Globally Optimal Pose from Orthographic Silhouettes

Agniva Sengupta[1,2] Dilara Kuş[1,2] Jianning Li[2] Stefan Zachow[2]
[1]Freie Universität Berlin [2]Zuse Institute Berlin

Abstract

We solve the problem of determining the pose of known shapes in $\mathbb{R}^3$ from their unoccluded silhouettes. The pose is determined up to global optimality using a simple yet under-explored property of the area-of-silhouette: its continuity w.r.t trajectories in the rotation space. The proposed method utilises pre-computed silhouette-signatures, modelled as a response surface of the area-of-silhouettes. Querying this silhouette-signature response surface for pose estimation leads to a strong branching of the rotation search space, making resolution-guided candidate search feasible. Additionally, we utilise the aspect ratio of 2D ellipses fitted to projected silhouettes as an auxiliary global shape signature to accelerate the pose search. This combined strategy forms the first method to efficiently estimate globally optimal pose from just the silhouettes, without being guided by correspondences, for any shape, irrespective of its convexity and genus. We validate our method on synthetic and real examples, demonstrating significantly improved accuracy against comparable approaches.[1]

1. Introduction

Estimating the global pose of 3D objects from a single image typically relies on point correspondences between the object template and the image. We address this pose-estimation problem using a much weaker cue – the object's projected silhouette; we call this the Pose-from-Silhouette (P*f*S) problem. Our approach discards all assumptions related to the object's shape, such as convexity or genus, and solve this P*f*S problem without additional information such as point correspondences or direct image intensities. Our input requirements are the projective silhouettes and a template of the object (in standard formats, e.g., triangulated mesh or semi-dense point cloud). Our solution is globally optimal when the shape-silhouette combination admits a unique solution. However, for symmetric shapes, global solutions for P*f*S may not be unique; in such cases, we estimate exactly one of these many redundant solutions – a reasonable outcome from silhouettes alone. This work presents the first globally optimal solution to the P*f*S problem.

Motivation. Solving P*f*S is appealing due to breadth of its motivating use-cases. Silhouettes are a key visual cue for diverse engineering applications, e.g.: autonomous driving [36], robotic manipulation [18], robotic navigation [2], Augmented Reality (AR) [7], surgical AR [8], space navigation [16], reconstruction from sparse views in medical imaging [30], automated flight control [39], and industrial quality control [27]. A closer inspection of these usages reveal silhouettes to be always used in conjunction with additional visual cues, be it feature correspondences, image intensities, or temporal priors – an unsurprising finding. P*f*S is indeed a fundamentally ill-posed problem; finding its global optima is challenging with no existing solutions.

Contributions. We solve the problem of *pose-estimation using just the unoccluded silhouettes of rigid objects* as input data. Our approach for solving this P*f*S problem involves pre-computing global shape signature responses for a given object-template and using this pre-computed shape-signature to infer pose from any arbitrary silhouette of this object. Specifically, we contribute the following:

- **Globally optimal silhouette-based pose estimation.** Building upon the framework of Hartley and Kahl [17], we introduce shape signatures of orthographic silhouettes that vary continuously with an object's orientation, enabling non-trivial branching strategies that lead to solutions converging to global optima, up to discretisation.
- **Geometric shape signatures and refinement.** We propose two intuitive yet powerful silhouette descriptors: the Area-of-Silhouettes (A*o*S) and the aspect ratio of ellipses fitted to the silhouettes; we show that they capture sufficient geometric information for accurate pose estimation. Orientation is recovered numerically, followed by a post-hoc non-linear refinement on the $\mathbb{SE}(3)$ manifold.
- **Generalisation to perspective imaging.** Assuming depth priors as input, the same formulation achieves *near-optimal accuracy* for perspective silhouettes, demonstrating strong practical performance while preserving an explainable, initialisation-free design.

We validate our approach on many synthetic and real datasets, strongly outperforming compared approaches.

[1]**Code and data:** agnivsen.github.io/pose-from-silhouette/

2. Background

Pose estimation from 2D silhouettes has been widely studied when considered in conjunction with 3D-to-2D point correspondences [22, 30] or with alternative correspondence cues such as contour generators [25], image textures [7, 10, 32], or both [28]. However, correspondence-free pose from just the 2D silhouettes/contours has not been solved up to global optimality for general shapes. There exist specialised approaches for specific shapes, such as the case of pose-estimation from ellipse-ellipsoid matches [13] (which can be considered as pose estimation from elliptical silhouettes), for surfaces of revolution [38], and for cylindrical objects [15]. None of these approaches, however, can be generalised to arbitrary shapes. There exist some older methods that utilise the notion of 'silhouette-lookup' [20, 21] to estimate poses of deformable structures from learned silhouette appearances along with some other methods [23, 34] that recognize objects based on the pencil of viewing tangent planes along the silhouette from smooth objects, but no efforts have been made to solve the pose estimation problem to global optimality, even in the rigid case. Recently, deep-learning based approaches [35] have been proposed for the PfS problem; unfortunately, they happen to be local methods necessitating an initial pose for their estimation process and requires colour information along object boundary. A Particle Swarm Optimisation (PSO) based approach has recently been proposed for perspective silhouettes [10], though it remains dependent on approximate depth bounds and remains stochastic with no optimality guarantees. **No globally optimal method exists for pose estimation from silhouettes alone.**

3. Method

Our methodological description begins by explaining the problem setup (Sec. 3.1) and problem statement (Sec. 3.2) followed by our strategy of branching the rotation space with global shape signatures (Sec. 3.3) which culminates in a globally optimal and efficient pose estimation approach (Sec. 3.4).

3.1. Problem Setup

We assume the availability of any standard shape prior (e.g.: triangulated/tetrahedral mesh, explicit/implicit surfaces, etc.) of the object that is to be localized by its silhouettes. A dense point cloud of 3D surface points is obtained from this shape prior via standard methods (e.g.: sampling [9] for meshes). We denote this point cloud as $\mathbf{Q} = [\mathbf{P}_1, \dots, \mathbf{P}_M] \in \mathbb{R}^{3\times M}$ maintaining M at a sufficiently large value to ensure accurate extraction of silhouettes from projections of $\mathbf{Q}$. We denote by $\Pi_{\mathrm{O}}(\mathbf{Q})$, the point cloud orthographically projected into the XY-plane, obtained by simply eliminating the third row of $\mathbf{Q}$. Silhouettes are represented discretely by the ordered sequence $\mathbf{G} = \langle \mathbf{s}_k \in \mathbb{R}^2 | k \in \mathbb{Z} \rangle$ of 2D points within the XY-plane. There exists a function $S(\cdot)$ that acts on a 2D point cloud computing the outer boundary of this point cloud with standard methods [11], i.e., $\mathbf{G} = S\big(\Pi_{\mathrm{O}}(\mathbf{Q})\big)$ if $\mathbf{G}$ is the orthographic silhouette of $\mathbf{Q}$. The input silhouette is given by $\mathbf{G}^*$

3.2. Problem Statement

The posed problem is defined in terms of the Hausdorff metric. Given an input silhouette $\mathbf{G}^*$ and an orthographic silhouette $\mathbf{G}$ of the template $\mathbf{Q}$, the Hausdorff distance between the two silhouettes is:

$$H(\mathbf{G}, \mathbf{G}^*) = \max\Big(\max_{\mathbf{s}_k \in \mathbf{G}} \big(\min_{\mathbf{s}_k^* \in \mathbf{G}^*} \|\mathbf{s}_k - \mathbf{s}_k^*\| \big), \max_{\mathbf{s}_k^* \in \mathbf{G}^*} \big(\min_{\mathbf{s}_k \in \mathbf{G}} \|\mathbf{s}_k^* - \mathbf{s}_k\| \big) \Big). \tag{1}$$

However, the orthographic silhouette of any model point cloud $\mathbf{Q}$ that has been rotated by $\mathbf{R} \in \mathbb{SO}(3)$ and translated by $(\mathbf{t}^\top, 0)^\top$ for some $\mathbf{t} \in \mathbb{R}^2$ is $\tilde{S}(\mathbf{Q}, \mathbf{R}, \mathbf{t}) = S\big(\Pi_{\mathrm{O}}(\mathbf{R}\mathbf{Q} + (\mathbf{t}^\top, 0)^\top)\big)$. Therefore, the P$f$S problem we want to solve relates to a combination of Eq. (1) and $\tilde{S}$ as follows:

$$\boxed{\min_{\mathbf{R}\in\mathbb{SO}(3), \mathbf{t}\in\mathbf{R}^2} H(\tilde{\mathbf{G}}, \mathbf{G}^*), \quad \text{s.t.:} \quad \tilde{\mathbf{G}} = \tilde{S}(\mathbf{Q}, \mathbf{R}, \mathbf{t})}. \tag{2}$$

Solving Eq. (2) is challenging due to non-convexity of the $\mathbb{SO}(3)$ manifold and the Hausdorff distance $H(\tilde{\mathbf{G}}, \mathbf{G}^*)$, exacerbated by the possible non-uniqueness of globally-optimal solutions. Importantly, we do not make any assumptions about shape symmetry, convexity, or genus for our proposed method. However, we do make a mild assumption about the solution space of Eq. (2):

Assumption 1. *There exists at least one pair of roto-translations* $(\mathbf{R}_{\mathrm{opt}}, \mathbf{t}_{\mathrm{opt}})$ *such that the Hausdorff distance* $H(\tilde{\mathbf{G}}, \mathbf{G}^*) \sim 0$ *for* $\tilde{\mathbf{G}} = \tilde{S}(\mathbf{Q}, \mathbf{R}_{\mathrm{opt}}, \mathbf{t}_{\mathrm{opt}})$.

Assumption 1 is a practically reasonable emphasis on presence of at least one solution in the solution space.

Input shape requirement. A necessary condition: projected silhouette of $\mathbf{Q}$ admits well-defined area for all $\mathbf{R} \in \mathbb{SO}(3)$, ruling out degenerate projections collapsing to lower-dimensional sets, e.g.: points, lines, or curves.

3.3. Rotation Space Branching

A natural approach for solving Eq. (2) is to identify a small subset of $\mathbb{SO}(3)$ that encompasses its feasible set. Our proposed method identifies such subsets by leveraging global shape signatures of silhouettes, irrespective of their non-uniqueness, to partition the $\mathbb{SO}(3)$ search space.

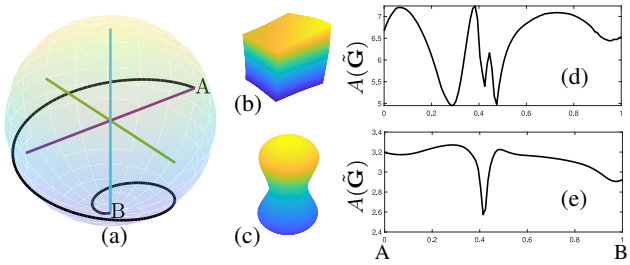


Figure 1. When a L-c trajectory from A to B on an unit sphere (a) is mapped to $\mathbb{SO}(3)$ and applied to two arbitrary shapes (b,c), the resulting evolution of orthographic AoS, as shown in (d,e) with its abscissa mapped as $[A, B] \mapsto [0, 1]$, is L-c

3.3.1. Area of silhouettes as shape signature

We use the area of a region enclosed by the silhouette as a global geometric feature, to subdivide the search space. We introduce the function A that acts on a silhouette to compute the area enclosed by it. This area can be regarded as a signature of the shape that is given by the silhouette. We begin with some simple observations about this shape signature:

Theorem 1. *$A(\tilde{\mathbf{G}})$ is Lipschitz-continuous (L-c) w.r.t. any sequence of rotations of $\mathbf{Q}$ in $\mathbb{R}^3$ as long as the sequence of rotations is L-c and $\mathbf{Q}$ is representable with a finite collection of triangles (e.g., a triangulated mesh).*

Proof. See section 1 of supplementary. □

As a consequence of Theorem 1, A is differentiable 'almost everywhere' in the compact subset of $\mathbb{R}^9$ embedding $\mathbb{SO}(3)$ due to Rademacher's theorem, and more importantly, have bounded gradients. Although silhouette boundary computation via numeric methods [11] from discrete representations (e.g.: dense point cloud) for complicated shapes may introduce artefacts in the computed AoS, this does not appear to be a problem in practical cases; we offer two examples of the L-c-ness of computed AoS in arbitrary shapes, shown in Fig. 1. Thus, if the translation $\mathbf{t}$ in Eq. (2) is accounted for and if there exists a map $\vartheta : \mathbb{SO}(3) \mapsto \mathbb{R}$ mapping every point of the $\mathbb{SO}(3)$ manifold to the corresponding area of $\tilde{\mathbf{G}}$, then the intersection of the input silhouette's area $A(\mathbf{G}^*)$ with the continuous surface of all possible values of $A(\tilde{\mathbf{G}})$, must contain the global optima, since $H(\tilde{\mathbf{G}}, \mathbf{G}^*) \sim 0$ implies $|A(\mathbf{G}^*) - A(\tilde{\mathbf{G}})| \sim 0$. Nonetheless, a conventional Branch-and-Bound (BnB) search over $\mathbb{SO}(3)$ (e.g.: [17]) is a very expensive proposition. Section 3.3.2 demonstrates a strategy for significant reduction in search space.

3.3.2. Search space reduction

Our goal now is to find the subspace of $\mathbb{SO}(3) \times \mathbb{R}^2$ where $|A(\mathbf{G}^*) - A(\tilde{\mathbf{G}})| \sim 0$ is satisfied. However, the translational component $\mathbf{t} \in \mathbb{R}^2$ is uninteresting, since: I) it can be determined by simply comparing the centroids of modelled and input silhouettes, and II) more importantly, for any silhouette $\tilde{\mathbf{G}}$ modelled as $\tilde{\mathbf{G}} = \tilde{S}(\mathbf{Q}, \mathbf{R}, \mathbf{t})$, the area function $A(\tilde{\mathbf{G}})$ is invariant w.r.t $\mathbf{t} \in \mathbb{R}^2$ and rotation along Z-axis. This motivates solving $\mathbf{R} \in \mathbb{SO}(3)$ independently of $\mathbf{t} \in \mathbb{R}^2$. Use of area-based shape signature causes $\mathbb{SO}(3)$ to further split into two categories: a) rotation along X and Y axes ($\mathfrak{R}_{XY}$) that causes $A(\tilde{\mathbf{G}})$ to vary smoothly and b) rotation along Z axis ($\mathfrak{R}_Z$) that leaves $A(\tilde{\mathbf{G}})$ invariant (projection is on the XY-plane). Thus, the area-based search can be restricted to $\mathfrak{R}_{XY}$, while $\mathfrak{R}_Z$ requires a different signature (given later in Sec. 3.4). Trivially deriving $\mathbf{R}$ from Euler angles is problematic due to non-commutativity of rotations in $\mathfrak{R}_{XY}$ and $\mathfrak{R}_Z$, thus necessitating a re-parametrisation.

Postel projection. The 'azimuthal-equidistant' projection, more commonly known as Postel projection, is a map from a rotation of magnitude α along some unit vector $\hat{\mathbf{v}}$ to a point $\alpha\hat{\mathbf{v}} \in [-\pi, \pi]^3 \subset \mathbb{R}^3$ [1, 17]. The Postel map is typically represented as a map via the quaternion space, given as $\left(\cos\left(\frac{\alpha}{2}\right),\ \sin\left(\frac{\alpha}{2}\right)\hat{\mathbf{v}}^\top\right)^\top \mapsto \alpha\hat{\mathbf{v}}$. But the entire $[-\pi, \pi]^3$ space is redundant as well; there exists a sphere of radius π centred at origin, we term this sphere as the Postel ball $\mathcal{S}_\pi$, and every point in $[-\pi, \pi]^3$ outside $\mathcal{S}_\pi$ represents a rotation duplicated with some other point inside $\mathcal{S}_\pi$. We represent by $F : \mathbb{R}^3 \mapsto \mathbb{SO}(3)$, a function that takes any point inside $\mathcal{S}_\pi$ (except its centre) to its equivalent rotation matrix in $\mathbb{SO}(3)$ by a function $F\left((\alpha, \mathbf{v}^\top)^\top\right)$, detailed in section 6 of the supplementary.

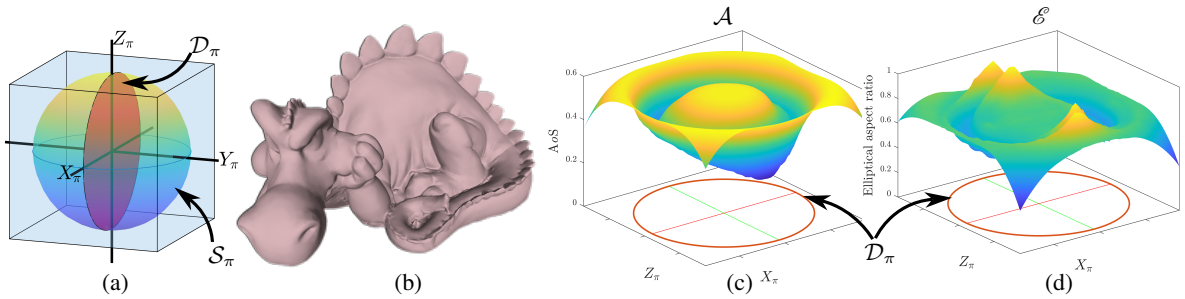
Rotation sufficiency over a disc. The search for some specific AoS can therefore be confined to $\mathcal{S}_\pi$ without loss of generality. For some $(\alpha, \mathbf{v}) \in \mathcal{S}_\pi$, the A$o$S is obtained by computing $\mathbf{R} \in \mathbb{SO}(3)$ using $F(\cdot)$, and applying this rotation to $\mathbf{Q}$ before computing $A\left(\tilde{S}(\mathbf{Q}, \mathbf{R}, \mathbf{0}_3)\right)$. However, there are still some redundancies in $\mathcal{S}_\pi$, since it represents both $\mathfrak{R}_{XY}$ and $\mathfrak{R}_Z$, i.e.,:

Lemma 1. *Given some $F : (\alpha, \mathbf{v}) \mapsto \mathbf{R}$, we have $A\left(\tilde{S}(\mathbf{Q}, \mathbf{R}, \mathbf{0}_3)\right) = A\left(\tilde{S}(\mathbf{Q}, \mathbf{R}_x, \mathbf{0}_3)\right)$ for all $\mathbf{R}_x = F\left((\alpha, \mathbf{v}_x^\top)^\top\right)$, if $\langle \mathbf{v}, (0, 0, 1)^\top \rangle = \langle \mathbf{v}_x, (0, 0, 1)^\top \rangle$.*

Proof. See section 2 of supplementary. □

Thus, instead of sampling the highly redundant $\mathcal{S}_\pi$, we sample only the disc of the intersection of $\mathcal{S}_\pi$ with its XZ-plane, we call this the Postel disc $\mathcal{D}_\pi$ and, for any point $\mathbf{d} \in \mathcal{D}_\pi \subset \mathbb{R}^2$, we denote by the function $G(\mathbf{d}) = (\alpha, \mathbf{v}^\top)^\top = (\|\mathbf{d}\|, \mathbf{d}_1/\|\mathbf{d}\|, 0, \mathbf{d}_2/\|\mathbf{d}\|)^\top$, the invertible map from $\mathcal{D}_\pi$ to the $\mathcal{S}_\pi$.

Silhouette area signature. Given that the set of all possible projected area values can be pre-computed from just $\mathcal{D}_\pi$, we do so by semi-densely sampling $\mathcal{D}_\pi$ and recording the projected area profile as a shape signature, we term this signature Projected Area Response Surface (PARS); which is a non-injective map $\mathcal{A} : \mathcal{D}_\pi \mapsto \mathbb{R}$ and surjective to some subset of $\mathbb{R}$. Pre-computation of PARS can be done as given in



(a) (b) (c) (d)

Figure 2. (a) Shows the Postel ball ($\mathcal{S}_\pi$) and disc ($\mathcal{D}_\pi$) inside a cube of length π with axes (X_π, Y_π, Z_π), (b) shows the 3D template of PD, (c) shows the mapping of $\mathcal{D}_\pi$ to AoS for PD, and (d) shows the mapping of $\mathcal{D}_\pi$ to elliptical aspect ratio for PD

algorithm-1 of supplementary. An example of PARS for the triangulated 3D model of Phlegmatic Dragon (PD) [12] is shown in Fig. 2.

3.4. Pose Estimation

We now show how to utilise pre-computed AoS signatures to estimate pose up to global optimality.

3.4.1. Globally optimal pose from area of silhouettes

With pre-computed PARS, pose inference from a silhouette involves: a) intersecting the input AoS with PARS, b) identifying 'candidate' rotations for pose estimation, and c) efficiently searching among them for the optimal pose.

Area-PARS intersection. First the translation between $\mathbf{G}^*$ and $\tilde{\mathbf{G}}$ is computed as the difference of centroids $\mathbf{t} = C(\tilde{S}(\mathbf{Q}, \mathbf{I}_3, \mathbf{0})) - C(\mathbf{G}^*)$, where $C(\cdot)$ computes the centroid of the 2D points. Next, the iso-contour of the intersection of $A(\mathbf{G}^*)$ with the map $\mathcal{A}$ is determined by standard contouring operations, namely, by first upgrading the semi-dense $\mathcal{A}$ to a continuous surface and then computing the intersection with $A(\mathbf{G}^*)$ using marching squares [31], giving us $U_\mathcal{A} = \{\mathbf{d}_j \in \mathcal{D}_\pi,\ j \in [1, N_D]\}$ discrete set of intersecting points, such that:

$$|A\Big(\tilde{S}\Big(\mathbf{Q}, F\big(G(\mathbf{d}_j)\big), \mathbf{t}\Big)\Big) - A(\mathbf{G}^*)| \leqslant \epsilon_{xy}, \qquad (3)$$

where ϵ_{xy} is a tunable threshold.

Candidate solution search. Since $\{\mathbf{d}j\}$ is derived from $\mathcal{D}_\pi$, the solution set spans $\mathfrak{R}_{XY}$ but not $\mathfrak{R}_Z$, allowing arbitrary Z-axis rotations without affecting projected AoS (Lemma 1). Evaluating AoS alone in 2D provides no further template pose validation. Aligning $\tilde{\mathbf{G}}$ and $\mathbf{G}^*$ via dominant singular values is a potential approach, but computationally expensive for large N_D. Instead, we leverage the 1D projections of $\tilde{\mathbf{G}}$ and $\mathbf{G}^*$ along X and Y, where their lengths, $L_x(\tilde{\mathbf{G}})$ and $L_y(\tilde{\mathbf{G}})$, provide useful additional information. We thus seek a rotation matrix $\mathbf{R}_z$ such that, with $\mathbf{R}_c = \mathbf{R}_z F\big(G(\mathbf{d}_j)\big)$, we have:

$$\begin{aligned} |L_x\Big(\tilde{S}\Big(\mathbf{Q}, \mathbf{R}_c, \mathbf{t}\Big)\Big) - L_x(\mathbf{G}^*)| \leqslant \epsilon_z \ \wedge \\ |L_y\Big(\tilde{S}\Big(\mathbf{Q}, \mathbf{R}_c, \mathbf{t}\Big)\Big) - L_y(\mathbf{G}^*)| \leqslant \epsilon_z, \end{aligned} \qquad (4)$$

where ϵ_z is a threshold. With uniform samples along Z-axis $\theta_{z,k} \in U(0, 2\pi)$, we accept a pose as a 'candidate' for a good solution if substituting $\mathbf{R}_z = M(\theta_{z,k}),\ \forall k \in [1, N_Z]$ in Eq. (4), for some N_Z, satisfies both of its conditions; $M(\theta_{z,k})$ gives the rotation matrix from the Euler angle $\theta_{z,k}$ along Z-axis. Thus, for every point $\mathbf{d}_j$, we get a set of rotation matrices:

$$C_j = \{M(\theta_{z,k'}) F\big(G(\mathbf{d}_j)\big),\ \forall k' \in [1, N'_{Z,j}]\}, \qquad (5)$$

where every rotation matrix in C_j satisfies Eq. (4). Thus $\tilde{C} = \bigcup_{j=1}^{N_D} C_j$ gives us the global set of candidate solutions. Thereafter, the filtering of $\tilde{C}$ into actual solutions of Eq. (2) is done by exhaustive search in this reduced feasible set $\tilde{C}$. Importantly, we show the existence of ϵ_o-global optimality, meaning a candidate solution exists whose distance to global optima in $\mathbb{SO}(3)$ is bounded by ϵ_o.

Theorem 2. *(ϵ_o-global optimality) There must exist an element of $\tilde{C}$ which lies within a ball of finite radius ϵ_o in $\mathbb{SO}(3)$ from the global optima of Eq. (2) with $\lim_{\epsilon_{xy}, \epsilon_z \to 0} \epsilon_o = 0$ and $(\epsilon_{xy}, \epsilon_z)$ are free parameters of sampling, controllable to approach zero.*

Proof. See section 3 of supplementary. □

In practice, we maintain an upper bound on $|\tilde{C}| \leqslant \lambda_c$, randomly eliminating candidates from $\tilde{C}$ if its count exceeds λ_c, to aid faster convergence; more details in algorithm-2 (supplementary). The guarantee in Theorem 2 trivially extends to scaled-orthographic projection; the optimality conditions and convexity properties of the orthographic case remain unchanged under global scaling.

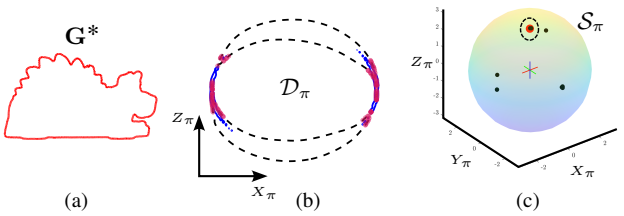


Figure 3. Example of pose estimation: (a) shows input silhouette from PD, (b) shows $U_{\mathcal{A}}$ in black-dashes, $U_{\mathscr{E}}$ in blue, and $U_{\mathcal{A}\cap\mathscr{E}}$ in red on $\mathcal{D}_\pi$, and (c) shows $\tilde{C}$ in black and the Groundtruth (G*t*) pose in red (encircled) inside $\mathcal{S}_\pi$; the PARS and PEARS for PD are given in Fig. 2

3.4.2. Ellipse aspect-ratio driven acceleration

$\tilde{C}$, obtained by sampling the Postel ball for equality of projected silhouette area, already strongly branches the search space, yet further branching remains possible - which can be verified by observing that any global shape signature, if L-*c* w.r.t rotation, can serve to subdivide the Postel ball in a manner analogous to the strategy (with area) in Lemma 1.

Given $\tilde{\mathbf{G}}$, we fit an ellipse $\mathscr{E}$ algebraically [4], noting that its aspect ratio $AR_{\mathscr{E}}$, i.e., its major-to-minor axis ratio, is heuristically L-*c* w.r.t. $\mathbf{R}$ and $\mathbf{t}$ in most cases. Since $AR_{\mathscr{E}}$ is used solely for acceleration, the lack of a formal proof of its L-*c*-ness w.r.t. $\mathbf{R}$ does not affect global optimality, as shown in algorithm-2 (supplementary).

Silhouette elliptical aspect ratio signature. Following the same strategy as algorithm-1 (supplementary), we learn the mapping from $\mathcal{D}_\pi$ to $AR_{\mathscr{E}}$ via semi-dense random sampling of $\mathcal{D}_\pi$, creating a new shape signature Projected Elliptical Aspect Response Surface (PEARS), which we denote by the non-injective map $\mathscr{E} : \mathcal{D}_\pi \mapsto \mathbb{R}$.

Accelerated pose estimation. The pose estimation follows two steps: we *first* estimate the contour of the intersection of $AR_{\mathscr{E}}(\mathbf{G}^*)$, denoted as a function computing $AR_{\mathscr{E}}$ of $\mathbf{G}^*$ with a mild abuse of notation, with $\mathcal{D}_\pi$, giving us $U_{\mathscr{E}} = \{\mathbf{h}_{j'} \in \mathcal{D}_\pi,\ j' \in [1, N_E]\}$ discrete set of intersecting points such that:

$$|A\Big(\tilde{S}\Big(\mathbf{Q}, F\big(G(\mathbf{h}_{j'})\big), \mathbf{t}\Big)\Big) - A(\mathbf{G}^*)| \leqslant \epsilon_e, \qquad (6)$$

where ϵ_e is a tunable threshold. For the *second* step, we find the nearest neighbours between $\{\mathbf{d}_j\}$ and $\{\mathbf{h}_{j'}\}$, giving us $U_{\mathcal{A}\cap\mathscr{E}} = \{\tilde{\mathbf{d}}_h \in \mathcal{D}_\pi,\ h \in [1, N_{\mathcal{A}\cap\mathscr{E}}]\}$, such that $U_{\mathcal{A}\cap\mathscr{E}} \subseteq U_{\mathcal{A}}$ and :

$$\exists j'' \in [1, N_E] \text{ s.t.: } \|\tilde{\mathbf{d}}_h - \mathbf{h}_{j''}\| \leqslant \epsilon_\cap, \forall h \in [1, N_{\mathcal{A}\cap\mathscr{E}}], \quad (7)$$

where $\epsilon_\cap$ is a parameter denoting an infinitesimal circle around each $\tilde{\mathbf{d}}_h$ which is considered part of the intersection $U_{\mathscr{E}} \cap U_{\mathcal{A}}$. We show an example of the pose estimation outcome with PD in Fig. 3.

Resolution pyramid. Our final estimation strategy follows a pyramidal approach, initialising search bounds $(\epsilon_{xy}, \epsilon_z, \epsilon_e, \epsilon_\cap)$ to satisfy $H(\tilde{\mathbf{G}}, \mathbf{G}^*) \leqslant \epsilon_H$, a desired precision. If violated, typically due to noisy silhouettes, only $(\epsilon_\cap, \epsilon_z)$ are incremented and retried. This iterative algorithm is detailed in algorithm-2 (supplementary).

We denote the pose from algorithm-2 (supplementary) as Globally Optimal P*f*S (GlOptiPoS) and algorithm-2 (supplementary) followed by non-linear refinement (Sec. 3.5) as GlOptiPoS + Non-linear Refinement (GlOptiPoS+).

3.5. Non-linear Refinement

The solution pose $(\mathbf{R}_{\text{opt}}, \mathbf{t}_{\text{opt}})$ from algorithm-2 (supplementary) may deviate from the global optimum of Eq. (2) due to sampling resolutions $(\epsilon_{xy}, \epsilon_z, \epsilon_\cap, \epsilon_H)$ and noise. To mitigate, we apply local, non-linear refinement, optimizing Eq. (2) while initializing $(\mathbf{R}, \mathbf{t})$ with $(\mathbf{R}_{\text{opt}}, \mathbf{t}_{\text{opt}})$. The cost is iteratively minimized via steps along the $\mathbb{SE}(3)$ tangent plane, followed by manifold retraction using standard optimization tools [6]. The final refined pose is $(\mathbf{R}_{\text{ref}}, \mathbf{t}_{\text{ref}})$.

3.6. Perspective Approximation

The optimality guarantee of Theorem 2 does not extend directly to perspective projection, since perspectivity couples silhouette-based signatures with object translation. Incorporating translation into the $\mathbb{R}^3$ search space would cause exponential growth, rendering B*n*B-style global optimization impractical. Hence, following prior work [10], we *assume availability of a coarse depth prior*, obtainable from RGB-D or monocular estimation (e.g.:[5]). The perspective silhouettes are $\mathbf{G}_\Pi = S(\Pi(\mathbf{Q})), \Pi(\cdot)$ being perspective projection, leading to the perspective equivalent of Eq. (2), with $\mathbf{t} \in \mathbb{R}^3$. Theorem 1 can be extended to perspective projection, since:

Lemma 2. *$A(\tilde{\mathbf{G}}_\Pi)$ is L-c w.r.t. any sequence of rotations of $\mathbf{Q}$ in $\mathbb{R}^3$ as long as the sequence of rotations is L-c, $\mathbf{Q} \notin \{Z = 0\}$, and $\mathbf{Q}$ is representable with a finite collection of triangles*

Proof. See section 4 of supplementary. □

PEARS is equivalently heuristically L-*c* and therefore used for acceleration following Sec. 3.4.2. Given a prior depth, PARS and PEARS are pre-computed offline perspectively at that depth while $\mathbf{G}^*_\Pi$ is transformed to normalised image coordinates, assuming known intrinsics. Non-linear refinement follows Sec. 3.5, replacing $\mathbf{G}$ with $\mathbf{G}_\Pi$. The details of the pipeline are omitted for brevity, obtained as the perspective variants of algorithm 2 (supplementary) given by GlOptiPoS Perspective (GlOptiPoS$_\Pi$) and GlOptiPoS$_\Pi$ followed by perspective variant of Sec. 3.5 as GlOptiPoS+ Perspective (GlOptiPoS$_\Pi$+).

		NlR			Nl-PaR			Ms-GO			STI-Pose$_{\Pi_O}$			GlOptiPoS+		
		RMSE ↓	OE ↓	TE ↓	RMSE ↓	OE ↓	TE ↓	RMSE ↓	OE ↓	TE ↓	RMSE ↓	OE ↓	TE ↓	RMSE ↓	OE ↓	TE ↓
SB	Mean	42.35	54.66	8.85	43.13	55.62	7.97	28.89	33.32	**3.87**	**9.75**	**3.12**	9.74	**0.46**	**0.32**	**0.14**
	SD	8.18	25.48	8.85	**4.18**	25.2	6.51	14.53	30.64	**1.99**	9.07	**13.63**	9.06	**0.35**	**0.36**	**0.15**
	Max.	64.23	**101.03**	46.3	50.37	110.55	27.2	**48.46**	104.25	**9.91**	67.77	111.45	67.76	**1.6**	**2.2**	**0.79**
PD	Mean	38.22	44.93	5.9	36.77	40.17	5.18	**32.58**	38.73	**4.18**	101.55	**4.29**	101.41	**0.91**	**0.61**	**0.32**
	SD	**7.85**	23.41	3.67	8.96	26.54	4.42	11.88	31.63	**2.29**	248.78	**16.9**	248.68	**0.92**	**0.88**	**0.32**
	Max.	50.01	**94.51**	17.3	48.4	118.91	18.82	**48.38**	114.81	**10.56**	1879.94	118.87	1879.57	**7**	**7.43**	**2.16**
PB	Mean	35.86	41.87	8.66	37.72	39.25	12.41	**32.86**	37.17	**5.41**	78.99	**3.47**	78.9	**0.76**	**0.5**	**0.26**
	SD	11.03	28.98	6.54	**9.3**	31.44	7.17	13.44	29.21	**3.3**	180.81	**15.15**	180.7	**0.95**	**0.93**	**0.27**
	Max.	50.22	100.79	31.7	**49.78**	106.18	25.86	51.21	**97.92**	**14.92**	1362.52	105.64	1362.14	**7.41**	**8.58**	**1.48**

Table 1. Accuracy of the methods N*l*R, N*l*-P*a*R, M*s*-GO, STI-Pose$_{\Pi_O}$, and GlOptiPoS+ on SB, PD, and PB (best values in each category are marked with **bold-underlines**, second-best values are in **bold**, very large error values have been highlighted in red)

4. Experimental Results

We empirically verify our approach below.

Experimental setup. Using three 3D models Stanford Bunny (SB), PD, and Pelvic Bone (PB) (resp.), visualized in section 7 of supplementary, we use point clouds of 29072, 29120, and 28976 points randomly sampled on their surface to obtain $\mathcal{A}$ and $\mathcal{E}$ via algorithm-1 (supplementary). For pose estimation, independently sampled point clouds are used to ensure a separation between offline pre-computation and pose estimation data. Each experiment applies roto-translation via $\mathbb{SE}(3)$ random-sampling, followed by orthographic projection and silhouette extraction. We also use the 20 objects in the real data from *B*inocular *O*bject *T*racking (B*c*OT) benchmark dataset [24], specifically in two modes: I) for statistical validation, random poses near the provided G*t* are simulated to render silhouettes perspectively (all methods evaluated under identical conditions), II) segmented objects from the '*complex_movable_handheld*' sequence of B*c*OT are used to estimate pose. Randomly sampled surfaces of the 20 objects in B*c*OT are used to learn $\mathcal{A}$ and $\mathcal{E}$ using perspective variant of algorithm-1 (supplementary) at a constant empirical depth prior of 80 *cm* – all perspective experiments are given without re-learning this depth prior.

Metric. We use three error metrics: I) Orientation Error (OE), the mean absolute Euler angles of the optimal orientation between G*t* and estimated posed shape, computed with Horn's method [19], II) Translational Error (TE), computed as $\|\mathbf{t}_{\text{gt}} - \mathbf{t}_{\text{est}}\|_2$, and III) Root Mean Square Error (RMSE), measuring the difference between $\mathbf{Q}_{\text{gt}}$ and $\mathbf{Q}_{\text{est}}$. Here, the suffix 'gt' and 'est' denotes the G*t* and estimated values (resp.). TE and RMSE are expressed as percentages of Largest Diagonal of Bounding-Box (LD*o*BB) for the 3D templates of PD, SB, and PB and for B*c*OT, they are expressed in *mm*.

Compared methods. We compare against three variants of P*f*S solutions that are intuitive: I) Non-linear Refinement (N*l*R), a non-convex solver using Levenberg-Marquardt (LM) [26] parametrised on Lie algebra [14] of the rotation group, II) Non-linear Project-and-Refine (N*l*-P*a*R), an iterative projection and refinement method aligning $\mathbf{G}^*$ and $\tilde{\mathbf{G}}$ via Euclidean distance minimisation, and III) Multi start - Global Optimization (M*s*-GO) following [33] while solving Eq. (2) directly; *details of these approaches are given in section-7-of-supplementary*. Importantly, N*l*R, N*l*-P*a*R, and M*s*-GO are 'pose-estimation methods' (and not reconstruction or registration methods). We do extensive comparison against [10], denoted by Silhouette Texture Independent Pose (STI-Pose), as our closest recent baseline method. We also adapt STI-Pose into an orthographic equivalent STI-Pose (orthographic) (STI-Pose$_{\Pi_O}$) for fair comparison with orthographic silhouettes.

4.1. Orthographic Silhouettes

We present our results on orthographic silhouettes below.

Experiment. We validate our approach by estimating P*f*S with GlOptiPoS+, repeated 200 times on randomly posed silhouettes of PD, SB, and PB, with additive silhouette noise of Standard-Deviation (SD) $\sim 1\%$ of LD*o*BB. The compared methods N*l*R, N*l*-P*a*R, M*s*-GO, and STI-Pose$_{\Pi_O}$ undergo identical experiments, with results in Tab. 1.

Comparison. STI-Pose$_{\Pi_O}$ is the second-best method in all experiments. Problems due to its stochastic nature is clear from the maximum error values of STI-Pose$_{\Pi_O}$, which are higher than N*l*R, N*l*-P*a*R, and M*s*-GO (OE of $\sim 110^\circ$ is practically unusable). Our method is 89.74%, 85.78%, and 85.59% better than STI-Pose$_{\Pi_O}$ in mean OE – a significant improvement – while our worst case OE is $\sim 8.6^\circ$ for PB, which is due to numerical artefacts and not typically 'catastrophic' for most use-cases. More importantly, mean OE of our method for all shapes are $\ll 1^\circ$, confirming the power of global optimality in P*f*S. Error histograms of GlOptiPoS and GlOptiPoS+, given in section 7 (supplementary), confirm robustness across all experiments. Notably, GlOptiPoS exhibits higher TE accuracy since its translation is in a closed-form, while GlOptiPoS+ optimises on the $\mathbb{SE}(3)$ manifold, prioritising OE at minor expense to TE.

Symmetry as proxy for algorithmic complexity. A formal derivation of GlOptiPoS's algorithmic complexity is intractable, however, the number of candidate solutions $|\tilde{C}|$ intuitively scales with the object's rotational symmetry, e.g.: a perfect sphere yields theoretically infinite solutions. To validate this intuition, we synthesize real spherical-harmonic shapes, denoted with a mild notational-abuse as

Y_l^m following [37], with $l \in [2, 6] \subset \mathbb{Z}^+$ and $m = l-1$, ordered reverse-chronologically by decreasing symmetry (labels *B–F* in Fig. 6I; *A* denotes the sphere). The number of candidates obtained by GlOptiPoS over 25 trials (Fig. 6II) shows approximately logarithmic decay with diminishing symmetry, while the sphere exhibits exponentially higher $|\tilde{C}|$. This confirms: (I) $|\tilde{C}|$ contracts as symmetry reduces, implying higher efficiency for real-world shapes; and (II) for highly symmetric objects, silhouette-only pose estimation is inherently ambiguous – a geometric truism.

Effect of noisy silhouettes on accuracy. Evaluating silhouette noise impact on pose estimation is done with PB, a genus-1 shape (Tab. 1). Three noise regimes are considered: *low* (SD $\approx$ 1% of LD*o*BB), *medium* (SD $\approx$ 2% of LD*o*BB), and *high* (SD $\approx$ 4% of LD*o*BB); exemplar noisy silhouettes are shown in section 7 of supplementary. A pose estimate is deemed 'successful' if OE $\leqslant 6^\circ$ and TE $\leqslant$ 2% of LD*o*BB. Unlike Tab. 1, we analyse not only the best solution but seven best candidate poses from $\tilde{C}$ of GlOptiPoS, corresponding to the lowest $H(\mathbf{G}, \mathbf{G}^*)$ values (Eq. (1)). Results in Fig. 4 gives success percentages. Increasing noise shifts optimal candidates away from G*t*. For *low* and *medium* noise, top-ranked candidate is successful in 100% of trials (confirming Tab. 1). *High* noise shifts candidates deeper in hierarchy, sometimes beyond the first seven levels. The behaviour is consistent with graceful noise degradation: with sufficient candidate sampling (algorithm-2, supplementary), valid poses are eventually recovered.

Impact of parameters on runtime and accuracy. We analyse GlOptiPoS varying thresholds for filtering candidate poses, detecting intersections, and adjusting template point-count for signature computation, focusing on PD as the toughest case (Tab. 1). Figure 5a show RMSE and time changes with $\epsilon_z \in [0, 0.15]$ in 7 steps; accuracy remains stable, but outliers vanish at $\epsilon_z = 0.15$, albeit with increased runtime. Figure 5b reveal a V-shaped RMSE response to $\epsilon_\cap \in [0, 0.15]$, minimizing at $\epsilon_\cap \sim 0.08$ due to maintaining $\lambda_c \sim 10^2$. Figure 5c indicate accuracy rises with increasing $P \in [100, 29121]$, though for greater runtime – an expected trade-off. Our implementation is in MATLAB executed on Intel® Core™ i9-10920X (24-core) CPU, 128 GB RAM. Importantly, such runtimes are typical of B*n*B methods and remains parallelisable as a future work.

4.2. Perspective Silhouettes

We present our results on perspective silhouettes below.

Experiment. For GlOptiPoS$_\Pi$+, two depth-prior settings are used: I) the nominal pre-computation depth for $\mathcal{A}$ and $\mathcal{E}$, and II) a perturbed setting (GlOptiPoS$_\Pi \pm 8$) with random ± 8 *cm* deviation from learned signatures, exceeding typical depth-sensor noise. For STI-Pose, authors' provided depth bounds ($[5, 100]$ *cm*) yielded unstable results; hence, bounds were improved to provide meaningful results: within $\pm$ 10 *cm* and $\pm$ 5 *cm* of G*t* depth, denoted as STI-Pose-A and STI-Pose-B (resp.). The results are summarised in Tabs. 2 to 4. The B*c*OT objects were grouped into 6 symmetric and 14 asymmetric objects, colour-coded in Tabs. 2 to 4; sample qualitative results on asymmetric shapes from *complex_movable_handheld* sequence are shown in Fig. 7, with extensive qualitative results in section

	STI-Pose-A		STI-Pose-B		GlOptiPoS$_\Pi \pm 8$		GlOptiPoS$_\Pi$+	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ape	32.74	28.07	**10.01**	**1.55**	54.61	28.99	**1.6**	**2.25**
Cat	31.71	21.98	**19.29**	**9.87**	78.8	10.88	**0.72**	**0.46**
Jack	32.04	35.14	10.71	19.05	**10.04**	**4.01**	**1.75**	**2.22**
Squirrel	**36.2**	42.85	38.08	**42.64**	121.91	79.19	**2.1**	**2.14**
Stitch	29.07	33.34	18.87	27.74	**8.26**	**1.14**	**0.71**	**0.41**
Vampire queen	49.53	32.61	**26.98**	**6.77**	57.6	8.07	**2.24**	**1.76**
3D Touch	48.61	31.62	33.4	19.99	**21.21**	**7.44**	**3.99**	**12.21**
Auto GPS	93.44	47.81	88.39	41.51	**52.24**	**27.05**	**1.57**	**2.05**
Bracket	47.35	50.1	**15.06**	**1.79**	17.41	**11**	**10.26**	27.87
Deadpool	40.74	36.63	**15.08**	**11.74**	67.12	38.58	**4**	**12.27**
Driller	76.93	61.11	62.74	51.7	**8.22**	**1.21**	**1.31**	**1.36**
Flashlight	57.25	42.57	**42.51**	33.59	88.76	**9.31**	**1.37**	**1.35**
Lamp clamp	**73.2**	**49.22**	77.03	**44.52**	113.25	61.66	**36.55**	50.63
Lego	58.35	28.78	**29.5**	**25.1**	47.3	32.09	**19.28**	**23.47**
RJ45	47.62	48.01	**32.22**	38.76	94.57	**24.81**	**5.39**	**19.25**
RTI Arm	110.85	59.98	79.31	66.97	**41.53**	**27.78**	**32.37**	**43.8**
Standtube	36.68	42.23	**29.05**	**34.91**	109.56	65.78	**1.08**	**0.71**
Teapot	46.92	40.43	**25.41**	30.98	81.4	**10.12**	**10.51**	**24.57**
Tube	47.84	40.6	27.96	29.2	**21.71**	**3.6**	**1.36**	**0.82**
Wall Shelf	48.32	39.88	37.3	34.58	**10.24**	**2.13**	**0.76**	**0.72**

Table 2. RMSE(↓) in *mm*: mean and SD for STI-Pose-A, STI-Pose-B, GlOptiPoS$_\Pi$+, and GlOptiPoS$_\Pi \pm 8$ for the 20 B*c*OT objects. Asymmetric and symmetric objects have been colour-coded **blue** and **red** respectively (best values in each category are marked with **bold-underlines**, second-best values are in **bold**)

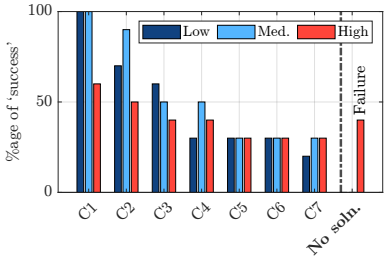


Figure 4. Success rate of pose estimation across the seven best solution candidates, denoted Cx ($x \in [1, 7]$), including some failure cases only for 'high' noise

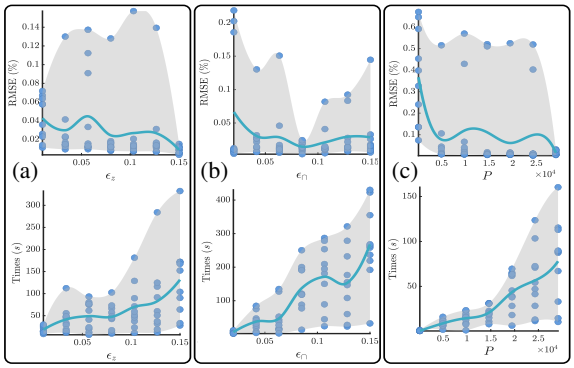


Figure 5. RMSE (top-row) and runtime (bottom-row) of GlOptiPoS under varying (a) ϵ_z, (b) $\epsilon_\cap$, and (c) P. Blue curves denote interpolated mean-curves (please zoom-in).

	STI-Pose-A		STI-Pose-B		GlOptiPoS$_\Pi \pm 8$		GlOptiPoS$_\Pi$+	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ape	8.94	16.93	**0.86**	**1.04**	4.7	2.86	**0.29**	**0.33**
Cat	3.12	**4.74**	**3.07**	9.44	37.54	25.67	**0.18**	**0.15**
Jack	7.21	12.83	6.57	22.55	**0.37**	**0.15**	**0.38**	**0.44**
Squirrel	9.67	25.22	10.32	15.46	**0.75**	**0.43**	**0.5**	**0.55**
Stitch	10.09	23.98	10	20.24	**1.13**	**0.39**	**0.21**	**0.2**
Vampire queen	14.62	25.45	4.82	10.99	**1.96**	**0.57**	**0.55**	**0.56**
3D Touch	14.75	27.09	8.14	14.38	**1.6**	**0.85**	**3.31**	**13.36**
Auto GPS	24.93	35.16	25.73	34.36	**3.28**	**1.74**	**0.4**	**0.58**
Bracket	15	26.46	**0.63**	**1.03**	**0.76**	**0.47**	6.06	17.84
Deadpool	9.28	18.2	**1.08**	**3.6**	**1.4**	**0.65**	1.44	5.13
Driller	33.64	38.54	15.5	25.23	**0.3**	**0.04**	**0.13**	**0.11**
Flashlight	32.87	40.72	22.88	34.23	**3.89**	**1.53**	**0.31**	**0.32**
Lamp clamp	27.11	33.53	**23.62**	**25.93**	23.94	**16.22**	**19.18**	29.86
Lego	47.21	35.36	**17.06**	**24.89**	60.26	41.75	**3.25**	**13.67**
RJ45	**15.6**	25.66	17.44	35.28	45.76	**15.27**	**2.42**	**9.58**
RTI Arm	48.17	35.63	29.89	33.35	**0.22**	**0.04**	**21.65**	**29.94**
Standtube	20.09	34.56	**16.42**	**28.48**	45.78	31.31	**0.29**	**0.28**
Teapot	22.65	27.2	**11.91**	**25.58**	59.19	36.22	**6.02**	**17.34**
Tube	13.43	23.03	14.23	24.13	**2.4**	**0.62**	**0.29**	**0.2**
Wall Shelf	25.43	37.86	17.47	25.08	**0.31**	**0.03**	**0.16**	**0.13**

Table 3. OE(↓) in degrees, following same scheme as Tab. 2

	STI-Pose-A		STI-Pose-B		GlOptiPoS$_\Pi \pm 8$		GlOptiPoS$_\Pi$+	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ape	25.39	21.67	**9.68**	1.41	11.94	**0.65**	**0.25**	**0.22**
Cat	29.28	18.71	**17.72**	**4.57**	26.67	6.04	**0.16**	**0.13**
Jack	26.58	27.6	5.81	1.77	**4.17**	**0.15**	**0.34**	**0.52**
Squirrel	25.62	21.39	**14.72**	**8.22**	18.98	8.56	**0.23**	**0.24**
Stitch	23.34	25.36	7.26	9.49	**2.87**	**0.06**	**0.03**	**0.02**
Vampire queen	38.64	20.87	**24.41**	**0.88**	48.31	4.28	**0.59**	**0.59**
3D Touch	36.45	19.41	27.51	11.56	**8.93**	**0.12**	**0.39**	0.89
Auto GPS	64.29	20.3	64.39	22.01	**12.31**	**0.74**	**0.58**	**0.78**
Bracket	36.28	40.46	**14.87**	**1.59**	16.28	10.21	**2.24**	**6.65**
Deadpool	31.68	26.02	12.68	**1.21**	**7.52**	**1.55**	**0.81**	2.57
Driller	32.47	24.58	21.85	12.96	**8.08**	**1.28**	**0.21**	**0.21**
Flashlight	34.63	22.59	22.09	8.5	**11.46**	**3.74**	**0.14**	**0.12**
Lamp clamp	29.71	23.51	**11.87**	**2.32**	30.34	5.34	**0.45**	**0.28**
Lego	33.42	28.99	8.28	5.28	**0.23**	**0.14**	**0.02**	**0.02**
RJ45	28.98	28.04	**13.13**	6.04	16.09	**4.23**	**0.72**	**2.13**
RTI Arm	37.98	43.27	4.8	2.63	**2.24**	**0.04**	**0.22**	**0.69**
Standtube	27.05	29.22	**13.51**	13.28	15.31	**0.23**	**0.21**	**0.14**
Teapot	24.38	23.73	**6.02**	**1.2**	18.83	8.1	**0.91**	**1.74**
Tube	32.38	28.02	17.04	11.22	**10.06**	**3.99**	**0.19**	**0.15**
Wall Shelf	34.64	28.77	20.44	18.25	**9.38**	**3.46**	**0.16**	**0.16**

Table 4. TE(↓) in *mm*, following same scheme as Tab. 2

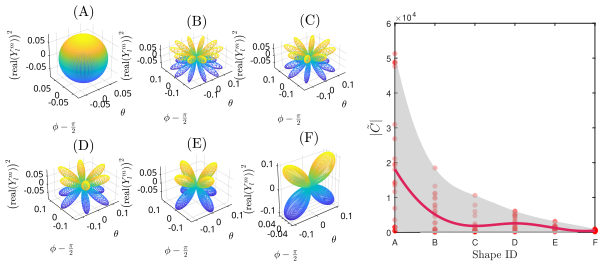
7 (supplementary).

Comparison. The accuracy of GlOptiPoS$_\Pi$+ is very high for asymmetric shapes, dropping for symmetric shapes owing to to multiple ambiguous solutions – consistent with STI-Pose-A, STI-Pose-B and GlOptiPoS$_\Pi$ $\pm$ 8. GlOptiPoS$_\Pi$+ is the best method across all shapes while the second position is split between STI-Pose-B and GlOptiPoS$_\Pi$ $\pm$ 8 for RMSE and TE. For OE, however, GlOptiPoS$_\Pi \pm 8$ is ahead of STI-Pose-B, confirming its advantage for orientation estimation. STI-Pose-A lags behind significantly on all metrics, hinting at higher sensitivity of STI-Pose to inaccurate depth bounds.

Effect of depth prior on accuracy. We vary the deviation of pre-computed and pose-estimation depth from 0 to 70 *mm* in 10 steps and record $H(\mathbf{G}_\Pi, \mathbf{G}^*_\Pi)$ along with RMSE, OE, and TE for an asymmetric and symmetric shape from B*c*OT: *Ape* and *AutoGPS* (resp.) – plots in section 7 (supplementary). We observe: I) *AutoGPS* is less accurate than *Ape* as expected, and II) for *Ape*, increasing deviation of pre-computed and pose-estimation depth results in outlying errors which could be large, however, the median accuracy remains low.

Partial comparisons - non-uniform baselines. For completeness, brief illustrative comparisons with Deep Active Contours (DAC) [35] and Perspective-1-Ellipsoid (P$_1$E) [13] are included in section 7 of supplementary, acknowledging their distinctly different and incomparable problem assumptions.

Additional analysis. We offer examples of ambiguities induced by symmetric shapes, variance of Hausdorff distance and A*o*S with point cloud sampling density and silhouette noise, and experiment on thin shell objects from the Bramante39M [3, 29] dataset in section 7 (supplementary).



(I) (II)

Figure 6. I) The six shapes *A* through *F* using spherical harmonics, and II) $|\tilde{C}|$ plotted against shapes *A* through *F*, the red curve passes through mean values, the shaded area gives the range

5. Discussion

Despite our method's accuracy, strong occlusion or heavy noise can still induce failures. Such settings are fundamentally unsolvable, no existing silhouette-only approach is reliable under such data corruption. Our contribution targets the regime where silhouettes alone remain informative, and within this scope, our method delivers consistently strong performance. Section 8 (supplementary) collects *additional clarifications and responses to typical questions*.

6. Conclusion

We introduce the first globally optimal solution to the P*f*S problem, delivering high accuracy and broad applicability across domains such as robotics, medical imaging, and AR. While extreme occlusion and high noise remain unsolvable theoretical bottlenecks, within the practically relevant regime of unoccluded silhouettes, our approach decisively advances the field, establishing benchmarks that surpass all prior methods.

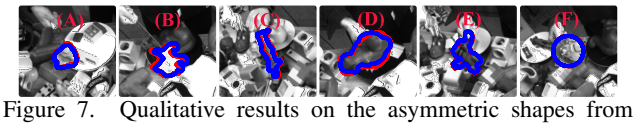


Figure 7. Qualitative results on the asymmetric shapes from *complex_movable_handheld* sequence of B*c*OT dataset: *Ape*, *Cat*, *Jack*, *Squirrel*, *Stitch*, and *Vampire queen* shown in (A) through (F); red curves are G*t* silhouettes, blue curves are from GlOptiPoS$_\Pi$+ (please zoom-in for details)

Acknowledgements

This research was conducted in the research campus MODAL, funded by the Federal Ministry of Research, Technology and Space (BMFTR), Germany, grant no. 3FO18501. This research is also connected with the Competence Center for Excellent Technologies (COMET , grant no. 911654), administered by the Austrian Research Promotion Agency (FFG).

References

[1] Wikipedia: azimuthal equidistant projection, Dec 2024. URL https://en.wikipedia.org/wiki/Azimuthal_equidistant_projection#cite_note-4. 3

[2] Georgios Albanis, Nikolaos Zioulis, Anastasios Dimou, Dimitrios Zarpalas, and Petros Daras. Dronepose: photorealistic uav-assistant dataset synthesis for 3d pose estimation via a smooth silhouette loss. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 663–681. Springer, 2020. 1

[3] Adrien Bartoli and Agniva Sengupta. Camera pose in sft and nrsfm under isometric and weaker deformation models. *Computer Vision and Image Understanding*, page 104488, 2025. 8

[4] Sebahattin Bektas. Least squares fitting of ellipsoid using orthogonal distances. *Boletim de ciências geodésicas*, 21(2): 329–339, 2015. 5

[5] Aleksei Bochkovskii, Amaël Delaunoy, Hugo Germain, Marcel Santos, Yichao Zhou, Stephan R. Richter, and Vladlen Koltun. Depth pro: Sharp monocular metric depth in less than a second. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.02073. 5

[6] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15(42):1455–1459, 2014. URL https://www.manopt.org. 5

[7] Jixiang Chen, Jing Chen, Kai Liu, Haochen Chang, Shanfeng Fu, and Jian Yang. Robust 6dof pose tracking considering contour and interior correspondence uncertainty for ar assembly guidance. *arXiv preprint arXiv:2502.11971*, 2025. 1, 2

[8] Toby Collins, Daniel Pizarro, Simone Gasparini, Nicolas Bourdel, Pauline Chauvet, Michel Canis, Lilian Calvet, and Adrien Bartoli. Augmented reality guided laparoscopic surgery of the uterus. *IEEE Transactions on Medical Imaging*, 40(1):371–380, 2020. 1

[9] Massimiliano Corsini, Paolo Cignoni, and Roberto Scopigno. Efficient and flexible sampling with blue noise properties of triangular meshes. *IEEE transactions on visualization and computer graphics*, 18(6):914–924, 2012. 2

[10] Xiao Cui, Nan Li, Chi Zhang, Qian Zhang, Wei Feng, and Liang Wan. Silhouette-based 6d object pose estimation. In *International Conference on Computational Visual Media*, pages 157–179. Springer, 2024. 2, 5, 6

[11] Herbert Edelsbrunner, David Kirkpatrick, and Raimund Seidel. On the shape of a set of points in the plane. *IEEE Transactions on information theory*, 29(4):551–559, 1983. 2, 3

[12] Jiří Filip, Radek Holub, Vlastimil Havran, Jaroslav Křivánek, and Daniel Sýkora. 3d model of a dragon released during eurographics 2007, 2007. URL www.dcgi.fel.cvut.cz/eg07/index.php?page=dragon. 4

[13] Vincent Gaudillière, Gilles Simon, and Marie-Odile Berger. Perspective-1-ellipsoid: Formulation, analysis and solutions of the camera pose estimation problem from one ellipse-ellipsoid correspondence. *International Journal of Computer Vision*, 131(9):2446–2470, 2023. 2, 8

[14] Robert Gilmore. *Lie groups, Lie algebras, and some of their applications*. Courier Corporation, 2006. 6

[15] Anna Gummeson and Magnus Oskarsson. Relative pose from cylinder silhouettes. In *Proceedings of the Asian Conference on Computer Vision*, pages 2545–2561, 2024. 2

[16] Wulong Guo, Weiduo Hu, Chang Liu, and Tingting Lu. Pose initialization of uncooperative spacecraft by template matching with sparse point cloud. *Journal of Guidance, Control, and Dynamics*, 44(9):1707–1720, 2021. 1

[17] Richard I Hartley and Fredrik Kahl. Global optimization through rotation space search. *International Journal of Computer Vision*, 82(1):64–79, 2009. 1, 3

[18] Paul Hebert, Nicolas Hudson, Jeremy Ma, Thomas Howard, Thomas Fuchs, Max Bajracharya, and Joel Burdick. Combined shape, appearance and silhouette for simultaneous manipulator and object tracking. In *2012 IEEE International Conference on Robotics and Automation*, pages 2405–2412. IEEE, 2012. 1

[19] B KP Horn, H M Hilden, and S Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *JOSA A*, 5(7):1127–1135, 1988. 6

[20] N R Howe. Silhouette lookup for automatic pose tracking. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 15–22. IEEE, 2004. 2

[21] Nicholas R Howe. Silhouette lookup for monocular 3d pose tracking. *Image and Vision Computing*, 25(3):331–341, 2007. 2

[22] Chen Kong, Chen-Hsuan Lin, and Simon Lucey. Using locally corresponding cad models for dense 3d reconstructions from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4857–4865, 2017. 2

[23] Svetlana Lazebnik, Amit Sethi, Cordelia Schmid, David Kriegman, Jean Ponce, and Martial Hebert. On pencils of tangent planes and the recognition of smooth 3d shapes from silhouettes. In *European Conference on Computer Vision*, pages 651–665. Springer, 2002. 2

[24] Jiachen Li, Bin Wang, Shiqiang Zhu, Xin Cao, Fan Zhong, Wenxuan Chen, Te Li, Jason Gu, and Xueying Qin. Bcot: A markerless high-precision 3d object tracking benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6697–6706, 2022. 6

[25] JF Menudet, JM Becker, T Fournel, and C Mennessier. Model-based shape from silhouette: A solution involving a

small number of views. In *Proceedings of the Second International Conference on Computer Vision Theory and Applications*, pages 379–386, 2007. 2

[26] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis: proceedings of the biennial Conference held at Dundee, June 28–July 1, 1977*, pages 105–116. Springer, 2006. 6

[27] Javier Pérez Soler, Jose-Luis Guardiola, Alberto Perez Jimenez, Pau Garrigues Carbó, Nicolás García Sastre, and Juan-Carlos Perez-Cortes. Optimal coherent point selection for 3d quality inspection from silhouette-based reconstructions. *Mathematics*, 11(21):4419, 2023. 1

[28] Mukta Prasad, Andrew W Fitzgibbon, and Andrew Zisserman. Fast and controllable 3d modelling from silhouettes. In *Eurographics (Short Presentations)*, pages 9–12, 2005. 2

[29] Agniva Sengupta and Adrien Bartoli. Convex solutions to sft and nrsfm under algebraic deformation models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 8

[30] Agniva Sengupta and Stefan Zachow. Shape-from-template with generalised camera. *Image and Vision Computing*, page 105579, 2025. 1, 2

[31] Inc. The Mathworks. *MATLAB Function Reference*. The Mathworks, Inc., 2024. 4

[32] Eno Töppe, Martin R Oswald, Daniel Cremers, and Carsten Rother. Silhouette-based variational methods for single view reconstruction. In *Video Processing and Computational Video: International Seminar, Dagstuhl Castle, Germany, October 10-15, 2010. Revised Papers*, pages 104–123. Springer, 2011. 2

[33] Zsolt Ugray, Leon Lasdon, John Plummer, Fred Glover, James Kelly, and Rafael Martí. Scatter search and local nlp solvers: A multistart framework for global optimization. *INFORMS Journal on computing*, 19(3):328–340, 2007. 6

[34] B Vijayakumar, David Kriegman, and Jean Ponce. Invariant-based recognition of complex curved 3d objects from image contours. *Computer Vision and Image Understanding*, 72(3): 287–303, 1998. 2

[35] Long Wang, Shen Yan, Jianan Zhen, Yu Liu, Maojun Zhang, Guofeng Zhang, and Xiaowei Zhou. Deep active contours for real-time 6-dof object tracking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14034–14044, 2023. 2, 8

[36] Rui Wang, Nan Yang, Joerg Stueckler, and Daniel Cremers. Directshape: Direct photometric alignment of shape priors for visual vehicle pose and shape estimation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11067–11073. IEEE, 2020. 1

[37] Wikipedia. Spherical harmonics — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Spherical%20harmonics&oldid=1321561306, 2025. [Online; accessed 13-November-2025]. 7

[38] Ming Zhang, Yinqiang Zheng, and Yuncai Liu. Using silhouette for pose estimation of object with surface of revolution. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 333–336. IEEE, 2009. 2

[39] Zhuo Zhang, WANG Qiufu, BI Daoming, SUN Xiaoliang, and YU Qifeng. Mc-lrf based pose measurement system for shipborne aircraft automatic landing. *Chinese Journal of Aeronautics*, 36(8):298–312, 2023. 1