

Estimation de pose de caméra à partir de correspondances ellipse - ellipsoïde

Vincent Gaudillière¹

Gilles Simon¹

Marie-Odile Berger¹

¹ Inria Nancy - Grand Est / Loria, Nancy, France

vincent.gaudilliere@inria.fr

Résumé

En vision par ordinateur, l'estimation de pose de caméra à partir de correspondances entre des entités géométriques 3D et leurs projections dans l'image est un problème largement étudié. Alors que la plupart des méthodes de l'état de l'art exploitent des primitives simples comme des points ou des droites, et nécessitent donc des modèles de scène relativement denses, les performances atteintes dernièrement par les détecteurs d'objets à base de réseaux de neurones convolutifs permettent d'envisager l'utilisation de modèles 3D beaucoup plus légers, composés seulement de quelques éléments sémantiquement pertinents. Dans ce contexte, nous proposons une nouvelle approche basée modèle pour l'estimation de pose de caméra, dans laquelle la scène est représentée par un ensemble d'ellipsoïdes. Nous montrons qu'il est possible d'optimiser uniquement les trois paramètres d'orientation de la caméra pour résoudre le problème, et que deux correspondances entre ellipses 2D et ellipsoïdes 3D sont nécessaires en pratique.

Mots Clef

Estimation de pose, quadriques, coniques.

Abstract

In computer vision, estimating camera pose from correspondences between 3D geometric entities and their projections into the image is a widely investigated problem. Although most state-of-the-art methods exploit simple primitives such as points or lines, and thus require dense scene models, the appearance of very effective CNN-based object detectors in the recent years pave the way to the use of much lighter 3D models composed solely of a few semantically relevant features. In that context, we propose a novel model-based camera pose estimation method in which the scene is represented by a set of ellipsoids. We show that the problem can be solved by optimizing only the three camera orientation parameters, and that two correspondences between 2D ellipses and 3D ellipsoids are necessary in practice.

Keywords

Pose estimation, quadrics, conics.

1 Introduction

Estimer la position et l'orientation d'une caméra par rapport à son environnement est une tâche fondamentale en vision par ordinateur. Dans ce problème, il est nécessaire de construire et de conserver une représentation de l'environnement dans lequel évolue l'observateur. En pratique, cette connaissance se présente le plus souvent sous la forme d'un modèle tridimensionnel, par rapport auquel la caméra est positionnée [1].

Lorsque la scène est modélisée par un nuage de points 3D, il est possible de calculer la pose de la caméra à partir de trois correspondances entre points de l'image et points du modèle (problème P3P) [2]. Pour obtenir une plus grande précision, la plupart des méthodes considèrent un nombre arbitraire $n > 3$ de correspondances 2D-3D (PnP). Ces méthodes consistent soit à encapsuler un algorithme P3P au sein d'un processus itératif de rejet des correspondances aberrantes (RANSAC [3]), soit à estimer directement la pose à partir des n correspondances (DLT [4],[5]). Enfin, certaines méthodes minimisent l'erreur de reprojection à partir d'une solution initiale ([4], page 181). Des méthodes équivalentes existent pour des modèles constitués de droites (PnL), ou d'un mélange de points et de droites [6]. Dans tous les cas, la construction du modèle est coûteuse en temps de calcul (très grand nombre d'indices locaux extraits des images avant de les mettre en correspondance, puis d'estimer leurs positions en 3D), et sensible à certaines caractéristiques telles que le manque de texture, la présence de motifs répétés, ou encore de spécularités. De plus, la construction d'un modèle fiable nécessite souvent un grand nombre d'images présentant des recouvrements importants. Dans le cas où l'observateur évolue dans un environnement inconnu, d'autres méthodes permettent de construire le modèle et d'estimer la pose simultanément (SLAM : Simultaneous Localization and Mapping), mais ces méthodes reposent principalement sur les mêmes principes que les méthodes PnP [7], et présentent donc les mêmes limitations.

Parallèlement, ces dernières années ont vu une explosion des performances de détection automatique d'objets dans les images, portée par des méthodes à base de réseaux de neurones convolutifs telles que R-CNN [8, 9, 10], SSD [11], ou encore YOLO [12, 13, 14]. Ce saut qualitatif a

permis l'émergence de nouvelles approches dans la résolution des problèmes classiques de vision par ordinateur. En particulier, l'extraction et la mise en correspondance de primitives haut-niveau (objets), en lieu et place des traditionnelles primitives bas-niveau (points d'intérêt visuels, segments de droites), est déjà à la base de plusieurs méthodes de la littérature. La modélisation des projections d'objets par des ellipses virtuelles a permis à Crocco *et al.* de proposer une solution analytique pour la reconstruction par *Structure from Motion* (SfM) de la scène sous forme d'un nuage d'ellipsoïdes [15]. Cette méthode se limite cependant au cas d'une projection orthographique, de même que son extension intégrant des modèles CAO des objets pour une plus grande précision de reconstruction [16]. La projection perspective est prise en compte dans [17], mais cette méthode de SLAM propose de minimiser une erreur géométrique de reprojection des ellipsoïdes en fonction des six degrés de liberté de la caméra, à partir d'une solution initiale fournie par des capteurs odométriques. Or, dans cet article, nous montrons que cette optimisation est sur-paramétrée, puisqu'il est possible d'optimiser uniquement l'orientation de la caméra, pour en déduire ensuite sa position.

Notre algorithme d'estimation de pose nécessite une modélisation de la scène sous forme d'un ensemble d'ellipsoïdes virtuelles, correspondant à des objets d'intérêt. Les projections de ces objets, une fois détectées dans l'image, sont utilisées pour calculer la pose de la caméra. Dans [18], Rubino *et al.* ont proposé une solution analytique pour construire un tel modèle, à partir d'au moins trois caméras perspectives calibrées. Par ailleurs, Wokes et Palmer ont proposé une méthode de calcul de pose d'un objet modélisé par un ellipsoïde de révolution (cas particulier d'ellipsoïde) [19, 20]. Les auteurs montrent que, dans le cas d'une projection perspective, le problème d'estimation de pose d'un ellipsoïde de révolution par rapport à la caméra n'admet que deux solutions distinctes. Pour le cas plus général de l'ellipsoïde, une mise en équation du même problème a été proposée par Eberly [21], sans toutefois qu'une méthode explicite de calcul des solutions ne soit présentée.

Dans cet article, nous reprenons le formalisme introduit par Eberly, et montrons que le problème initial est équivalent à un second problème, dans lequel seule l'orientation de l'ellipsoïde intervient (Section 2). La position de l'ellipsoïde peut alors être déduite de son orientation sans ambiguïté (Section 3.1). Ces considérations théoriques nous permettent de proposer une méthode d'estimation de pose de caméra à partir de correspondances entre ellipses 2D et ellipsoïdes 3D, sous la forme d'un problème d'optimisation dont l'unique paramètre est l'orientation de la caméra. Nous mettons en évidence numériquement le fait que la pose ne peut être déterminée de manière unique à partir d'une seule correspondance (Section 3.2), avant de formuler le problème pour un nombre arbitraire d'ellipsoïdes (Section 3.3). Enfin, une étude de la robustesse de la méthode est proposée (Section 4).

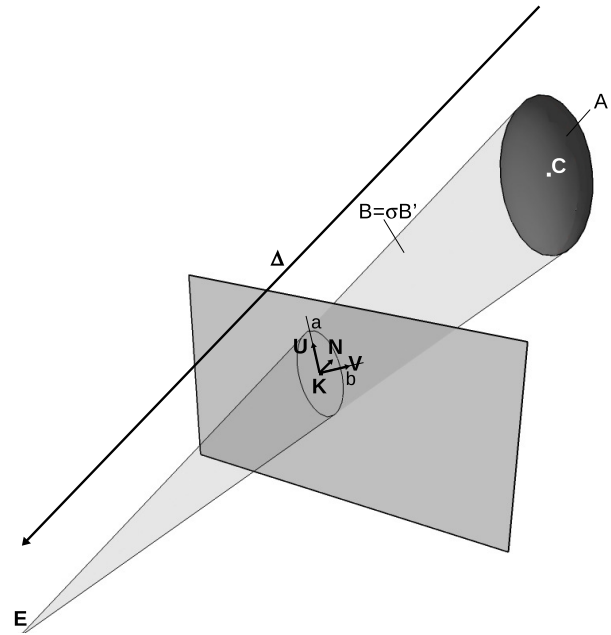


FIGURE 1 – Illustration du plan image, du centre de projection, de l'ellipsoïde, et de son ellipse projetée.

2 Réduction du problème

En premier lieu, nous nous intéressons au problème d'estimation de pose à partir d'une scène composée d'un seul ellipsoïde. Nous considérons en fait le problème équivalent, qui consiste à calculer l'orientation et la position de l'ellipsoïde par rapport à la caméra, et montrons qu'il peut se réduire à l'estimation de l'orientation seulement. Dans le prolongement du travail de Wokes et Palmer sur les ellipsoïdes de révolutions [20], nous proposons donc ici une analyse du cas général (ellipsoïdes quelconques). Toutefois, le formalisme que nous utilisons diffère de celui de la référence, car ce dernier repose sur l'hypothèse que l'ellipsoïde possède un axe de révolution, ce qui est faux dans le cas général.

2.1 Equation d'alignement des cônes

Sauf mention contraire, toutes les grandeurs introduites ci-après sont exprimées dans le repère caméra.

Suivant les notations introduites dans [21] et présentées Figure 1, on considère un ellipsoïde de centre C défini par l'équation

$$(\mathbf{X} - \mathbf{C})^\top A (\mathbf{X} - \mathbf{C}) = 1$$

dans laquelle A est une matrice réelle symétrique définie positive caractérisant l'orientation et la taille de l'ellipsoïde, et \mathbf{X} un point de l'espace appartenant à l'ellipsoïde. Etant donné un centre de projection E et un plan de projection de normale \mathbf{N} auquel E n'appartient pas, la projection de l'ellipsoïde sur ce plan est une ellipse de centre K et de demi-axes de longueurs a et b . Les directions principales de l'ellipse sont représentées par les vecteurs unitaires \mathbf{U} et \mathbf{V} , de telle sorte que $\{\mathbf{U}, \mathbf{V}, \mathbf{N}\}$ forment une

base orthonormale de l'espace.

Expression du cône de projection. On définit le cône de projection comme étant le cône de sommet \mathbf{E} tangent à l'ellipsoïde. Toujours d'après [21], en notant $\mathbf{\Delta} = \mathbf{E} - \mathbf{C}$ le vecteur reliant le centre de l'ellipsoïde au centre de projection, le cône de projection est défini par la matrice

$$B \stackrel{def}{=} A\mathbf{\Delta}\mathbf{\Delta}^\top A - (\mathbf{\Delta}^\top A\mathbf{\Delta} - 1)A$$

de sorte que les points \mathbf{X} du cône de projection sont ceux qui satisfont l'équation $(\mathbf{X} - \mathbf{E})^\top B(\mathbf{X} - \mathbf{E}) = 0$. On notera que B est une matrice symétrique réelle, inversible et de signature (1,2) ou (2,1), c'est-à-dire qu'une de ses valeurs propres est du signe opposé à celui des deux autres.

Expression du cône de rétroprojection. On définit le cône de rétroprojection comme le cône généré par l'ensemble des droites passant par \mathbf{E} et par un point de l'ellipse. Eberly montre que ce cône est caractérisé par une matrice B' , définie de la manière suivante

$$B' \stackrel{def}{=} P^\top M P - Q$$

avec

$$\begin{aligned} M &= \mathbf{U}\mathbf{U}^\top/a^2 + \mathbf{V}\mathbf{V}^\top/b^2 \\ \mathbf{W} &= \mathbf{N}/(\mathbf{N} \cdot (\mathbf{K} - \mathbf{E})) \\ P &= I - (\mathbf{K} - \mathbf{E})\mathbf{W}^\top \\ Q &= \mathbf{W}\mathbf{W}^\top \end{aligned}$$

Là encore, les points \mathbf{X} du cône de rétroprojection sont ceux qui vérifient l'équation $(\mathbf{X} - \mathbf{E})^\top B'(\mathbf{X} - \mathbf{E}) = 0$, et la matrice B' possède les mêmes propriétés que B .

Mise en équation du problème. Dès lors, étant donné un ellipsoïde, une projection centrale (centre et plan de projection), ainsi qu'une ellipse inscrite dans le plan de projection, l'ellipse est l'image projetée de l'ellipsoïde si et seulement si les cônes de projection et de rétroprojection sont confondus [21], *i.e.* si et seulement s'il existe un scalaire non nul σ tel que $B = \sigma B'$, soit

$$A\mathbf{\Delta}\mathbf{\Delta}^\top A - (\mathbf{\Delta}^\top A\mathbf{\Delta} - 1)A = \sigma B' \quad (1)$$

De plus, on peut définir un repère dans lequel la forme quadratique associée à l'ellipsoïde a pour représentation matricielle une matrice diagonale D (à termes diagonaux strictement positifs). En notant R la rotation entre ce repère et le repère image, la matrice A s'écrit alors $A = R^\top D R$. Ainsi, résoudre le problème d'estimation de pose d'un ellipsoïde à partir de son ellipse projetée revient à résoudre l'équation suivante, d'inconnues $(R, \mathbf{\Delta}, \sigma)$:

$$DR\mathbf{\Delta}\mathbf{\Delta}^\top R^\top D - (\mathbf{\Delta}^\top R^\top D R \mathbf{\Delta} - 1)D = \sigma R B' R^\top$$

Dans cette formulation, R caractérise l'orientation de l'ellipsoïde, $\mathbf{\Delta}$ sa position, et σ est un facteur d'échelle.

2.2 Théorème d'équivalence

Le problème d'estimation de pose d'un ellipsoïde fait donc intervenir trois variables. Dans cette partie, nous montrons que $\mathbf{\Delta}$ et σ sont en fait des variables secondaires. Pour cela, nous montrons que le problème formalisé dans la section 2.1 est équivalent à un second problème, dans lequel l'unique inconnue est la matrice R .

Par abus de langage, on dira dans la suite qu'une matrice A est une génératrice du cône B' à une translation près lorsqu'il existe un vecteur $\mathbf{\Delta}$ et un scalaire non nul σ tels que le triplet $(A, \mathbf{\Delta}, \sigma)$ est solution de l'équation (1).

Théorème 1. *A est une génératrice de B' à une translation près si et seulement si le discriminant du polynôme caractéristique généralisé du couple $\{A, B'\}$ est nul.*

Démonstration. **Supposons qu'il existe un vecteur $\mathbf{\Delta}$ et un scalaire non nul σ tels que le triplet $(A, \mathbf{\Delta}, \sigma)$ soit solution de l'équation 1.** Alors, en multipliant l'équation (1) à droite par $\mathbf{\Delta}$ (voir calcul en Annexe 1), on obtient

$$A\mathbf{\Delta} = \sigma B' \mathbf{\Delta} \quad (2)$$

A partir de là, trouver les paires $(\sigma, \mathbf{\Delta})$ qui satisfont l'équation 2 revient à résoudre un problème aux valeurs propres généralisé [22]. De telles valeurs de σ sont appelées des *valeurs propres généralisées* du couple $\{A, B'\}$, et les vecteurs $\mathbf{\Delta}$ correspondant leurs *vecteurs propres généralisés*. En particulier, les valeurs propres généralisées du couple $\{A, B'\}$ sont les racines du polynôme caractéristique $P_{\{A, B'\}}(x) = \det(A - xB')$. Or, comme la matrice B' est inversible, on remarque aisément que les éléments propres généralisés du couple $\{A, B'\}$ sont les mêmes que les éléments propres de la matrice $B'^{-1}A$. On peut ensuite remarquer que le polynôme $Q(x) = \mu x^2 - (\mu + 1)\sigma x + \sigma^2$, avec $\mu = 1 - \mathbf{\Delta}^\top A \mathbf{\Delta}$, est un polynôme annulateur de $B'^{-1}A$ (voir Annexe 2). Comme Q est de degré 2, on en déduit que la matrice $B'^{-1}A$, et a fortiori le couple $\{A, B'\}$, ne possèdent qu'au plus deux valeurs propres distinctes. Ainsi, on en déduit que le polynôme $P_{\{A, B'\}}$, qui ne possède donc qu'au plus deux racines distinctes, a son discriminant égal à 0 (voir Annexe 3).

Supposons maintenant que le discriminant du polynôme caractéristique du couple $\{A, B'\}$ soit égal à 0. On peut tout d'abord remarquer que, A étant symétrique définie positive et B' étant symétrique, le couple $\{A, B'\}$ possède les propriétés suivantes [22] :

1. Les valeurs propres généralisées du couple sont réelles,
2. Chaque sous-espace propre généralisé est de dimension égale à la multiplicité de la valeur propre associée,
3. Les vecteurs propres généralisés forment une base de \mathbb{R}^3 , et ceux associés à des valeurs propres différentes sont A -orthogonaux.

Comme le discriminant de $P_{\{A, B'\}}$ est égal à 0, le couple $\{A, B'\}$ ne possède qu'au plus deux valeurs propres distinctes. De plus, si le couple ne possédait qu'une seule valeur propre de multiplicité 3 appelée σ_0 , alors d'après la propriété 2 ci-dessus, on aurait $\dim(\text{Ker}(A - \sigma_0 B')) = 3$, c'est-à-dire $A = \sigma_0 B'$, ce qui est impossible car A représente un ellipsoïde tandis que B' représente un cône. Donc **le couple possède exactement deux valeurs propres généralisées.**

Appelons ensuite σ_1 (multiplicité 1) et σ_2 (multiplicité 2) ces deux valeurs propres. En remarquant que $\frac{1}{\sigma_1}$ et $\frac{1}{\sigma_2}$ sont les valeurs propres généralisées du couple $\{B', A\}$, on peut écrire, d'après [23] (Théorème 3)

$$\forall \mathbf{X} \in \mathbb{R}^3 \setminus \{0\}, \min\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}\right) \leq \frac{\mathbf{X}^\top B' \mathbf{X}}{\mathbf{X}^\top A \mathbf{X}} \leq \max\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}\right)$$

Si σ_1 et σ_2 étaient de même signe, alors $\forall \mathbf{X} \in \mathbb{R}^3 \setminus \{0\}$, $\mathbf{X}^\top B' \mathbf{X}$ serait de ce signe (car $\mathbf{X}^\top A \mathbf{X} > 0$). Or B' n'est ni définie positive ni définie négative (cône), donc c'est impossible. Ainsi, on en déduit que **les deux valeurs propres généralisées sont de signes opposés.**

Appelons $\Delta, \mathbf{U}, \mathbf{V}$ trois vecteurs propres associés à σ_1 et σ_2 . On a alors, par définition

$$\begin{aligned} A\Delta &= \sigma_1 B' \Delta \\ A\mathbf{U} &= \sigma_2 B' \mathbf{U} \\ A\mathbf{V} &= \sigma_2 B' \mathbf{V} \end{aligned}$$

En appelant $B_A(\mathbf{X})$ la matrice $A\mathbf{X}\mathbf{X}^\top A - (\mathbf{X}^\top A \mathbf{X} - 1)A$, on peut remarquer que $\forall \mathbf{X} \in \mathbb{R}^3$, on a

$$\begin{aligned} B_A(\mathbf{X})\mathbf{X} &= A\mathbf{X}(\mathbf{X}^\top A \mathbf{X}) - \mathbf{X}^\top A \mathbf{X} A \mathbf{X} + A\mathbf{X} \\ &= (\mathbf{X}^\top A \mathbf{X})A\mathbf{X} - \mathbf{X}^\top A \mathbf{X} A \mathbf{X} + A\mathbf{X} \\ &= A\mathbf{X} \end{aligned}$$

En particulier, $B_A(\Delta)\Delta = A\Delta$, d'où

$$B_A(\Delta)\Delta = \sigma_1 B' \Delta$$

Par ailleurs, si on multiplie cette matrice à droite par \mathbf{U} , sachant que Δ et \mathbf{U} sont A -orthogonaux (cf propriété 3 ci-dessus), on obtient

$$\begin{aligned} B_A(\Delta)\mathbf{U} &= A\Delta\Delta^\top A\mathbf{U} - (\Delta^\top A\Delta - 1)A\mathbf{U} \\ &= A\Delta(\Delta^\top A\mathbf{U}) + (1 - \Delta^\top A\Delta)A\mathbf{U} \\ &= (1 - \Delta^\top A\Delta)A\mathbf{U} \\ &= (1 - \Delta^\top A\Delta)\sigma_2 B' \mathbf{U} \end{aligned}$$

Dès lors, on peut choisir la norme de Δ telle que $(1 - \Delta^\top A\Delta)\sigma_2 = \sigma_1$. En effet, il suffit de choisir $\|\Delta\|$ telle que $\Delta^\top A\Delta = 1 - \frac{\sigma_1}{\sigma_2}$, ce qui est rendu possible par le fait que σ_1 et σ_2 sont de signes opposés, d'où $1 - \frac{\sigma_1}{\sigma_2} > 1 > 0$. En choisissant un tel Δ , on obtient

$$B_A(\Delta)\mathbf{U} = \sigma_1 B' \mathbf{U}$$

et, de la même façon

$$B_A(\Delta)\mathbf{V} = \sigma_1 B' \mathbf{V}$$

Comme $(\Delta, \mathbf{U}, \mathbf{V})$ forment une base de \mathbb{R}^3 , on a finalement

$$B_A(\Delta) \stackrel{def}{=} A\Delta\Delta^\top A - (\Delta^\top A\Delta - 1)A = \sigma_1 B'$$

Ainsi, A fait bien partie d'un triplet de solutions (A, Δ, σ_1) de l'équation (1). \square

En conclusion, on peut réduire l'estimation de pose (orientation et position) d'un ellipsoïde à la seule estimation de son orientation. Les orientations R solutions sont celles qui annulent le discriminant du polynôme caractéristique généralisé du couple $\{A = R^\top D R, B'\}$.

3 Méthode d'estimation de la pose

Dans la section précédente, nous avons vu qu'estimer la pose d'un ellipsoïde dans le repère caméra - et donc, de manière équivalente, estimer la pose de la caméra à partir d'un ellipsoïde - revient à trouver une orientation R de l'ellipsoïde dans le repère caméra telle que le discriminant du couple $\{A = R^\top D R, B'\}$ soit nul. Dans cette partie, nous montrons tout d'abord que la position peut ensuite être déterminée à partir de l'orientation sans ambiguïté (section 3.1), puis nous mettons en évidence numériquement le fait qu'il existe un continuum de solutions pour R (section 3.2). Nous formulons finalement le problème d'estimation de pose de la caméra dans le cas d'une scène composée d'au moins deux ellipsoïdes (section 3.3).

3.1 Position connaissant l'orientation

Le Théorème 1 a montré que Δ et σ sont des variables secondaires de l'équation (1). Dans cette partie, nous montrons de plus qu'elles peuvent être déterminées de manière unique à partir de A et B' .

Corollaire 1. *Si A est une génératrice de B' à une translation près, alors (i) le couple $\{A, B'\}$ possède exactement deux valeurs propres généralisées de signes opposés, (ii) σ est la valeur propre de multiplicité 1, et (iii) Δ est unique et est un vecteur propre généralisé associé à σ .*

Démonstration. Nous avons vu que le problème aux valeurs propres généralisé (équation 2) possède deux solutions. Appelons (σ_1, Δ_1) et (σ_2, Δ_2) ces solutions, de telle sorte que σ_i soit la valeur propre de multiplicité i , et que $\|\Delta_i\| = 1$. Supposons maintenant qu'il existe $k \in \mathbb{R}^*$ tel que $(A, \sigma_2, k\Delta_2)$ soit solution de l'équation (1). On a alors

$$A - \sigma_2 B' = M A$$

avec $M = k^2(\Delta_2^\top A \Delta_2 I - A \Delta_2 \Delta_2^\top)$, où I est la matrice identité. Vu la propriété 2 de la preuve du Théorème 1, on a $\dim(\text{Ker}(A - \sigma_2 B')) = 2$, d'où, puisque A est inversible, $\dim(\text{Ker}(M)) = 2$. Or, en définissant

$\Delta_2^\perp = \{X \in \mathbb{R}^3 / X \perp \Delta_2\}$ le sous-espace de dimension 2 orthogonal à Δ_2 , on observe que

$$\begin{aligned} \forall X \in \Delta_2^\perp, MX &= k^2 \Delta_2^\top A \Delta_2 X - k^2 A \Delta_2 \Delta_2^\top X \\ &= k^2 \Delta_2^\top A \Delta_2 X - k^2 A \Delta_2 (\Delta_2 \cdot X) \\ &= k^2 \Delta_2^\top A \Delta_2 X \end{aligned}$$

D'où, comme A est définie positive, $MX \neq 0$ lorsque $X \neq 0$. Ainsi, $\Delta_2^\perp \cap \text{Ker}(M) = \{0\}$, ce qui implique que $\dim(\Delta_2^\perp) + \dim(\text{Ker}(M)) \leq 3$. Or c'est impossible puisque $\dim(\Delta_2^\perp) = \dim(\text{Ker}(M)) = 2$.

Ainsi, les triplets $(A, \sigma_2, k\Delta_2)$ ne peuvent pas être solutions de l'équation (1), donc les solutions possibles s'écrivent nécessairement $(A, \sigma_1, k\Delta_1)$, avec $k \in \mathbb{R}^*$. L'équation (1) s'écrit alors

$$k^2(A\Delta_1\Delta_1^\top A - \Delta_1^\top A\Delta_1 A) = \sigma_1 B' - A$$

Ainsi, les deux matrices exhibées ci-dessus sont proportionnelles, et k^2 est leur coefficient de proportionnalité. Finalement, le seul k possible est celui qui permet au centre de l'ellipsoïde d'être devant la caméra (contrainte de chiralité). \square

Ainsi, σ et Δ peuvent être calculés sans ambiguïté à partir de A . En pratique, le discriminant de $P_{\{A, B'\}}$ n'est pas toujours égal à 0. Dans ce cas, on commence par calculer les rapports entre les valeurs propres généralisées du couple $\{A, B'\}$, puis on détermine les deux valeurs dont le ratio est le plus proche de 1, σ étant alors défini comme la troisième valeur. Ensuite, on définit Δ_0 comme étant le vecteur propre de norme 1 associé à σ , puis on pose $M = \sigma B' - A$ et $N = A\Delta_0\Delta_0^\top A - \Delta_0^\top A\Delta_0 A$. On définit alors K_2 la matrice de terme général $M(i, j)/N(i, j)$, puis $\text{vec}(K_2)$ le vecteur formé par les six éléments de sa partie triangulaire supérieure (K_2 est symétrique par construction). On calcule finalement k^2 par la moyenne $k^2 = \overline{\text{vec}(K_2)}$, et on détermine le signe de k en appliquant la contrainte de chiralité.

3.2 Poses à partir d'un seul ellipsoïde

On définit $\mathcal{R}_w = (O; \mathcal{B}_w)$ le repère monde, et $\mathcal{R}_c = (E; \mathcal{B}_c)$ le repère caméra, avec \mathcal{B}_w et \mathcal{B}_c deux bases orthonormées directes. On appelle A_w la représentation matricielle, dans la base \mathcal{B}_w , de la forme quadratique associée à l'ellipsoïde, et A_c sa représentation dans la base \mathcal{B}_c . De la même façon, on appelle B'_c la représentation matricielle, dans la base \mathcal{B}_c , de la forme quadratique associée au cône de rétroprojection. On appelle, de plus, Δ_w l'expression, dans la base \mathcal{B}_w , du vecteur reliant le centre de l'ellipsoïde au centre de la caméra, et Δ_c son expression dans la base \mathcal{B}_c . On désigne enfin par wR_c la matrice de passage de \mathcal{B}_w à \mathcal{B}_c , grâce à laquelle on peut écrire

$$\begin{cases} A_c = {}^wR_c^\top A_w {}^wR_c \\ \Delta_w = {}^wR_c \Delta_c \end{cases}$$

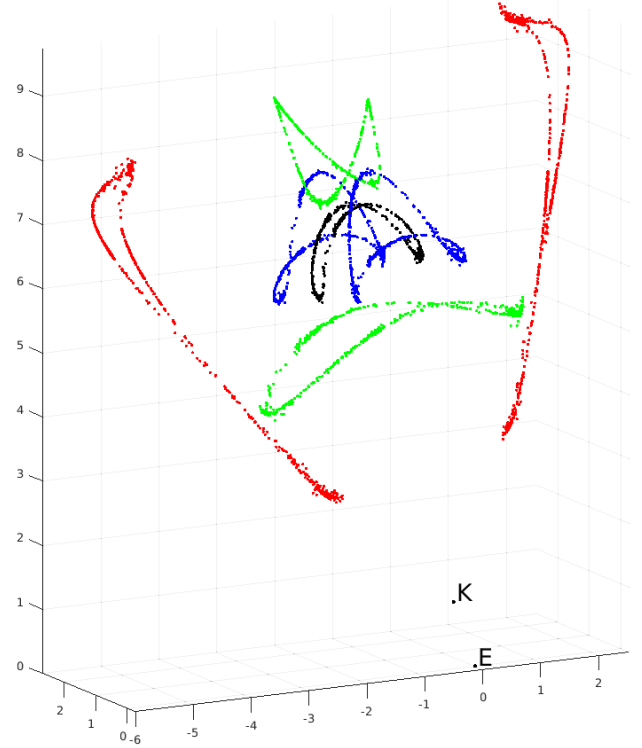


FIGURE 2 – Illustration, dans le repère caméra, des solutions du problème d'estimation de pose à un seul ellipsoïde, obtenues après minimisation de la fonction f_{discr} à partir de 1728 rotations initiales différentes. En noir : centres des ellipsoïdes reconstruits. En rouge, vert, et bleu : sommets des axes principaux des ellipsoïdes.

Supposant A_w connue (modèle), le problème consiste alors à trouver une matrice de rotation wR_c telle que $\text{discriminant}(P_{\{A_c, B'_c\}}) = 0$. On peut donc le voir comme un problème d'optimisation. Pour cela, il est nécessaire de supposer connue une estimée initiale de la rotation entre le repère monde et le repère caméra (par exemple calculée à partir des points de fuite), caractérisée par les angles d'Euler $\Theta_0 = (\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)})$. On définit ensuite la fonction de coût à minimiser :

$$f_{discr}(\Theta) = |\text{discriminant}(P_{\{A_c(\Theta), B'_c\}})|^2$$

Une illustration des solutions du problème est proposée en Figure 2. Pour obtenir ces solutions, nous avons considéré un cas de synthèse, et discrétisé l'espace des angles d'Euler avec un pas de 30 degrés dans chaque direction, afin d'obtenir 1728 triplets de paramètres Θ_0 différents. Nous avons ensuite appliqué un algorithme Levenberg-Marquardt pour minimiser la fonction $f_{discr}(\Theta)$ à partir de cette position Θ_0 . Enfin, pour chaque orientation, la position du centre de l'ellipsoïde dans le repère caméra a été calculée de la manière décrite dans la section 3.1.

Sachant que les valeurs finales de la fonction de coût sont toutes comprises entre 2.4×10^{-13} et 2.9×10^{-2} , on

constate qu'il existe un lieu de solutions pour le problème d'estimation de pose à un seul ellipsoïde.

3.3 Pose à partir de N ellipsoïdes

La figure 2 montre qu'il existe une infinité de solutions au problème d'estimation de pose à un seul ellipsoïde. Nous considérons donc dans la suite un modèle de scène composé de $N \geq 2$ ellipsoïdes. En reprenant la philosophie de notation précédente, on ajoute un exposant (i) aux grandeurs relatives à l'ellipsoïde i , de telle sorte que, pour tout i entre 1 et N , on puisse écrire

$$\begin{cases} A_c^{(i)} = {}^wR_c^\top A_w^{(i)} {}^wR_c \\ \Delta_w^{(i)} = {}^wR_c \Delta_c^{(i)} \end{cases} \quad (3)$$

Là encore, les matrices $A_w^{(i)}$ sont supposées connues (modèle), et la rotation wR_c est caractérisée par les angles d'Euler $\Theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)})$, dont dépendent donc les matrices $A_c^{(i)}$. On définit alors la nouvelle fonction de coût à minimiser de la manière suivante :

$$f'_{discr}(\Theta) = \sum_{i=1}^N |\text{discriminant}(P_{\{A_c^{(i)}(\Theta), B_c^{(i)}\}})|^2$$

Calcul de l'orientation Partant d'une solution initiale Θ_0 (calculée par exemple à l'aide des points de fuite), on applique un algorithme Levenberg-Marquardt à la fonction $f'_{discr}(\Theta)$ pour faire tendre simultanément vers 0 les discriminants des polynômes caractéristiques de tous les couples $\{A_c^{(i)}, B_c^{(i)}\}$. Les paramètres Θ_f obtenus permettent de définir une nouvelle orientation pour la caméra : ${}^wR_c(\Theta_f)$.

Calcul de la position On appelle $C_w^{(i)}$ les coordonnées, supposées connues, des centres des ellipsoïdes dans le repère monde \mathcal{R}_w . A partir de la rotation ${}^wR_c(\Theta_f)$, on calcule les $\Delta_c^{(i)}$ et les $K_2^{(i)}$ de la manière expliquée dans la section 3.1. Puis E_w , position de la caméra dans le repère monde \mathcal{R}_w , est donnée par la moyenne pondérée suivante :

$$E_w = \frac{\sum_{i=1}^N \alpha_i (\Delta_w^{(i)} + C_w^{(i)})}{\sum_{i=1}^N \alpha_i} \quad (4)$$

$$= \frac{\sum_{i=1}^N \alpha_i ({}^wR_c(\Theta_f) \Delta_c^{(i)} + C_w^{(i)})}{\sum_{i=1}^N \alpha_i} \quad (5)$$

avec

$$\alpha_i = \frac{1}{s(\sqrt{\text{vec}(K_2^{(i)})})}$$

$s(\cdot)$ désignant l'écart-type des éléments d'un vecteur.

4 Résultats

Nous avons procédé à l'évaluation de la méthode à partir d'un environnement simulé, et avons comparé les performances d'optimisation de notre fonction de coût (f'_{discr}) avec celles de deux autres erreurs présentes dans l'état de l'art. Plus précisément, la méthode QuadricSLAM [17]

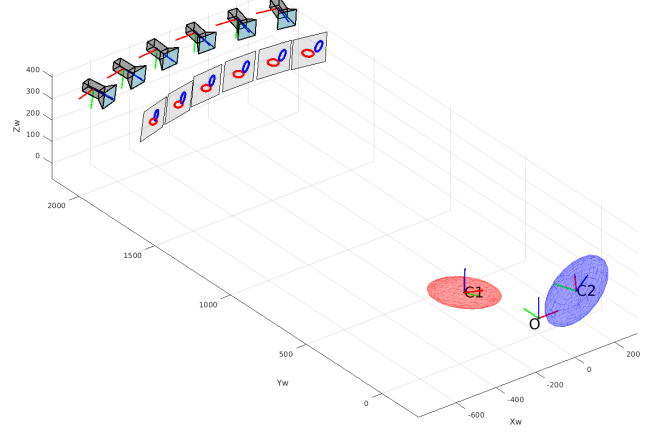


FIGURE 3 – Scène synthétique utilisée pour les tests, composée de deux ellipsoïdes et de six caméras, dont des plans images virtuels contenant les ellipses projetées ont été représentés.

minimise itérativement une erreur géométrique définie comme la distance entre les boîtes englobantes des ellipses détectées et reprojetées, en fonction des six paramètres de pose, tandis que les solutions analytiques présentées dans [15, 18] reposent sur une erreur algébrique définie comme la distance entre les vecteurs formés par les 5 paramètres caractérisant les ellipses. Dans notre évaluation, nous avons utilisé ces deux erreurs, désignées par *QuadricSLAM* et *Err. algebr.* dans la Figure 4, comme fonctions de coût du processus d'optimisation, afin de les comparer avec notre méthode. Il est important de noter que le calcul (et donc la minimisation) de ces deux erreurs ne peut se faire qu'à partir des six paramètres de pose, et donc que ces méthodes nécessitent la connaissance d'un a priori sur la position de la caméra, contrairement à la nôtre qui ne requiert que la donnée d'une orientation approximative, plus facile à obtenir (gyromètre, calcul à partir des points de fuite, etc).

En pratique, nous nous sommes intéressés au cas minimal d'estimation de pose à partir de deux correspondances ellipses - ellipsoïdes. La scène (représentée en Figure 3) est constituée de deux ellipsoïdes de rayons (18, 12, 6 cm) et (20, 10, 8 cm) placés et orientés arbitrairement, et de six caméras disposées approximativement en arc de cercle à une distance moyenne de 2.1 m du barycentre des ellipsoïdes.

Pour les tests, nous avons bruité les ellipses véritables en ajoutant un bruit de loi uniforme compris entre -1 (respectivement -3) et 1 (3) pixels sur les coordonnées de six points pris sur l'ellipse, avant d'interpoler ces nouveaux points par la meilleure ellipse possible (au sens des moindres carrés). Nous avons également fourni une estimée initiale de l'orientation de la caméra, calculée à partir des angles d'Euler de la véritable orientation auxquels nous avons ajouté un bruit de loi uniforme compris entre -10° et $+10^\circ$. De la même façon, l'estimée initiale de la position, nécessaire aux méthodes auxquelles nous nous comparons,

est calculée à l'aide d'une perturbation de loi uniforme et d'amplitude variable : $[-5\%; +5\%]$, $[-10\%; +10\%]$, et $[-30\%; +30\%]$ de la distance entre les caméras et le barycentre des ellipsoïdes.

Les résultats, présentés en Figure 4, montrent que dans le cas où les ellipses reçoivent la plus faible perturbation (1 pixel), notre méthode présente les meilleurs résultats en termes d'erreur absolue sur l'orientation de la caméra et d'erreur relative sur sa position, avec dans le pire cas 1.61° et 2.62% d'erreurs respectivement. La minimisation de l'erreur algébrique ne permet pas toujours de converger (au minimum 10% des expériences conduisent à une pose aberrante), mais permet d'aboutir à des résultats plus précis lorsqu'elle converge, tandis que la minimisation de l'erreur géométrique *QuadricSLAM* est sensible à la précision de l'estimation initiale de la pose. Dans le second cas (3 pixels de perturbation sur les ellipses), les performances de *QuadricSLAM* ne deviennent supérieures aux nôtres qu'à condition de fournir une très bonne estimée initiale de la position (5% de perturbation). Nos erreurs finales sur la pose sont cette fois inférieures à 5.24° et 8.38% . Enfin, les trois méthodes sont équivalentes en termes de temps de calcul (quelques centièmes à quelques dixièmes de secondes pour la minimisation).

Ainsi, notre méthode présente l'avantage important de ne nécessiter qu'une estimée initiale des trois paramètres d'orientation de la caméra, au lieu des six paramètres de pose requis par les autres méthodes de l'état de l'art. De plus, les résultats montrent une bonne convergence de notre algorithme malgré des perturbations initiales volontairement exagérées : jusqu'à 10° par angle d'Euler, alors qu'il est possible d'obtenir une précision plus grande (1 à 2° maximum d'erreur) en pratique, en utilisant les points de fuite présents dans les images.

5 Conclusion

Dans cet article, nous avons présenté une nouvelle méthode d'estimation de pose de caméra capable d'opérer à partir d'un modèle de scène composé d'au minimum deux ellipsoïdes, et ne nécessitant que la donnée d'une estimée initiale de son orientation. Nous avons, pour cela, proposé une formulation optimale du problème, dans laquelle l'orientation est d'abord optimisée, avant d'être utilisée pour calculer la position. Enfin, nous avons démontré la robustesse de notre méthode sur un cas de synthèse, et avons montré son intérêt par rapport aux autres méthodes de l'état de l'art. Au vu de ses caractéristiques, cette méthode pourrait être utilisée pour calculer une pose approximée de caméra à partir d'une orientation initiale même grossière, avant d'optimiser l'erreur algébrique de reprojection des ellipses à partir de cette pose pour en obtenir une plus précise.

Annexe 1 : $A\Delta = \sigma B'\Delta$

Multiplier l'équation (1) à droite par Δ revient à écrire

$$B\Delta = \sigma B'\Delta$$

Or, en remarquant que $\Delta^\top A\Delta$ est un scalaire, et donc commute avec les vecteurs et matrices, on a

$$\begin{aligned} B\Delta &= (A\Delta\Delta^\top A - (\Delta^\top A\Delta - 1)A)\Delta \\ &= A\Delta(\Delta^\top A\Delta) - \Delta^\top A\Delta A\Delta + A\Delta \\ &= (\Delta^\top A\Delta)A\Delta - \Delta^\top A\Delta A\Delta + A\Delta \\ &= A\Delta \end{aligned}$$

D'où $A\Delta = \sigma B'\Delta$.

Annexe 2 : $Q(B'^{-1}A) = 0$

On remarque tout d'abord qu'en injectant (2) dans (1), on obtient :

$$\sigma^2 B'\Delta\Delta^\top B' - (\sigma\Delta^\top B'\Delta - 1)A = \sigma B'$$

d'où on déduit l'expression suivante pour A :

$$A = \frac{\sigma}{1 - \sigma\Delta^\top B'\Delta} (B' - \sigma B'\Delta\Delta^\top B')$$

Ainsi, en notant I la matrice identité, $B'^{-1}A$ s'écrit

$$B'^{-1}A = \frac{\sigma}{1 - \sigma\Delta^\top B'\Delta} (I - \sigma\Delta\Delta^\top B')$$

D'où, en posant $f = \frac{\sigma}{1 - \sigma\Delta^\top B'\Delta}$, et en élevant au carré :

$$\begin{aligned} (B'^{-1}A)^2 &= f^2 (I - \sigma\Delta\Delta^\top B')^2 \\ &= f^2 (I - 2\sigma\Delta\Delta^\top B' + \sigma^2\Delta(\Delta^\top B'\Delta)\Delta^\top B') \\ &= f^2 (I - 2\sigma\Delta\Delta^\top B' + \sigma^2(\Delta^\top B'\Delta)\Delta\Delta^\top B') \\ &= f^2 (I - \sigma(2 - \sigma\Delta^\top B'\Delta)\Delta\Delta^\top B') \end{aligned}$$

On pose alors $\mu = 1 - \sigma\Delta^\top B'\Delta = 1 - \Delta^\top A\Delta$:

$$\begin{aligned} (B'^{-1}A)^2 &= f^2 (I - \sigma(\mu + 1)\Delta\Delta^\top B') \\ &= f^2 ((\mu + 1)(I - \sigma\Delta\Delta^\top B') - \mu I) \\ &= f(\mu + 1)B'^{-1}A - f^2\mu I \\ &= \frac{\sigma}{\mu}(\mu + 1)B'^{-1}A - \frac{\sigma^2}{\mu}I \end{aligned}$$

Finalement, on a :

$$\mu(B'^{-1}A)^2 = \sigma(\mu + 1)B'^{-1}A - \sigma^2 I$$

D'où, en posant $Q(x) = \mu x^2 - (\mu + 1)\sigma x + \sigma^2$:

$$Q(B'^{-1}A) = 0$$

Annexe 3 : Lien entre les racines d'un polynôme réel de degré 3 et le signe de son discriminant

On considère le polynôme de degré 3 $P(x) = ax^3 + bx^2 + cx + d$, avec $(a, b, c, d) \in \mathbb{R}^4$. Le discriminant du polynôme P est donné par la formule :

$$D = 18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2$$

Le lien entre les racines de P et le signe de D est le suivant :

- $D > 0$: P possède trois racines réelles distinctes,
- $D = 0$: P possède une ou deux racines réelles, distinctes dans le dernier cas,
- $D < 0$: P possède trois racines distinctes, dont une réelle et deux complexes conjuguées.

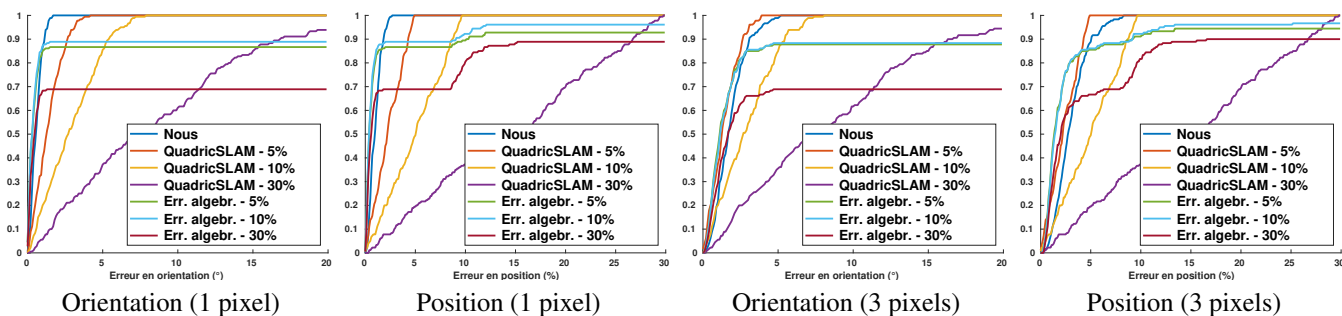


FIGURE 4 – Fonctions de répartition des erreurs finales en orientation et en position des caméras, selon le bruit (en pixels) appliqué sur les ellipses. En légende : "xxx - n%" désigne l'optimisation de l'erreur xxx à partir d'une position initiale perturbée de maximum n%.

Références

- [1] E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality : A hands-on survey," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 12, pp. 2633–2651, Dec 2016.
- [2] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *CVPR*, 2011, pp. 2969–2976.
- [3] M. A. Fischler and R. C. Bolles, "Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [4] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN : 0521540518, second edition, 2004.
- [5] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Eppnp : An accurate $O(n)$ solution to the pnp problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, 2009.
- [6] C. Xu, L. Zhang, L. Cheng, and R. Koch, "Pose estimation from line correspondences : A complete analysis and a series of solutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1209–1222, 06 2017.
- [7] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms : a survey from 2010 to 2016," *IPSJ Transactions on Computer Vision and Applications*, vol. 9, no. 1, pp. 16, Jun 2017.
- [8] R. B. Girshick, "Fast R-CNN," in *ICCV*, 2015, pp. 1440–1448.
- [9] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN : towards real-time object detection with region proposal networks," in *NIPS*, 2015, pp. 91–99.
- [10] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," in *ICCV*, 2017, pp. 2980–2988.
- [11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C.-Y. Fu, and A. C. Berg, "SSD : single shot multibox detector," in *ECCV*, 2016, pp. 21–37.
- [12] J. Redmon, S. Kumar Divvala, R. B. Girshick, and A. Farhadi, "You only look once : Unified, real-time object detection," in *CVPR*, 2016, pp. 779–788.
- [13] J. Redmon and A. Farhadi, "YOLO9000 : better, faster, stronger," in *CVPR*, 2017, pp. 6517–6525.
- [14] J. Redmon and A. Farhadi, "Yolov3 : An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018.
- [15] M. Crocco, C. Rubino, and A. Del Bue, "Structure from motion with objects," in *CVPR*, June 2016.
- [16] P. Gay, V. Bansal, C. Rubino, and A. Del Bue, "Probabilistic structure from motion with objects (psfmo)," in *ICCV*, 2017, pp. 3094–3103.
- [17] L. Nicholson, M. Milford, and N. Sünderhauf, "Quadricslam : Dual quadrics from object detections as landmarks in object-oriented slam," *IEEE Robotics and Automation Letters*, vol. 4, no. 1, pp. 1–8, Jan 2019.
- [18] C. Rubino, M. Crocco, and A. Del Bue, "3d object localisation from multi-view image detections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1281–1294, June 2018.
- [19] D. S. Wokes and P. L. Palmer, "Autonomous pose determination of a passive target through spheroid modelling," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, August 2008.
- [20] D. S. Wokes and P. L. Palmer, "Perspective reconstruction of a spheroid from an image plane ellipse," *International Journal of Computer Vision*, vol. 90, no. 3, pp. 369–379, 2010.
- [21] D. Eberly, "Reconstructing an ellipsoid from its perspective projection onto a plane," <https://www.geometrictools.com/>, May 2007, Updated version : March 1, 2008.
- [22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, third edition, 1996.
- [23] H. Avron, E. Ng, and S. Toledo, "A generalized courant-fischer minimax theorem," <https://escholarship.org/uc/item/4gb4t762>, August 2008.