.

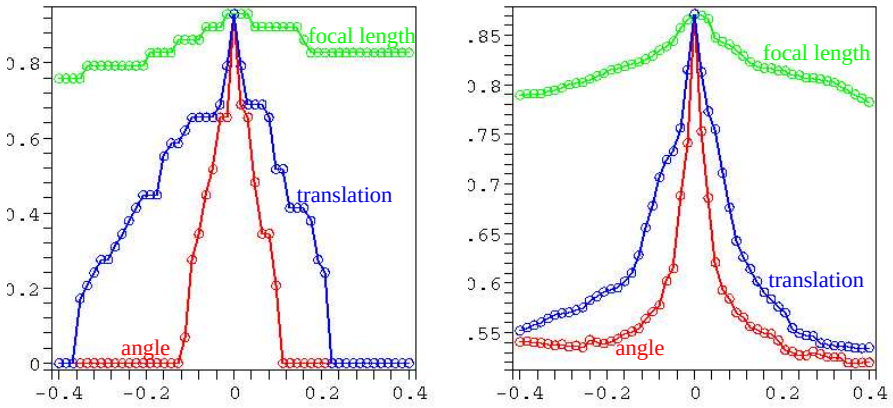

Figure 7: The silhouette calibration ratio of 2 (left) and 7 (right) viewing cones of a synthetic body model as a function of the same camera parameter values than in figure 6.

5.1 Silhouette calibration ratio

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 image contribution intervals along ray r, and let us call $\mathcal{N}(\omega_r)$ the number of image contribution 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 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:

$$C_r = \frac{1}{m(n-1)^2} \sum_r \sum_i \max_{\omega_r^i}(\mathcal{N}(\omega^i)), \tag{6}$$

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 as illustrated in figure 7.

5.2 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

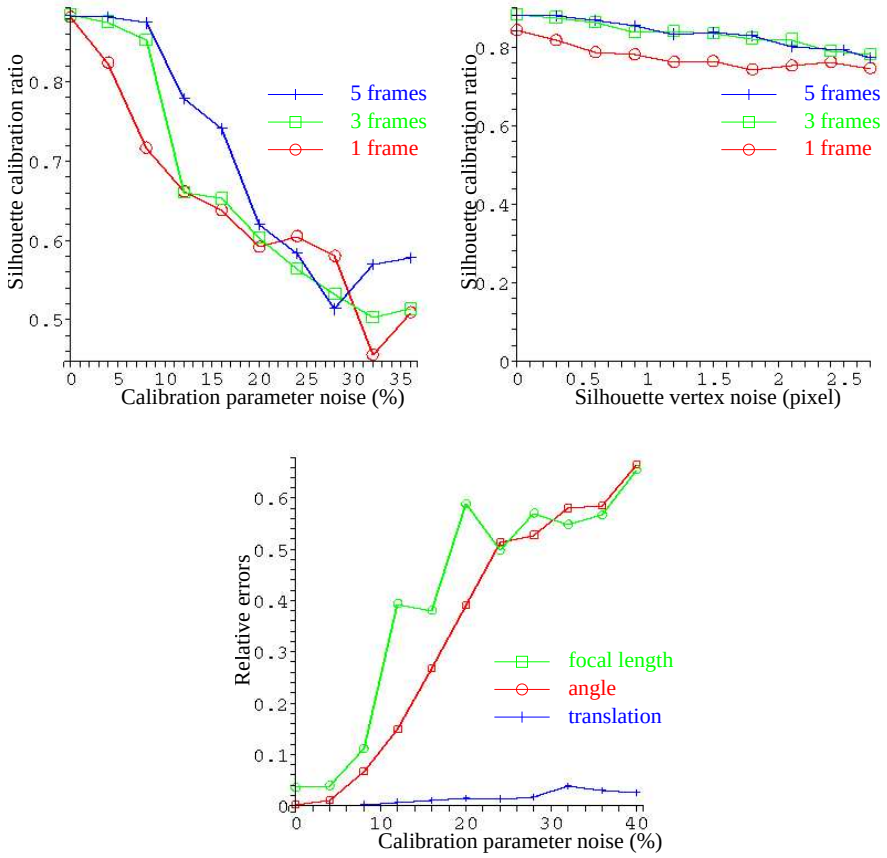

Figure 8: Top-Right, robustness to the initial solution; Top-left, robustness to the silhouette information; Bottom, the relative errors in the estimated camera parameters for the 5 frame case: errors relative to the true value for the focal length, errors relative to the distance camera-scene for the translation parameter and errors relative to π for the angle parameter.

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 8 shows the silhouette calibration ratios obtained by adding uniform noise to the initial solution (left) or Gaussian noise to silhouette vertices (middle); and on the right, relative errors in the estimated camera parameters after optimization using 5 frames per cameras. Each point, in all graphs, represents the mean value obtained over 20 trials.

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 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.

5.3 Real data

Our approach was also tested in real conditions with 6 firewire cameras viewing a moving person. The configuration is shown in figure 9. For this configuration, ground truth for camera parameters exists. It was obtained using a known calibration stick, and by optimizing an initial solution through a bundle adjustment. In the following experiments, we use the same initial solution for optimization. As for the synthetic case, all camera extrinsic parameters and one focal length per camera are optimized. Figure 9 shows the results of our calibration at different iterations. Figure 10 shows, on top, the input images and a visual hull model obtained using ground truth values for calibration; 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 the calibration stick dimension was imposed for the ground truth solution, the 2 most-right models are very close to the ground truth one. Figure 11 shows, silhouette calibration ratios obtained with ground truth values and with our method. Each point in the graph corresponds to the mean value obtained over 20 trials, each trial being achieved by randomly choosing frames in the sequence (200 frames). It shows that over 3 frames per camera, our approach behaves similarly to, or even better than, a classical approach with respect to the given silhouettes.

6 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. As most of the parameter optimization approaches, ours requires an initial solution for camera parameters. This operation can be difficult, and we are currently investigating this issue.

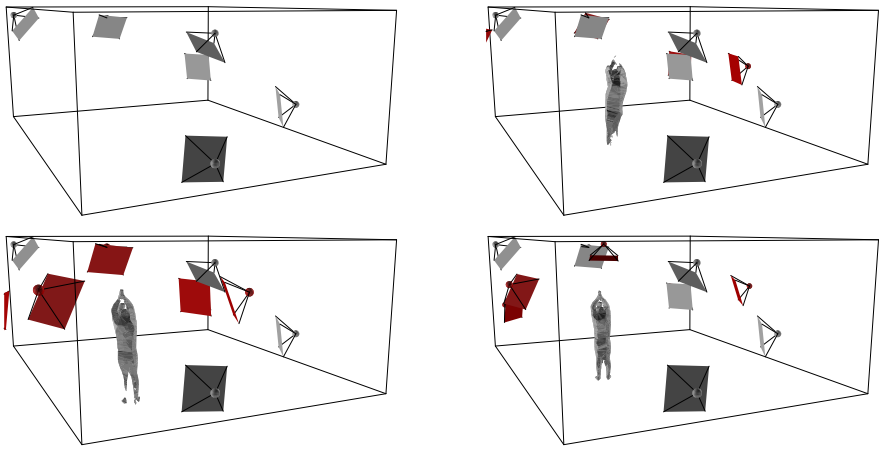

Figure 9: Real data: the original and optimized (in red) 6 camera configurations, as well as 1 reconstructed silhouette model. Top-Left, before optimization; Top-right and bottom, during and after optimization.

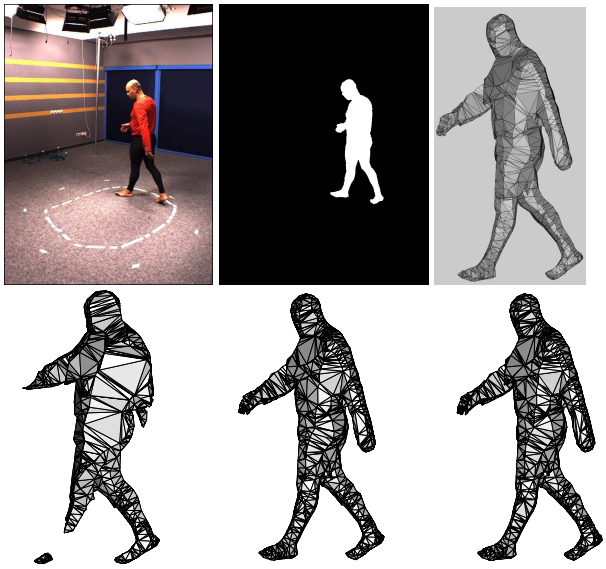

Figure 10: On top, one of the original image, the corresponding silhouette and the visual hull model obtained with ground truth calibration. Below, 3 models which correspond to calibrations obtained with our method and using respectively 1, 3 and 5 frames per camera.

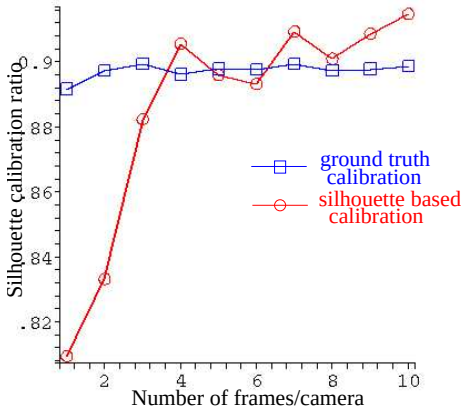

Figure 11: Mean value of the silhouette calibration ratio over 20 trials and as a function of the number of frames considered.

References

[Åström96] K. Åström, R. Cipolla, and P.J. Giblin. Generalised Epipolar Constraints. In *Proceedings of Fourth European Conference on Computer Vision, Cambridge, (England)*, pages 97–108, April 1996. Lecture Notes in Computer Science, volume 1065.

[Baumgart75] B.G. Baumgart. A polyhedron representation for computer vision. In *AFIPS National Computer Conference*, pages 589–596, 1975.

[Bottino03] A. Bottino and A. Laurentini. Introducing a New Problem: Shape-from-Silhouette when the Relative Positions of the Viewpoints is Unknown. *IEEE Transactions on PAMI*, 25(11): 1484–1493, 2003.

[Boyer97] E. Boyer and M.-O. Berger. 3D Surface Reconstruction Using Occluding Contours. *International Journal of Computer Vision*, 22(3): 219–233, 1997.

[Cipolla95] R. Cipolla, K.E. Åström, and P.J. Giblin. Motion from the Frontier of Curved Surfaces. In *Proceedings of 5th International Conference on Computer Vision, Boston (USA)*, pages 269–275, June 1995.

[Fitzgibbon98] A.W. Fitzgibbon, G. Cross, and A. Zisserman. 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*, pages 154–170, June 1998.

[Jiang04] G. Jiang, L. Quan, and H. Tsui. Circular Motion Geometry Using Minimal Data. *IEEE Transactions on PAMI*, 26(6): 721–731, 2004.

[Joshi95] T. Joshi, N. Ahuja, and J. Ponce. Structure and Motion Estimation from Dynamic Silhouettes under Perspective Projection. In *Proceedings of 5th International Conference on Computer Vision, Boston (USA)*, pages 290–295, June 1995.

[Laurentini94] A. Laurentini. The Visual Hull Concept for Silhouette-Based Image Understanding. *IEEE Transactions on PAMI*, 16(2): 150–162, February 1994.

[Mendoça01] P.R.S. Mendoça, K.-Y.K. Wong, and R. Cipolla. Epipolar Geometry from Profiles under Circular Motion. *IEEE Transactions on PAMI*, 23(6): 604–616, 2001.

[Rieger86] J.H. Rieger. Three-Dimensional Motion from Fixed Points of a Deforming Profile Curve. *Optics Letters*, 11: 123–125, 1986.

[Sinha04] S.N. Sinha, M. Pollefeys, and L. McMillan. Camera Network Calibration from Dynamic Silhouettes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington, (USA)*, 2004.

[Stolfi91] J. Stolfi. *Oriented Projective Geometry: A Framework for Geometric Computations* . Academic Press, 1991.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399