Consistency of silhouettes and their duals

Matthew Trager, Martial Hebert, Jean Ponce

Consistency of silhouettes and their duals

Matthew Trager* Inria
Martial Hebert Carnegie Mellon University
Jean Ponce* École Normale Supérieure / PSL Research University

Abstract

Silhouettes provide rich information on three-dimensional shape, since the intersection of the associated visual cones generates the "visual hull", which encloses and approximates the original shape. However, not all silhouettes can actually be projections of the same object in space: this simple observation has implications in object recognition and multi-view segmentation, and has been (often implicitly) used as a basis for camera calibration. In this paper, we investigate the conditions for multiple silhouettes, or more generally arbitrary closed image sets, to be geometrically "consistent". We present this notion as a natural generalization of traditional multi-view geometry, which deals with consistency for points. After discussing some general results, we present a "dual" formulation for consistency, that gives conditions for a family of planar sets to be sections of the same object. Finally, we introduce a more general notion of silhouette "compatibility" under partial knowledge of the camera projections, and point out some possible directions for future research.

1. Introduction

When can a set of 2D silhouettes be projections of the same object in space? This seemingly simple question is related to a variety of practical problems in computer vision, such as multi-view segmentation [9], object recognition [34], and multi-view stereo [11]. Geometric consistency is sometimes taken for granted, when appearance-based features give reasonable evidence that the silhouettes are associated to the same object. However, it is clear that incorporating geometric constraints can be important, either in the process of inference, or for correcting the effects of noisy data.

In this paper, we analyze the notion of "consistency" for silhouettes and for more general closed image sets. We consider opaque objects projected in different images and assume, in our initial setting, the knowledge of all camera parameters. The theory can be seen as a natural extension of classical multi-view geometry, which provides conditions for *points* (and sometimes *lines*) to be consistent (*i.e.*, to correspond) in terms of given camera projections [10, 14]. This also expands an analogy initiated by the generalized epipolar constraint introduced in [1]. Throughout our discussion, we point out several results that hold for arbitrary closed image sets (or sometimes convex image sets), that have identical counterparts in the theory of point correspondences.

As consistency is clearly not a metric property of silhouettes, it is natural to adopt the framework of *projective geometry* [6], since this eliminates various degenerate situations and allows, for example, to unify the cases of orthographic and perspective projections. This is also typical in multi-view geometry.[1] Another advantage of the projective language is that it provides a homogeneous formulation of *duality*. In particular, we exploit the fact that perspective projections are related to *planar sections* in the dual space [24], to define a very natural "dual" notion of consistency, expressing conditions for a family of planar sets to be sections of the same object.

Finally, we also consider the case of having only *partial* knowledge of the camera parameters, and discuss a more general concept of "compatibility" for silhouettes. This extends a setting first considered in [2].

The overall goal of the paper is not to give the "final" answer to the complex problem of silhouette consistency, but rather to make a first formal foray in that field, with precise definitions that have been missing so far in a general setting, spelling out rigorously what is known in this area and adding a set of new results. For example, we do not deal with algorithmic issues here, and we assume throughout the presentation an ideal setting with no noise. We believe this to be a necessary first step, much in the same way as multi-view geometry initially characterizes *exact* point correspondences, and then makes use of the theory to infer camera parameters from real world data.

*Willow project team. DI/ENS, ENS/CNRS/Inria UMR 8548.

[1]The most significant difference between the euclidean and projective frameworks in our setting is that in the latter case visual cones are *two-sided*. However, we will make one general assumption that will cause this distinction to be irrelevant. An alternative (but perhaps less natural) approach would have been the use of *oriented projective geometry* [27].

Previous work.

The most widespread application of silhouette consistency has been for designing alternatives to point-based methods for camera calibration, required for dealing with smooth and textureless surfaces. Indeed, ever since the seminal work of Rieger [25], the problem of estimating camera motion or calibration parameters using only silhouettes has received considerable attention: see, *e.g.*, [3, 12, 15, 18, 22, 32]. Albeit with some variations, all these methods exploit (more or less directly) the geometric constraints provided by the *epipolar tangencies* [1].

In addition to camera calibration, silhouette consistency has been enforced explicitly for other tasks; for example multi-view segmentation [4], or 3D-reconstruction [7] and recognition [21]. Another interesting "artistic" application is discussed in [23].

There exists limited theoretical work on silhouette consistency, and it is always restricted to special situations. In particular, the problem of determining whether a family of silhouettes can correspond to a real object is considered in [2, 17, 33], but results are only given for the case of orthographic projections and somewhat restricted camera motion. Some theoretical facts, *e.g.*, the fact that epipolar tangency conditions do not imply global consistency, can be found in [3, 5].

Finally, the duality between projections and planar sections is well known in both the euclidean (orthographic) setting [30, 26], as well as in the projective (perspective) case [24, 28]. However, we are not aware of work on consistency for planar sections.

Main contributions.

- We formally introduce a notion of *geometric consistency* for arbitrary closed sets in the case of general projective (perspective) cameras. We present some new results (Propositions 1, 3, 5), and collect others which are scattered in more applied work (Propositions 2, 4).
- We investigate in detail the relationship between pairwise consistency, epipolar tangencies, and our more general notion of consistency (Sections 2.2 and 2.3). We also discuss the (rather counterintuitive) topology of the visual hull associated to two views.
- We restate the notion of consistency in terms of duality, expressing the condition for planar sets to be sections of the same object. For convex silhouettes, we show that the dual of the visual hull coincides with the convex hull of the dual image of the silhouettes (Proposition 6).
- We define the notion of *compatibility* for silhouettes, which characterizes silhouettes that *may* be geometrically consistent for appropriate camera parameters. This setting generalizes a viewpoint first introduced in [2].

Proofs. Proof sketches are given in the main body of the presentation; full proofs and technical details can be found in the supplemental material.

Notation. Our analysis will be coordinate free, so we consider projective cameras as linear maps $\mathcal{M} : \mathbb{P}^3 \setminus \{\boldsymbol{c}\} \to \mathbb{P}^2$, where $\boldsymbol{c}$ is the camera pinhole or center. The action of a camera will be indicated with $\mathcal{M}(\boldsymbol{p}) = \boldsymbol{u}$ (no need to use proportionality as $\boldsymbol{u}$ is seen a *geometric point* in $\mathbb{P}^2$, rather than a vector of coordinates). For any camera $\mathcal{M}$ and set of points $T \subseteq \mathbb{P}^2$ in an image, we define the associated *visual cone* as $\mathcal{M}^{-1}(T)$, where $\mathcal{M}^{-1}$ denotes the pre-image set.

2. Consistency of image sets

In this section, we introduce a notion of geometric consistency for arbitrary closed sets in different images. Our definition is very natural, and similar concepts have previously been used to introduce "incoherence" measures for silhouettes [3, 15]. Compared to these works, we focus on analyzing some theoretical properties of consistency, rather than finding strategies for putting it into practice. After a general discussion, we consider the important case of two silhouettes (Section 2.2), pointing out how the consistency condition is basically equivalent to the popular "epipolar tangency" constraint, but only applied to *extremal* tangents. We then turn to the case of an arbitrary number of convex silhouettes (Section 2.3), where we note that all geometric impediments to consistency can only involve at most three silhouettes. We also introduce a special class of "tangential" triple points that help clarify the distinction between the epipolar tangency constraint and more general consistency.

2.1. Basic definitions

Let $\mathcal{M}_1, \ldots, \mathcal{M}_n$ be n perspective cameras with distinct centers $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$, and let $T_1, \ldots, T_n$ be a family of closed sets, one in each image. For example, the sets T_i could be a finite collection of points, curves or closed regions. For each $i = 1, \ldots, n$, we let $C_i = \mathcal{M}_i^{-1}(T_i)$ be the visual cone associated to T_i.

Definition 1. *The sets* $T_1, \ldots, T_n$ *are said to be* consistent *if there exists a non-empty set* $R \subseteq \mathbb{P}^3 \setminus \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}$ *such that* $\mathcal{M}_i(R) = T_i$ *for all* $i = 1, \ldots, n$.

When $T_1, \ldots, T_n$ *are consistent, the* visual hull *associated with* $T_1, \ldots, T_n$ *is given by* $H = \bigcap_i C_i$*, and it is the largest set that projects onto* $T_1, \ldots, T_n$.[2]

If all $T_1, \ldots, T_n$ are singletons, then consistency reduces to the classical notion of point correspondence [10, 14]; in this case, the visual hull is simply the triangulated 3D-point. Extending this analogy, consistent image sets can be seen as

[2]This notion of consistency is a property of the sets $T_1, \ldots, T_n$ *relative* to the cameras $\mathcal{M}_1, \ldots, \mathcal{M}_n$. However, in reality the condition is based on geometric properties of the visual cones C_i, rather than of the actual image sets. In fact, a more "geometric" definition of consistency only in terms of cones could have been given (analogous to requiring the convergence of lines). For our purposes, this approach would have probably been less natural.

n*-tuples of candidate projections* of the same object ("candidate" because, just as for point correspondences, there might not be an actual object occupying R). Moreover, it is clear that there is a one-to-one correspondence between n-tuples of consistent sets, and the (exact) visual hulls associated with a fixed set of n cameras.

In principle, the concept of visual hull is not well defined if the original silhouettes (or image sets) are not geometrically consistent. This is rarely taken into consideration, and it is customary to define the visual hull simply as the intersection of the cones $\bigcap_i C_i$ for arbitrary (non necessarily consistent) silhouettes: this operation can be justified by noting that if $\bigcap_i C_i$ is not empty, then it is the visual hull associated with the subsets $\tilde{T}_i = \mathcal{M}_i(\bigcap_j C_j) \subseteq T_i$, which will always be consistent. In fact, consistency is clearly equivalent to the fact that $\tilde{T}_i = T_i = \mathcal{M}_i(\bigcap_j C_j)$ for all $i = 1, \dots, n$, or to $T_i \subseteq \mathcal{M}_i(\bigcap_j C_j)$, since the opposite inclusion is always true. We collect a few other simple but useful properties:

Proposition 1. *Let $T_1, \dots, T_n$ be arbitrary closed image sets.*

1. *$T_1, \dots, T_n$ are consistent if and only if for each $i = 1, \dots, n$, and for all $\boldsymbol{u}_i \in T_i$, the visual ray $\mathcal{M}^{-1}(\boldsymbol{u}_i)$ intersects $\bigcap_{j \neq i} C_j$.*

2. *$T_1, \dots, T_n$ are consistent if and only if*

$$T_i \subseteq \mathcal{M}_i \left(\bigcap_{j \neq i} C_j \right), \quad \forall i \in \{1, \dots, n\}. \qquad (1)$$

3. *If $T_1, \dots, T_n$ are consistent, then any subfamily $T_{i_1}, \dots, T_{i_s}$ is consistent (for the associated cameras $\mathcal{M}_{i_1}, \dots, \mathcal{M}_{i_s}$).*

Proof. The first property follows from the fact that $T_i \subseteq \mathcal{M}_i(\bigcap_j C_j)$ can be expressed as $\mathcal{M}_i^{-1}(\boldsymbol{u}_i) \cap \bigcap_j C_j \neq \emptyset$ for all $\boldsymbol{u}_i \in T_i$, which in turn is equivalent to $\mathcal{M}_i^{-1}(\boldsymbol{u}_i) \cap \bigcap_{j \neq i} C_j \neq \emptyset$, since $\mathcal{M}_i^{-1}(\boldsymbol{u}_i) \subseteq C_i$. The second and third properties are consequences of the first one. □

This might be a good moment to point out that the notion of geometric consistency discussed in this paper is somewhat independent from a more intuitive (but less formal) concept of "similarity" of appearance. For example, Figure 1 shows that consistent silhouettes may actually look completely different; on the other hand, almost identical silhouettes may be geometrically inconsistent. Thus, the concept might be well suited for being used alongside more traditional feature-based methods for recognition.

Figure 1: In shadow art very different looking consistent silhouettes are exploited for artistic effects. Figure from [23].

Finally, for the rest of the paper we will make the following assumption for all n-tuples of image sets $T_1, \dots, T_n$ and cameras $\mathcal{M}_1, \dots, \mathcal{M}_n$:

(A) For each camera center $\mathbf{c}_i$ and visual cone C_j, with $i \neq j$, $\mathbf{c}_i$ does not belong to C_j.

This condition is useful for excluding various degenerate situations; for example, it guarantees that the visual hulls associated with all subfamilies of $T_1, \dots, T_n$ are closed sets.[3]

2.2. Pairwise consistency

Let us assume that we are given only two image sets T_1, T_2 (and, as usual, two cameras $\mathcal{M}_1, \mathcal{M}_2$). According to Proposition 1, we know that T_1, T_2 are consistent if and only if

$$T_1 \subseteq \mathcal{M}_1 (C_2), \quad \text{and} \quad T_2 \subseteq \mathcal{M}_2 (C_1) \qquad (2)$$

In other words, we require for each set to be contained in the projection of the visual cone associated to the other one (see Figure 2). Pairwise consistency is closely related to the popular *epipolar tangency constraint* [1, 32]. Indeed, we can restate the condition (2) in terms of epipolar geometry as follows.

Proposition 2. *Two arbitrary closed sets T_1, T_2 are consistent if and only if the set of epipolar lines in the first image intersecting T_1 is in epipolar correspondence with the set of epipolar lines in the second image intersecting T_2.*

Proof. The statement can be seen as a consequence of the first property in Proposition 1. In fact, the epipolar correspondence condition guarantees that for all $i = 1, 2$, and for every point $\boldsymbol{u}_i \in T_i$ there exists at least one corresponding point $\boldsymbol{u}_j \in T_j$ ($j \neq i$), so that triangulating all pairs of associated points (*i.e.*, intersecting the cones C_1, C_2) we obtain a set $R \subseteq \mathbb{P}^3 \setminus \{\boldsymbol{c}_1, \boldsymbol{c}_2\}$ that projects exactly onto T_1 and T_2. Note that assumption (A) guarantees that, in each image, the epipole lies outside of the given set. □

This result can also be stated in more geometric terms: in order for T_1, T_2 to be consistent, we consider the pencil of planes through the two centers $\boldsymbol{c}_1, \boldsymbol{c}_2$, and require the set

[3] This condition is not actually necessary for all of our results, and weaker assumptions may often be considered. However, for the sake of simplicity, we give a single condition that is valid throughout the paper.

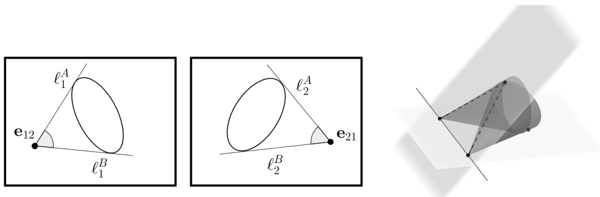

Figure 2: Parwise consistency. Left: the lines ℓ_1^A, ℓ_2^A and the lines ℓ_1^B, ℓ_2^B must be in epipolar correspondence. Right: the set of planes through the centers $\boldsymbol{c}_1, \boldsymbol{c}_2$ that intersect C_1 must be the same as the set of planes intersecting C_2.

of planes which intersecting C_1 to be the same as the set intersecting C_2. See Figure 2.

If we assume that T_1, T_2 are connected closed regions bounded by smooth curves, then pairwise consistency basically reduces to the fact that *extremal epipolar tangents* (*i.e.*, "outermost" epipolar lines that are tangent to the contours) are in epipolar correspondence [3]. It is worth emphasizing that pairwise consistency *does not* require non-extremal epipolar tangents to be matched. Indeed, assuming that their extremal epipolar tangents correspond, *any* two silhouettes determine an actual visual hull, as can be seen from the proof of Proposition 2. In fact, perhaps unintuitively, the visual hull of two silhouettes with non-extremal epipolar tangencies, (matched or unmatched) will *always* exhibit the kind of self-occlusions shown in Figure 3 (although for close viewpoints the visual hull will be extremely "deep"). This can be argued investigating the topology of the visual hull, by walking along all possible paths α in H: these are equivalent to pairs of paths β_1 in T_1, β_2 in T_2 that are such that $\beta_1(t) \in T_1$ and $\beta_2(t) \in T_2$ remain in epipolar correspondence for all t; in particular, if the images are assumed to be rectified, then such pairs of paths are those that can be traveled along simultaneously in the two silhouettes while always remaining at the same "height" (interestingly, this setting is closely related to what is known as the "mountain climbing problem" [31]). See Figure 4 (and the supplemental material for more details). Using similar arguments we also observe that, again rather surprisingly, the visual hull associated to two connected and consistent sets may be disconnected: an example is shown in Figure 5.

Returning to the case of an arbitrary number of sets, we will say that $T_1, \ldots, T_n$ are *pairwise consistent* if each pair T_i, T_j, $i \neq j$ is consistent. From Proposition 1, we know that this holds whenever $T_1, \ldots, T_n$ are consistent. The converse is *not* true, as pointed out in [3, 5] (see also Figure 7). However, it would be useful to clarify the practical distinction between these two notions: much of our discussion in the following will be aimed at a better understanding of this issue. For example, in the case of three sets, pairwise consistency often implies that there is at least an "approximate" consistency: this property directly generalizes the fact that three *non-coplanar* visual rays that converge pairwise will always converge [14].

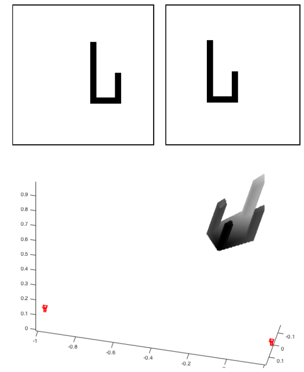

Figure 3: The visual hull of two "hook-shaped" silhouettes *always* presents self-occlusions. See Figure 4 for an explanation.

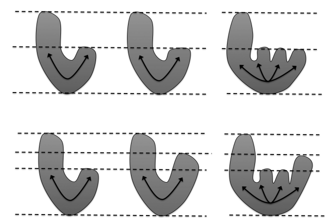

Figure 4: Topology of two silhouettes (left and center) and their visual hull (right). In both the top and bottom cases (matched and unmatched internal epipolar tangencies), the visual hull has four "bumps", that correspond to four pairs of paths in the projections ("long" and "short" side for each silhouette) that occlude each other. The difference between the two situations is that in the first case there *exist* objects (other than the visual hull) that project onto the two silhouettes with only *two* bumps, while in the second case any object that projects consistently must have at least three bumps (and self-occlusions *must* occur). See the supplemental material for a more detailed discussion.

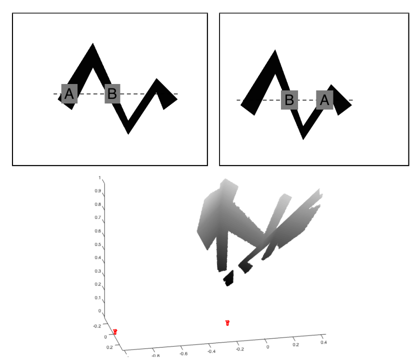

Figure 5: Two special "M" figures give rise to a (very complex) disconnected visual hull. To prove this, consider a pair of corresponding points located in the two sillhouettes at "A", and a pair of corresponding points located at "B". These may be viewed as two points on the visual hull, and no path on the visual hull can connect them, since there is no pair of admissible paths in the projected silhouettes (*i.e.*, it is not possible to go from "A" to "B" in both silhouettes while always remaining at the same height).

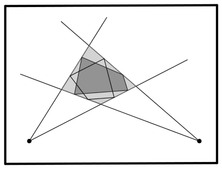

Figure 6: Proof of Proposition 3. The set T_1 (dark gray) and $\mathcal{M}_1(C_2 \cap C_3)$ are both "inscribed" in the quadrilateral $\mathcal{M}_1(C_2) \cap \mathcal{M}_1(C_3)$ (light gray), and must thus intersect.

Proposition 3. *Let T_1, T_2, T_3 be connected closed sets that are pairwise consistent. If the centers of the cameras $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ are not collinear, and if each visual cone C_i does not intersect the plane spanned by $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$, then $\bigcap_i C_i$ is not empty.*

Proof. It is sufficient to prove that, say, $T_1 \cap \mathcal{M}_1(C_2 \cap C_3)$ is not empty. Let $R = \mathcal{M}_1(C_2) \cap \mathcal{M}_1(C_3)$. The assumptions on the centers guarantee that R is a quadrilateral (it is the intersection of two connected projected cones; see Figure 6). From pairwise consistency (2), we know that $T_1 \subseteq R$ and, moreover, T_1 will contain points belonging to each of the four edges of R. The same actually holds for $\mathcal{M}_1(C_2 \cap C_3)$: indeed $\mathcal{M}_1(C_2 \cap C_3) \subseteq R$ is true from basic set theory, and it must contain points on every edge of R since C_2, C_3 are themselves pairwise consistent (an extremal epipolar line in $\mathcal{M}_1(C_2)$ is the projection of a line in C_2 which must intersect $C_2 \cap C_3$). The claim now follows from simple continuity arguments: for example, consider paths in T_1 and $\mathcal{M}_1(C_2 \cap C_3)$ connecting different pairs of opposite edges. □

2.3. Consistency for convex sets

Let us now assume that $T_1, \ldots, T_n$ are closed *convex* sets. The careful reader might object that the usual notion of convexity cannot be used in a projective setting, since two points do not determine a segment. In practice, this is not an issue: we can say that a closed set $T \subseteq \mathbb{P}^2$ is convex if its intersection with every line consists of at most one connected component, and there is at least a line in $\mathbb{P}^2$ that doesn't intersect T (in other words, T is convex if there exists an affine chart in which T is finite and convex). In particular, a set lying in a physical image that is convex in the traditional sense is clearly also convex according to our definition in the projective closure of the image. We refer to the supplemental material for a more detailed discussion on convexity in the projective setting. If $T_1, \ldots, T_n$ are consistent closed convex sets, then the associated visual hull H is also closed and convex.

Although considering the convex case may appear restrictive, we note that: 1) The pairwise consistency constraint discussed in the previous section is actually a condition on convex hulls, since Proposition 2 clearly implies that an n-tuple of connected sets is pairwise consistent if and only if the associated convex hulls are pairwise consistent. 2) If an n-tuple of connected sets is consistent, then their convex hulls will also be consistent, as a consequence of Proposition 1. 3) Even in the restricted case of convex sets, pairwise consistency *does not* imply global consistency, as can be seen in Figure 7 (contradicting a claim in [3]). We thus believe that it is useful to investigate this case first, and clarify the distinction between pairwise consistency and global consistency in this simplified setting.

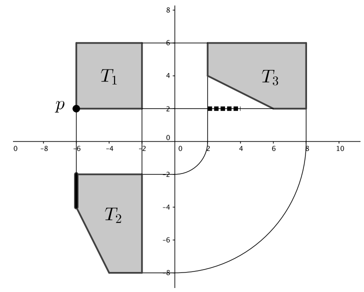

Figure 7: The convex sets T_1, T_2, T_3, with respect to appropriate orthographic projections, are pairwise consistent but not globally consistent. This figure can be understood as a multi-view orthographic drawing: in particular, the epipolar lines that are shown guarantee pairwise consistency, but, for example, the point $\boldsymbol{p}$ in T_1 corresponds to a segment in T_2 that reprojects outside of T_3: this implies that the visual ray associated to $\boldsymbol{p}$ does not intersect $C_2 \cap C_3$.

For example, although not equivalent to pairwise consistency, general consistency for convex sets is actually guaranteed by "triplet-wise" consistency (see also [33], where this result is stated for orthographic projections):

Proposition 4. *Let $T_1, \ldots, T_n$ be closed convex sets. If T_i, T_j, T_k are consistent for every $\{i, j, k\} \subseteq \{1, \ldots, n\}$ then $T_1, \ldots, T_n$ are consistent.*

Proof. From Proposition 1, it is sufficient to prove that for every $i = 1, \ldots, n$, and for all $\boldsymbol{u}_i \in T_i$, we have $\mathcal{M}^{-1}(\boldsymbol{u}_i) \cap \left(\bigcap_{j \neq i} C_j\right) \neq \emptyset$. Because of convexity, the sets $U_j^i = \mathcal{M}_i^{-1}(\boldsymbol{u}_i) \cap C_j$ are intervals; moreover, they intersect pairwise because of the assumption of "triplet-wise" consistency: this implies that they all intersect.[4] See Figure 8. □

This statement closely resembles "Helly-type" theorems in computational geometry [29], and is a generalization of the fact that point correspondence is always implied by triplet-wise point correspondence. An inspection of the proof also shows that it is not actually necessary for the sets T_i to be convex, but only for their intersections with epipolar lines (rather than *all* lines) to be intervals.

[4]To be precise, it should be noted that, although in projective space, none of the cones C_j contain $\boldsymbol{c}_i$ (because of assumption (A)), so we may treat the sets U_j^i as intervals on a real line.

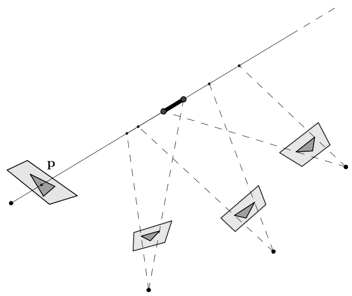

Figure 8: Proof of Proposition 4: the visual ray corresponding to $\boldsymbol{p}$ meets the other visual cones in intervals that intersect pairwise, and thus must all intersect.

If we assume that $T_1, \ldots, T_n$ are closed convex regions bounded by curves $\gamma_1, \ldots, \gamma_n$, then condition (1) for consistency can be simplified as

$$\gamma_i \subseteq \mathcal{M}_i\left(\bigcap_{j\neq i} C_j\right), \ \forall i \in \{1, \ldots, n\}, \tag{3}$$

since the sets $\mathcal{M}_i\left(\bigcap_{j\neq i} C_j\right)$ are convex. This condition has actually been used in [15] as a practical relaxation of (1) to measure consistency for general (non-convex) silhouettes: indeed, (3) is actually equivalent to general consistency whenever $\mathcal{M}_i\left(\bigcap_{j\neq i} C_j\right)$ is simply connected. While this is often a reasonable assumption, it is not clear how to easily verify it when the sets T_i are not convex.

Finally, we can use (3) to provide a condition that can be added to pairwise consistency in order to guarantee general consistency. We first need to observe that a visual ray can intersect two other convex silhouettes in several qualitative ways, which are shown in Figure 9. In particular, we will call a point $\boldsymbol{u} \in \gamma_1$, such that the visual ray $\mathcal{M}^{-1}(\boldsymbol{u})$ is as in case (e) a "tangential" triple point (our terminology). We now can state the following result:

Proposition 5. *Let $T_1, \ldots, T_n$ be pairwise consistent convex regions with smooth boundaries $\gamma_1, \ldots, \gamma_n$. Assume also that $\gamma_i \cap \mathcal{M}_i(\bigcap_{j\neq i} C_j) \neq \emptyset$ for all $i = 1, \ldots, n$. If $\gamma_1, \ldots, \gamma_n$ do not contain tangential triple points, then $T_1, \ldots, T_n$ are consistent.*

Proof sketch. By virtue of Proposition 4 we may assume $n = 3$. Consider $\boldsymbol{u} \in \gamma_1 \cap \mathcal{M}_1(C_2 \cap C_3)$: as $\boldsymbol{u}$ varies continuously along the contour γ_1, the visual ray $\mathcal{M}^{-1}(\boldsymbol{u})$ will always intersect C_2 and C_3 in two segments (possibly reduced to points, as shown in Figure 9) because of pairwise consistency; moreover these segments will "slide" continuously along the ray. It is clear that, in order for there to exist a point $\boldsymbol{u}' \in \gamma_1$ such that $\mathcal{M}^{-1}(\boldsymbol{u}') \cap C_2 \cap C_3 = \emptyset$, a tangential triple point must occur. A more formal argument is given in the supplemental material. □

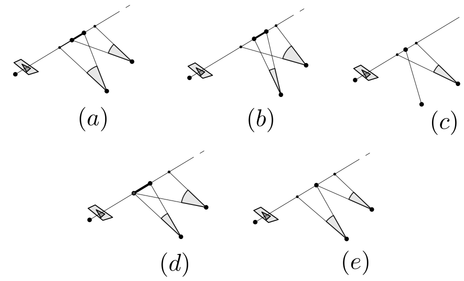

Figure 9: Intersections of a visual ray with two other convex silhouettes. Cases (a) and (b) are the "generic" intersections; case (c) is a ray tangent to another silhouette (corresponding to a point of epipolar tangency); case (d) is a "typical" triple point; case (e) is a "tangential" triple point.

3. A dual view of consistency

In this section we revisit the notion of consistency from the viewpoint of *duality*. In particular, we show how the consistency of projections relates to a different notion of consistency for planar sections. In this section, we focus mainly on the case of convex sets, for which this relationship is especially transparent. For completeness, we first recall in Section 3.1 some basic definitions of duality in the projective setting.

3.1. Projective duality

The basis of many similar notions of duality is the fact that points and hyperplanes in some n dimensional space can play symmetric roles. In $\mathbb{R}^n$, for example, a hyperplane through the origin can be described by its orthogonal vector. In projective space $\mathbb{P}^n$, any hyperplane $H \subseteq \mathbb{P}^n$ represents a point in the dual space $(\mathbb{P}^n)^*$. More generally, duality associates k-dimensional linear subspaces in $\mathbb{P}^n$ with $(n-k)$-dimensional linear subspaces in $(\mathbb{P}^n)^*$, by interchanging the role of "join" and "meet" [6].

Now if $S \subseteq \mathbb{P}^n$ is a *smooth hypersurface*, the set of tangent hyperplanes at points of S forms a *dual* hypersurface $S^* \subseteq (\mathbb{P}^n)^*$. Although seemingly intuitive, it is still remarkable that $(S^*)^* = S$ [28], so duality defines an involutive correspondence between smooth hypersurfaces. However, dual hypersurfaces will typically have self-intersections: for example, in the case of curves, crossings correspond to bitangents of the original curve. With some care, duality can also be extended to *piecewise smooth* surfaces and curves, assuming these are oriented. We refer to the supplementary material for details. For our purposes, we will only be interested in curves in $\mathbb{P}^2$ and surfaces in $\mathbb{P}^3$.

3.2. Duality and visual hulls

Projective duality is useful in studying vision because taking perspective projections of a surface S is equivalent

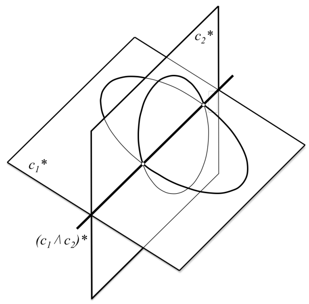

Figure 10: Dual-pairwise-consistency. The dual images of the silhouettes $\mathcal{M}_1^*(T_1^*)$ and $\mathcal{M}_1^*(T_2^*)$ must have the same intersection with the "dual baseline" $(\boldsymbol{c}_1 \wedge \boldsymbol{c}_2)^*$.

to taking *planar sections* of the corresponding dual surface S^*. Indeed, a perspective camera $\mathcal{M}$ with center $\boldsymbol{c}$ also defines a dual map $\mathcal{M}^*$ that associates lines in $\mathbb{P}^2$ to planes in $\mathbb{P}^3$ through $\boldsymbol{c}$: in particular, the image of $\mathcal{M}^*$ is a *plane* in the dual space $(\mathbb{P}^3)^*$, namely $\boldsymbol{c}^*$. If $\gamma \subseteq \mathbb{P}^2$ is the boundary of the projection of some surface $S \subseteq \mathbb{P}^3$, then $\mathcal{M}^*(\gamma^*) \subseteq (\mathbb{P}^3)^*$ is a planar curve that coincides with $S^* \cap \boldsymbol{c}^*$. We also refer to [24], where projective duality is used to investigate the qualitative relationship between image contours and projective shapes.

If $T_1, \ldots, T_n$ are silhouettes bounded by curves $\gamma_1, \ldots, \gamma_n$, then we can try to express the condition for consistency in terms of the dual curves $\gamma_1^*, \ldots, \gamma_n^*$. For simplicity, we assume that $T_1, \ldots, T_n$ are convex: in this case, $\gamma_1^*, \ldots, \gamma_n^*$ do not have self-intersections (because $\gamma_1, \ldots, \gamma_n$ do not have bitangents) and they bound convex regions $T_1^*, \ldots, T_n^*$ in $(\mathbb{P}^2)^*$. Fixing appropriate affine charts, the sets T_i^* are the *dual convex bodies* associated to T_i [26]. See the supplemental material for a more detailed discussion.

We now introduce the notion of "sectional consistency" for planar sets sets in $\mathbb{P}^3$:

Definition 2. *Let $S_1, \ldots, S_n$ be convex and planar sets in $\mathbb{P}^3$, so that $S_i \subset \pi_i$ for distinct planes $\pi_1, \ldots, \pi_n$. We say that such sets are* sectionally consistent *if there exists a convex set $K \subseteq \mathbb{P}^3$ such that $K \cap \pi_i = S_i$ for all $i = 1, \ldots, n$.*

Note that if $S_1, \ldots, S_n$ are sectionally consistent, the smallest set K from the previous definition is simply the convex hull of $S_1, \ldots, S_n$. In particular, this means that if $S_1, \ldots, S_n$ are bounded (relative to the planes π_i) by curves $\tau_1, \ldots, \tau_n$, then sectional consistency is equivalent to the fact that each curve τ_i lies on the boundary of their convex hull K.

The notion of sectional consistency can be seen as a dual version of consistency. Indeed, it is a simple consequence of duality that a set of convex silhouettes $T_1, \ldots, T_n$ is consistent if and only if the planar sets $\mathcal{M}_1^*(T_1^*), \ldots, \mathcal{M}_n^*(T_n^*)$ are sectionally consistent. Moreover, the following holds:

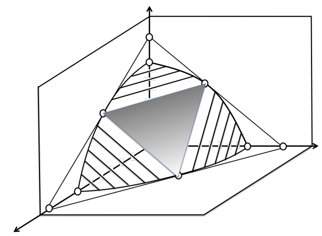

Figure 11: Structure of the dual visual hull. The dual visual hull is composed of planar curves, ruled surface patches and planar triangular patches: each of these are associated to specific components of the visual hull. See text for details.

Proposition 6. *If $T_1, \ldots, T_n$ are consistent convex silhouettes and H is their associated visual hull, then $H^* = K$, where K is the convex hull of $\mathcal{M}_1^*(T_1^*), \ldots, \mathcal{M}_n^*(T_n^*)$. Thus, K is the* "dual visual hull" *of $T_1, \ldots, T_n$.*

Indeed, H^* is the smallest possible set with the "correct" sections (since H is the largest possible set with the correct projections); thus it must coincide with K, because duality reverses containments for convex sets. See the supplemental material for a more detailed proof. Alternatively, it is a standard result in convex analysis that if K_1, K_2 are convex sets in $\mathbb{R}^n$, then $(K_1 \cap K_2)^* = Conv(K_1^* \cup K_2^*)$, assuming that $\mathbf{0} \in K_1$ and $\mathbf{0} \in K_2$ [26] (here $Conv$ denotes the convex hull): by choosing appropriate affine charts, this can be used to show Proposition 6.

Consistency is arguably more intuitive in its dual formulation, since planar regions in space are conceptually easier to grasp than families of intersecting cones. For example, the pairwise consistency constraint for two convex silhouettes T_1, T_2 reduces to the fact that $\mathcal{M}_1^*(T_1^*)$ and $\mathcal{M}_1^*(T_2^*)$ have the same intersection with the "dual baseline" $(\boldsymbol{c}_1 \wedge \boldsymbol{c}_2)^*$ (here $\wedge$ denotes the "join" of two points): see Figure 10.

We also briefly describe a "combinatorial" relationship between the visual hull H and its dual $K = H^*$. Let π be a plane tangent to H (*i.e.*, a supporting plane in an appropriate affine chart). If π is also tangent to a visual cone C_i, then it represents a point in the dual space that belongs to a planar curve $\mathcal{M}_i^*(\gamma_i^*)$; if π only intersects H along an *intersection curve* [20] belonging to $\partial C_i \cap \partial C_j$, then it represents a point on a ruled patch of H^* connecting $\mathcal{M}_i^*(\gamma_i^*)$ and $\mathcal{M}_j^*(\gamma_j^*)$; finally, if π only intersects H at a *triple point* belonging to $\partial C_i \cap \partial C_j \cap \partial C_k$, then it represents a point on a planar patch of H^* connecting $\mathcal{M}_i^*(\gamma_i^*), \mathcal{M}_j^*(\gamma_j^*), \mathcal{M}_k^*(\gamma_k^*)$ (ruled and planar patches are typical for convex hulls of curves in space [19]). See Figure 11.

The notion of sectional consistency given in Definition 2 is reminiscent of questions in geometric tomography [13], or stereology [16]. In tomography, for example, the duality between projections and sections is well studied, but typically in a euclidean setting that considers only orthographic projections. Nevertheless, it is quite possible that tools from these related fields could provide interesting new insight for problems in computer vision: for example, the corresponding algorithms may give alternative approaches for the construction of the (dual) visual hull.

4. Compatible silhouettes

Throughout the paper, we have always considered families of silhouettes (or image sets) $T_1, \dots, T_n$ together with known camera projections $\mathcal{M}_1, \dots, \mathcal{M}_n$. However, we can introduce a more general notion of "geometric compatibility", that can be applied in the case of incomplete knowledge of the camera parameters. More precisely, we define an arbitrary family of silhouettes (or image sets) $T_1, \dots, T_n$ to be *compatible with some partial knowledge* $\mathcal{P}$ of the camera parameters when there *exist* projections $\mathcal{M}_1, \dots, \mathcal{M}_n$ that agree with $\mathcal{P}$ and for which $T_1, \dots, T_n$ are consistent according to Definition 1.

Simply put, this notion characterizes the most general condition for a family of silhouettes (or sets) to be feasible projections of a single object. The study of similar issues was initiated in [2], where the authors assume the external parameters of the cameras to be unknown, and analyze in detail the geometric constraints for compatibility in a particular case (orthographic viewing directions parallel to the same plane).

The first theoretical problems raised by our definition of compatibility is to understand, in a given setting, 1) how "large" the space of compatible silhouettes is, and 2) the extent to which the complete camera parameters are determined by compatible silhouettes. For example, in [2] it is pointed out that if one considers internal as well as external parameters to be unknown, then *any* family of silhouettes will be compatible (they note that considering a convex object, and applying local protrusions with appropriate shapes, one is able to produce arbitrary silhouettes by placing cameras near the surface). This interesting observation, however, would seem to imply that SfM methods can never exploit the geometry of the silhouettes in order to recover camera parameters, since a particular family of silhouettes would provide no information on the viewing conditions. On the other hand, we note that the construction proposed in [2] violates our assumption (A) for *all* pairs of cameras and silhouettes (viewing cones must be extremely "wide", and thus contain all other centers which lie near the surface of the convex object): instead, we argue that by restricting ourselves to certain regions in the space of parameters so that (A) is satisfied by sufficiently many pairs of views (or, in practice, by considering appropriate initial estimates for the parameters), then a sufficient number of generic silhouettes *will* allow camera projections to be locally (over)determined, even with no prior knowledge about the parameters. See Figure 12 or the supplemental material.

In general, a better understanding of these issues can be important for several practical reasons, such as spelling out

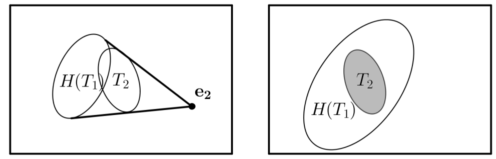

Figure 12: In the case that (only) two silhouettes T_1, T_2 are given, we can express the fundamental matrix as $F = [\boldsymbol{e}_2]_\times H$ [14], and note that if the homography H is such that $H(T_1)$ is not contained in or does not contain T_2, then $\boldsymbol{e}_2$ is basically uniquely determined by H (it must be at the intersection of two bitangents: see [8] that exploits this idea with parallax); on the other hand if, say, $H(T_1) \subseteq T_2$ then $\boldsymbol{e}_2$ is only constrained to belong to $H(T_1)$. We see that with appropriate assumptions camera geometry is constrained by silhouettes (in this case, 5 dof instead of 7).

conditions for when silhouettes may or may not be used to determine camera geometry (and possibly help design better algorithms), or similarly to give conditions for a family of silhouettes alone to provide a unique representation of (the visual hull of) an object. Duality might also prove to be a useful tool for investigating these kinds of questions: much in the same way as in Section 3, we realize that the notion of compatibility of silhouettes can be expressed in terms of the compatibility of planar regions, which need to be "assembled" consistently in order to be feasible sections of a single object.

5. Conclusions

We have analyzed in detail the notion of "geometric consistency" for arbitrary image sets, in a setting that can be seen as an extension of traditional multi-view geometry. In the case of convex silhouettes we have also discussed a "dual" interpretation of consistency, expressing conditions for planar sets to be sections of a single object. These concepts lead to a more general notion of silhouette "compatibility", that does not require (complete) knowledge of the camera parameters.

We plan to extend this work in various directions. On the practical side, our results need to be revisited for dealing with noisy data, and the theory may be used for comparing different measures of "inconsistency" such as the ones considered in [3, 15, 22]. On the theoretical side, in addition to the questions on "compatibility" discussed above, there remains to gain a better understanding of (primal and dual) consistency for non-convex silhouettes. In geometric tomography, for example, it is typical to study sections of convex bodies [13]; however, for applications in vision, this assumption may be restrictive. Solutions to all of these problems would be useful in many practical settings, and would help us clarify the fundamental relationship between two-dimensional projections and the natural concept of "shape".

References

[1] K. Åström, R. Cipolla, and P. J. Giblin. Generalised epipolar constraints. In *Computer Vision—ECCV'96*, pages 95–108. Springer, 1996. 1, 2, 3

[2] A. Bottino and A. Laurentini. Introducing a new problem: Shape-from-silhouette when the relative positions of the viewpoints is unknown. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(11):1484–1493, 2003. 1, 2, 8

[3] E. Boyer. On using silhouettes for camera calibration. In *Computer Vision–ACCV 2006*, pages 1–10. Springer, 2006. 2, 4, 5, 8

[4] N. D. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla. Automatic 3d object segmentation in multiple views using volumetric graph-cuts. *Image and Vision Computing*, 28(1):14–25, 2010. 2

[5] K. M. G. Cheung. *Visual hull construction, alignment and refinement for human kinematic modeling, motion tracking and rendering*. PhD thesis, Citeseer, 2003. 2, 4

[6] H. S. M. Coxeter. *Projective geometry*. Springer Science & Business Media, 2003. 1, 6

[7] D. Cremers and K. Kolev. Multiview stereo and silhouette consistency via convex functionals over convex domains. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(6):1161–1174, 2011. 2

[8] G. Cross, A. W. Fitzgibbon, and A. Zisserman. Parallax geometry of smooth surfaces in multiple views. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 323–329. IEEE, 1999. 8

[9] A. Djelouah, J.-S. Franco, E. Boyer, F. Le Clerc, and P. Perez. Multi-view object segmentation in space and time. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 2640–2647. IEEE, 2013. 1

[10] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. MIT press, 2004. 1, 2

[11] Y. Furukawa and J. Ponce. Carved visual hulls for image-based modeling. In *Computer Vision–ECCV 2006*, pages 564–577. Springer, 2006. 1

[12] Y. Furukawa, A. Sethi, J. Ponce, and D. Kriegman. Structure and motion from images of smooth textureless objects. In *Computer Vision-ECCV 2004*, pages 287–298. Springer, 2004. 2

[13] R. J. Gardner. *Geometric tomography*, volume 6. Cambridge University Press Cambridge, 1995. 7, 8

[14] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1, 2, 4, 8

[15] C. Hernández, F. Schmitt, and R. Cipolla. Silhouette coherence for camera calibration under circular motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(2):343–349, 2007. 2, 6, 8

[16] V. Howard and M. Reed. *Unbiased stereology: three-dimensional measurement in microscopy*. Garland Science, 2004. 7

[17] D. Jacobs, P. Belhumeur, and I. Jermyn. Judging whether multiple silhouettes can come from the same object. In *Visual Form 2001*, pages 532–541. Springer, 2001. 2

[18] T. Joshi, N. Ahuja, and J. Ponce. Structure and motion estimation from dynamic silhouettes under perspective projection. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 290–295. IEEE, 1995. 2

[19] J. J. Koenderink. *Solid shape*, volume 2. Cambridge Univ Press, 1990. 7

[20] S. Lazebnik, Y. Furukawa, and J. Ponce. Projective visual hulls. *International Journal of Computer Vision*, 74(2):137–165, 2007. 7

[21] S. Lazebnik, A. Sethi, C. Schmid, D. Kriegman, J. Ponce, and M. Hebert. On pencils of tangent planes and the recognition of smooth 3d shapes from silhouettes. In *Computer Vision—ECCV 2002*, pages 651–665. Springer, 2002. 2

[22] P. R. Mendonça, K.-Y. K. Wong, and R. Cippolla. Epipolar geometry from profiles under circular motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):604–616, 2001. 2, 8

[23] N. J. Mitra and M. Pauly. Shadow art. In *ACM Transactions on Graphics*, volume 28, pages 156–1, 2009. 2, 3

[24] J. Ponce and M. Hebert. On image contours of projective shapes. In *European Conference on Computer Vision*, 2014. 1, 2, 7

[25] J. Rieger. Three-dimensional motion from fixed points of a deforming profile curve. *Optics Letters*, 11(3):123–125, 1986. 2

[26] R. Schneider. *Convex bodies: the Brunn–Minkowski theory*. Number 151. Cambridge University Press, 2013. 2, 7

[27] J. Stolfi. *Oriented projective geometry: A framework for geometric computations*. Academic Press, 2014. 1

[28] E. A. Tevelev. *Projective duality and homogeneous spaces*, volume 133. Springer Science & Business Media, 2006. 2, 6

[29] C. D. Toth, J. O'Rourke, and J. E. Goodman. *Handbook of discrete and computational geometry*. CRC press, 2004. 5

[30] P. L. Van Hove. Silhouette-slice theorems. Technical report, DTIC Document, 1986. 2

[31] J. V. Whittaker. A mountain-climbing problem. *Canad. J. Math*, 18:873–882, 1966. 4

[32] K.-Y. K. Wong and R. Cipolla. Reconstruction of sculpture from its profiles with unknown camera positions. *Image Processing, IEEE Transactions on*, 13(3):381–389, 2004. 2, 3

[33] L. Yi and D. W. Jacobs. Efficiently determining silhouette consistency. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007. 2, 5

[34] C. L. Zitnick, J. Sun, R. Szeliski, and S. Winder. Object instance recognition using triplets of feature symbols. Technical report, Technical report, Microsoft Research, Redmond, WA, USA, 2007. 1