Utilisation de la cohérence globale entre silhouettes pour la modélisation d'objets 3D à partir d'images non-calibrées

Carlos Hernández Francis Schmitt Roberto Cipolla

Résumé

Nous présentons un système complet et opérationnel pour la numérisation d'objets 3D à partir d'images en rotation non-calibrées. Dans cet article nous introduisons et développons un nouveau critère de cohérence globale d'un ensemble de silhouettes générées par un objet 3D. Nous montrons comment la maximisation de la cohérence entre silhouettes peut être exploitée pour l'estimation du mouvement de la caméra et de sa distance focale.

La cohérence entre silhouettes apparaît comme une généralisation du critère bien connu des points de tangence épipolaire, qui permet d'estimer le mouvement d'une caméra à partir des silhouettes d'un objet 3D comme seule source d'information. La cohérence entre silhouettes permet notamment d'exploiter toute l'information le long des silhouettes et non pas seulement aux seuls points de tangence épipolaire. Elle peut en particulier être utilisée dans des cas pratiques où l'absence de points caractéristiques ou de points de tangence épipolaire rend l'estimation des caméras très difficile.

Nous proposons un algorithme qui exploite de façon robuste et efficace la cohérence globale entre silhouettes afin d'estimer le mouvement de la caméra et sa distance focale. Nous utilisons cet algorithme pour la modélisation 3D d'objets à partir d'images en rotation non-calibrées. Le système a été validé par la numérisation de plus de 100 objets de musées avec une très haute qualité. Cinq exemples sont montrés dans cet article. L'algorithme a aussi été évalué de manière quantitative en comparant ses performances avec celles obtenues par une des méthodes les plus performantes de l'état de l'art qui exploitent seulement le critère des points de tangence épipolaire.

1 Introduction

Les techniques de vision par ordinateur sont de plus en plus utilisées pour l'acquisition de modèles 3D de haute qualité à partir de séquences d'images. Ceci est particulièrement vrai dans le cadre de l'archivage numérique du patrimoine culturel, comme celui des objets de musée qui sont ainsi rendus plus accessibles, via internet, aux peuples du monde entier.

Récemment, un certain nombre de techniques de reconstruction 3D multi-stéréo ont été développées. Ces techniques sont maintenant capables de produire à partir d'images calibrées des modèles 3D très denses avec une texture de haute qualité. Un modèle 3D est alors généralement optimisé afin de rendre cohérentes les multiples vues d'une même surface en employant des méthodes de "space carving" [1], des modèles déformables [2], ou des méthodes de "graph-cut" [3].

Un élément clef pour rendre ces méthodes véritablement pratiques est qu'elles puissent être utilisables par des usagers non-experts en vision par ordinateur. Par exemple, un photographe de musée devrait pouvoir se consacrer seulement à l'acquisition de séquences de photos de haute qualité. Une étape supplémentaire de calibrage de la caméra serait pour lui un obstacle important à l'acquisition d'objets 3D de musée, d'autant plus qu'entre 12 et 72 images sont typiquement acquises par objet. Il apparaît donc qu'un calibrage automatique de la caméra est essentiel pour ce type d'application.

Parmi les techniques de calibrage de caméras, les méthodes basées sur des points caractéristiques sont les plus connues (voir [4] pour une revue de synthèse). Ces méthodes sont fondées sur la présence de points caractéristiques sur la surface de l'objet et peuvent fournir des résultats de calibrage très précis. Malheureusement, des points caractéristiques ne sont pas toujours disponibles ou fiables (voir par exemple les quatre images de la Fig. 6.d). Pour de telles séquences, il existe des algorithmes alternatifs qui utilisent le contour de l'objet comme seule source d'information. Ils exploitent la notion de tangente épipolaire et de point frontière [5–7]. Pour garantir des résultats de bonne précision, ces méthodes nécessitent des silhouettes de très bonne qualité, ce qui rend difficile en pratique leur intégration dans un système complet. Dans le cas particulier d'un mouvement en rotation, l'étape la plus difficile est dans la segmentation des silhouettes est la séparation de l'objet du plateau tournant sur lequel il est posé. Une solution possible est de couper le bas des silhouettes. De telles silhouettes coupées se rencontrent également dans l'acquisition d'un détail sur un objet de plus grande dimension (voir à la Fig. 6 divers exemples d'extraction de silhouette dans le cas d'un mouvement circulaire).

Nous présentons une nouvelle approche pour l'estimation du mouvement de la caméra et de sa distance focale à partir de silhouettes. Notre approche exploite la notion de *cohérence entre silhouettes*. Celle-ci nous permet d'imposer la contrainte géométrique clef entre les silhouettes d'un même objet 3D rigide, à savoir l'existence d'un objet 3D ayant généré ces silhouettes. La technique proposée étend les méthodes précédentes en traitant de manière naturelle les silhouettes partielles ou coupées pour lesquelles l'évaluation et l'appariement des points de tangence épipolaire peuvent être très difficiles. Elle exploite également plus d'information que celle disponible uniquement aux points de tangence épipolaire. La méthode proposée est particulièrement intéressante lorsqu'elle est combinée avec une technique de reconstruction 3D permettant de fusionner les silhouettes avec d'autres informations [2, 3, 8].

Cet article est organisé de la manière suivante : dans la section 2 nous passons en revue la littérature existante. Dans la section 3 nous formulons notre problème. Dans la section 4 nous présentons le concept de cohérence entre silhouettes. Dans la section 5 nous décrivons un algorithme pratique pour l'estimation du mouvement de la caméra. Enfin, nous présentons des résultats expérimentaux et illustrons les performances de la méthode avec des reconstructions de haute qualité dans la section 6.

2 Travaux précédents

Notre approche de cohérence entre silhouettes est liée à trois types de techniques différentes connues : l'estimation du mouvement d'une caméra par autocalibrage, le recalage entre un ensemble de silhouettes et un modèle 3D donné, et le calcul de l'enveloppe visuelle [9].

Il existe beaucoup d'algorithmes pour **l'estimation du mouvement d'une caméra et l'autocalibrage** [4]. Ils sont fondés sur la correspondance des mêmes primitives détectées sur des images différentes. Dans le cas particulier d'un mouvement circulaire, des méthodes dédiées [10, 11] donnent de bons résultats quand les images contiennent assez de texture pour permettre une détection robuste des primitives. Une alternative à ces méthodes consiste à exploiter les silhouettes à la place de la texture. Les silhouettes ont été principalement employées pour l'estimation des caméras en utilisant le concept de point de *tangence épipolaire* [5], [6], c'est-à-dire des points du contour de la silhouette où la tangente à la silhouette est une ligne épipolaire. Une littérature riche existe sur l'exploitation des tangentes épipolaires, à la fois pour des caméras orthographiques [5, 7, 12, 13] et des caméras perspectives [14–17]. En particulier, les travaux de Mendonça et al [15] et de Wong et Cipolla [16] utilisent seulement les deux tangentes épipolaires extérieures, ce qui élimine le besoin d'apparier les tangentes épipolaires entre différentes images. Bien que ces méthodes aient donné de bons résultats, leur principal inconvénient est le nombre très limité de points de tangence épipolaire par paire d'images, généralement réduit à deux : un au dessus et un en dessous de la silhouette. Dès que nous disposons d'un plus grand nombre de points de tangence épipolaire, le problème est alors de les apparier entre vues différentes et de gérer leur visibilité, comme cela est proposé dans [13, 17].

En ce qui concerne le *recalage entre un ensemble d'images et un modèle 3D*, il existe quelques algorithmes qui utilisent pour cela les silhouettes. Dans [18] les auteurs proposent un algorithme pour estimer la pose d'un objet 3D par rapport à un ensemble de silhouettes en minimisant la distance 3D entre l'objet et les rayons optiques générés par les contours des silhouettes. Au lieu d'utiliser une distance 3D, le recalage peut également être accompli en minimisant l'erreur entre les contours des silhouettes et les contours de l'objet projeté. Dans [19] et [20] l'erreur est définie comme la somme des distances entre quelques points échantillonnés sur un contour et les points les plus proches de l'autre contour. Dans [21] une méthode accélérée matériellement est utilisée pour calculer la similarité entre deux silhouettes définie comme le cardinal de leur intersection.

Le **calcul de l'enveloppe visuelle** est un domaine très actif puisqu'il est l'une des manières les plus rapides et robustes pour obtenir un premier modèle 3D d'un objet. Il peut être suffisamment précis pour des applications temps réel telles que [22] et [23], ou utilisé comme initialisation pour des algorithmes de reconstruction 3D plus poussé comme celui décrit dans [2].

Dans cet article nous présentons une approche dérivée de [21] mais avec la différence importante que nous n'avons pas besoin d'un modèle 3D explicite de l'objet réel. Dans notre cas le modèle 3D est reconstruit *implicitement* à partir des silhouettes par une méthode d'enveloppe visuelle en même temps que le calibrage. En particulier, l'utilisation de la technique décrite dans Matutsik et al [24] permet d'effectuer tous les calculs dans le domaine de l'image, ce qui permet d'éviter de recourir à une représentation 3D explicite.

Même si le concept de cohérence entre silhouettes apparaît dans la littérature sous des noms différents, elle n'a jamais été exploitée auparavant pour le problème de l'estimation de caméras. Bottino et Laurentini [25] étudient le problème de la *compatibilité* entre silhouettes dans le cas de la projection orthographique et donnent quelques règles pour déterminer si un ensemble de silhouettes correspond ou non à celles d'un objet réel. Ils ne fournissent cependant aucune manière de mesurer quantitativement la *compatibilité* d'un ensemble de silhouettes. Dans son manuscrit de thèse, Cheung [26] emploie le terme d'alignement *stables* (en anglais : "consistent") pour le recalage de deux enveloppes visuelles. Cependant, il rejette son utilisation parce qu'il l'estime trop cher à calculer pour être exploitable en pratique.

Dans cet article nous présentons le nouveau concept de *cohérence* entre silhouettes et établissons un lien avec la géométrie épipolaire, et plus précisément avec le critère des points de tangence épipolaire employé par Wong et Cipolla [16]. En particulier, le critère des points de tangence épipolaire peut être vu comme une mesure de cohérence entre silhouettes dans le cas particulier de seulement deux silhouettes. En utilisant un ensemble plus important de silhouettes, le critère proposée étend le critère des points de tangence épipolaire en exploitant toute l'information contenue dans les contours des silhouettes et pas simplement aux points de tangence épipolaire, comme c'est le cas dans [16]. Ce critère nous permet d'estimer correctement le mouvement de la caméra et sa distance focale même s'il n'y a aucune tangence épipolaire disponible.

3 Problématique

Nous considérons un modèle de caméra perspective où la relation entre un point 3D $\mathbf{M}$ et sa projection 2D $\mathbf{m}$ est entièrement représentée par la matrice 3×4 P [4] :

$$\mathbf{m} \simeq \mathsf{P}\mathbf{M} \simeq \mathsf{K}[\mathsf{R}|\mathbf{t}]\mathbf{M}. \tag{1}$$

La matrice de rotation 3×3 R et le vecteur $\mathbf{t}$ représentent l'orientation et la translation définissant la pose de le caméra. La matrice de calibrage K contient les paramètres intrinsèques de la caméra. Le facteur d'aspect et le *skew* étant supposés proches de l'idéal pour des caméras CMOS et CCD, les seuls paramètres intrinsèques que nous considérons sont la distance focale f (en pixel) et le point principal $(u_0, v_0)^\top$:

$$\mathsf{K} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

Etant donné que, sous l'hypothèse de mouvement circulaire, la translation $\mathbf{t}$ et le point principal $(u_0, v_0)^\top$ interviennent de façon très similaire dans l'équation de projection et peuvent donc ainsi se compenser mutuellement, nous considérons le point principal comme connu et égal au centre de l'image. Notre problème est donc d'estimer la pose $[\mathsf{R}|\mathbf{t}]$ et la distance focale f d'une séquence de caméras sous les hypothèses de mouvement circulaire et des paramètres intrinsèques constants. Les seules données en entrée sont les silhouettes d'un objet rigide.

4 La cohérence entre silhouettes

Supposons que nous disposions d'un ensemble de silhouettes d'un même objet 3D prises de différents points de vue et que les matrices de projection associées aux caméras soient connues. Nous voulons alors mesurer l'exactitude de la segmentation des silhouettes et des matrices de projection. Nous exploitons pour cela la principale information apportée par une silhouette : une classification binaire de tous les rayons optiques passant par le centre optique de la caméra correspondante. Ces rayons optiques sont labellisés par la silhouette comme *intersectant l'objet* (label **S**) s'ils appartiennent à l'intérieur de la silhouette, ou *non intersectant l'objet* (label **B**) s'ils appartiennent à l'extérieur de la silhouette.

Considérons un rayon optique appartenant à l´intérieur de la silhouette et donc classé comme **S**. Sa projection dans toute autre vue doit donc logiquement intersecter la silhouette correspondante de l'objet. La rétroprojection de

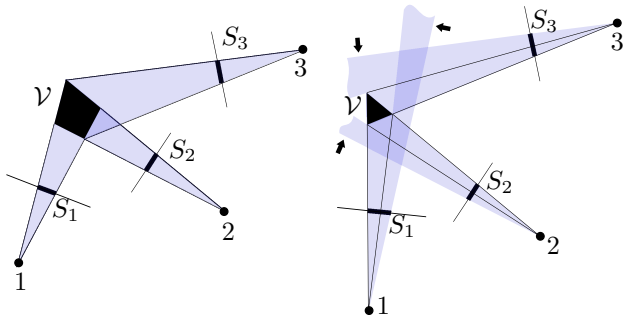


FIG. 1 – Exemples en 2D de différents degrés de cohérence entre silhouettes. L'enveloppe visuelle reconstruite $\mathcal{V}$ est le polygone noir. (a) Ensemble de silhouettes parfaitement cohérent. (b) Même ensemble de silhouettes mais avec une faible cohérence due à une pose imprécise des caméras. Les flèches indiquent les faisceaux de rayons optiques non-cohérents. Dans cet article le manque de cohérence est utilisé pour estimer le mouvement des caméras.

cette intersection sur le rayon optique 3D définit un ou plusieurs intervals de profondeur dans lesquels se situent les intersections du rayon avec la vraie surface 3D de l'objet.

A cause du bruit dans les silhouettes ou de l'imprécision des matrices de projection, l'affirmation précédente peut se révéler fausse, un rayon optique classé **S** par l'une des silhouettes pouvant avoir ses intervals de profondeur vides. Dans le cas de deux vues, les silhouettes correspondantes ne seront pas cohérentes s'il existe au moins un rayon optique classé **S** par une des silhouettes dont la projection dans l'autre vue n'intersecte pas la silhouette. Dans le cas de n vues, le manque de cohérence est défini par l'existence d'au moins un rayon optique dont les intervals de profondeur définis par les autres $n-1$ silhouettes ont une intersection vide. Ce manque de cohérence peut être mesuré simplement en comptant dans chaque silhouette le nombre de rayons optiques qui ne sont pas cohérents avec les $n-1$ autres silhouettes.

Deux exemples de différents degrés de cohérence entre silhouettes sont montrés dans le cas 2D à la Fig. 1. Les cônes correspondant aux parties des silhouettes qui sont incohérentes avec les autres silhouettes sont marqués par une flèche dans la Fig. 1.b .

Une façon simple de calculer la mesure de cohérence est la suivante :

- reconstruire l'enveloppe visuelle définie par les silhouettes,
- projeter l'enveloppe visuelle sur les caméras pour déterminer ses silhouettes, et
- comparer les silhouettes de l'enveloppe visuelle avec les silhouettes originales.

Dans le cas de données idéales (segmentation parfaite des silhouettes et matrices de projection exactes), les silhouettes de l'enveloppe visuelle reconstruite seront identiques aux silhouettes originales (voir Fig. 1a). Mais avec des données réelles les silhouettes et les matrices de projection ne sont plus parfaites, ce qui implique que les silhouettes originales et les silhouettes projetées ne sont plus identiques, les silhouettes de l'enveloppe visuelle reconstruite étant **toujours contenues** dans les silhouettes originales. Ceci est justifié mathématiquement dans la section suivante.

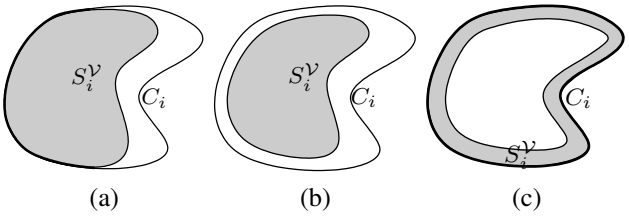


FIG. 2 – Trois scénarios différents pour la comparaison des silhouettes. La silhouette de l'enveloppe visuelle $S_i^{\mathcal{V}}$ est indiquée en gris. L'intersection $C_i \cap S_i^{\mathcal{V}}$ est dessiné en trait gras. (a) Scénario le plus fréquent. (b) Scénario où la cohérence calculée avec les contours est zéro tandis que la cohérence calculée avec les aires est beaucoup plus grande. (c) Scénario où la silhouette de l'enveloppe visuelle a un trou. La cohérence calculée avec les contours est 1 tandis que la cohérence calculée avec les aires est beaucoup plus petite.

4.1 Contrainte géométrique

Soit S_i la $i^{ème}$ silhouette et P_i la matrice de projection de la caméra correspondante. Nous pouvons définir le cône $\mathcal{V}_i$ généré par la silhouette S_i comme l'ensemble de points 3D $\mathbf{M}$ tels que :

$$\mathcal{V}_i = \{\mathbf{M} \in \mathbb{R}^3 \ : \ \mathsf{P}_i\mathbf{M} \in S_i\}. \tag{3}$$

L'enveloppe visuelle reconstruite $\mathcal{V}$ définit par l'ensemble de silhouettes S_i, $i = 1, \ldots, n$ peut être décrite comme l'intersection de cônes suivante :

$$\mathcal{V} = \bigcap_{i=1,\ldots,n} \mathcal{V}_i = \{\mathbf{M} \in \mathbb{R}^3 \ : \ \mathsf{P}_i\mathbf{M} \in S_i \ \forall i\}. \tag{4}$$

La silhouette de $\mathcal{V}$ dans la $i^{ème}$ image, notée $S_i^{\mathcal{V}}$, est définie comme l'ensemble de points 2D $\mathbf{m}$ tels que :

$$S_i^{\mathcal{V}} = \{\mathbf{m} = \mathsf{P}_i\mathbf{M} \ : \ \mathbf{M} \in \bigcap_{j=1,\ldots,n} \mathcal{V}_j\}. \tag{5}$$

A partir de (3) et (5), nous pouvons séparer la contribution de la silhouette S_i à $S_i^{\mathcal{V}}$ de la façon suivante :

$$S_i^{\mathcal{V}} = S_i \cap \{\mathbf{m} = \mathsf{P}_i\mathbf{M} \ : \ \mathbf{M} \in \bigcap_{j \neq i} \mathcal{V}_j\}. \tag{6}$$

Nous déduisons donc de (6) la relation suivante : $S_i^{\mathcal{V}} \subseteq S_i \ \forall i$. Si les silhouettes et les matrices de projection sont parfaites, alors :$S_i^{\mathcal{V}} = S_i \ \forall i$.

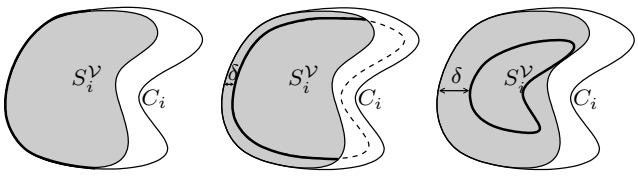


FIG. 3 – Comparaison de la mesure de cohérence pour des valeurs de δ croissantes de gauche à droite. La silhouette originale S_i correspond au contour externe. La silhouette de l'enveloppe visuelle $S_i^{\mathcal{V}}$ est en grisé. Le terme $(C_i \ominus \delta) \cap S_i^{\mathcal{V}}$ dans l'Equ. (9) est indiqué en trait gras.

4.2 Une mesure de cohérence

Nous cherchons donc à déterminer une mesure de cohérence $\mathcal{C}$ entre une silhouette S_i et la projection correspondante $S_i^{\mathcal{V}}$ de l'enveloppe visuelle. Une réponse rapide serait d'utiliser le rapport d'aires entre les deux silhouettes comme proposé dans [21] :

$$\mathcal{C}_1(S_i, S_i^{\mathcal{V}}) = \frac{\int S_i^{\mathcal{V}}}{\int S_i} = \frac{\int (S_i \cap S_i^{\mathcal{V}})}{\int S_i} \in [0,1]. \quad (7)$$

Mais cette mesure a deux inconvénients : une précision réduite et un très gros temps de calcul. La précision devient en effet problématique lorsque, dans le cas de grandes images, l'aire des silhouettes devient très élevée et que les différences entre les deux silhouettes s'amenuisent, la dynamique de la mesure devenant alors très petite et insuffisante pour les applications que nous envisageons. Le deuxième inconvénient est son temps de calcul, car l´évaluation de la mesure étant discrétisée, le temps de calcul est proportionnel à l'aire de la silhouette à évaluer.
La solution à ces deux problèmes est obtenue par une simple substitution dans (7) de la silhouette S_i par son contour, noté C_i :

$$\mathcal{C}(S_i, S_i^{\mathcal{V}}) = \frac{\int (C_i \cap S_i^{\mathcal{V}})}{\int C_i} \in [0,1]. \quad (8)$$

Une observation importante concernant l'utilisation des contours au lieu de la silhouette elle même doit être faite : les deux mesures peuvent être très différentes dans les cas particuliers montrés dans les Fig. 2b et Fig. 2c. Le fait d'utiliser les contours à la place des aires pénalisera les scénarios comme celui de la Fig. 2b tout en encourageant les scénarios de la Fig. 2c. Le cas b arrive beaucoup plus souvent que le cas c dans le problème qui nous concerne. Il peut se produire facilement lorsqu'une distance focale estimée est plus petite pour une des vues. Par contre, si aucune des silhouettes n'a des trous, le cas c devient impossible étant données les propriétés de l'enveloppe visuelle.
La faiblesse ci-dessus due à l'utilisation des contours peut être corrigée avec l'utilisation d'un offset δ dans (8) (voir Fig. 3). Pour un offset δ donné, nous substituons dans (8) le contour C_i par sa version érodée de δ pixels $C_i \ominus \delta$, ce qui donne :

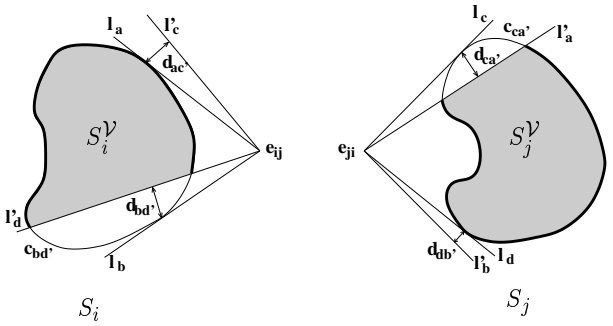


FIG. 4 – Comparaison du critère de cohérence entre silhouettes et du critère des points de tangence épipolaire pour $n = 2$ silhouettes. Les silhouettes de l'enveloppe visuelle $S_i^{\mathcal{V}}$ et $S_j^{\mathcal{V}}$ sont montrées en gris. Les termes $C_i \cap S_i^{\mathcal{V}}$ et $C_j \cap S_j^{\mathcal{V}}$ sont dessinés en trait gras. Les deux critères sont équivalents dans le cas de 2 vues : tous les deux minimisent les secteurs définis par l_b et l'_d, et l_c etl'_a.

$$\mathcal{C}^\delta(S_i, S_i^{\mathcal{V}}) = \frac{\int ((C_i \ominus \delta) \cap S_i^{\mathcal{V}})}{\int (C_i \ominus \delta)} \in [0,1]. \quad (9)$$

Plus δ est grand, plus la nouvelle mesure est robuste face à une mauvaise segmentation. Mais cette robustesse s'obtient au détriment de la précision. Pour un offset δ donné, la mesure ne pourra pas faire la différence entre la silhouette originale et la silhouette de l'enveloppe visuelle déterminée avec une erreur plus petit que δ. Des valeurs typiques de δ varient entre 0.25 et 1 pixel, ce choix dépendant de la qualité de la segmentation des silhouettes.
La mesure $\mathcal{C}^\delta(S_i, S_i^{\mathcal{V}})$ évalue la *cohérence* entre la silhouette S_i et toutes les autres silhouettes $S_{j\neq i}$ qui ont contribué à l'enveloppe visuelle. Etant donné que $S_i^{\mathcal{V}}$ est complètement déterminé par les contours des silhouettes $C_{i=1,\cdots,n}$, la mesure peut aussi être notée comme $\mathcal{C}^\delta(C_i, C_{j\neq i})$. Pour calculer la cohérence totale entre toutes les silhouettes, nous calculons simplement la cohérence moyenne entre chaque silhouette et les $n-1$ autres silhouettes :

$$\mathcal{C}^\delta(C_1, \ldots, C_n) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{C}^\delta(C_i, C_{j\neq i}) \in [0,1]. \quad (10)$$

4.3 Relation avec la géométrie épipolaire

Le critère de cohérence entre silhouettes proposé peut être vu comme une extension du critère des points de tangence épipolaire. Pour une paire de vues données (voir Fig.4), le critère des points de tangence épipolaire minimise la distance au carré entre les tangentes épipolaires dans une vue (l_a et l_b dans la vue i, l_c et l_d dans la vue j) et les tangentes épipolaires transférées de l'autre vue (l'_c et l'_d dans la vue i, l'_a et l'_b dans la vue j). Autrement dit, il minimise la somme des distances au carré $\mathcal{C}_{et}(C_i, C_j) = d^2_{ac'} + d^2_{bd'} + d^2_{ca'} + d^2_{db'}$. Pour la même paire de silhouettes, l'optimisation du

Algorithm 1 Cohérence entre silhouettes $\mathcal{C}^\delta(C_i, C_{j\neq i})$

Require: Matrices de projection $\mathsf{P}_i, \forall i$, contour de référence C_i, liste de contours $C_{j\neq i}$, offset δ, nombre N d'échantillons par contour
  Construire liste de points $\mathbf{m}^{(k)}$, en échantillonnant N points dans $C_i \ominus \delta$
  for all $\mathbf{m}^{(k)}$ **do**
    Initialiser interval 3D $I_{3D} = [0, \infty]$
    Initialiser compteur $N' = 0$
    for all $C_{j\neq i}$ **do**
      Projeter rayon optique $l = \mathsf{P}_j \mathsf{P}_i^{-1} \mathbf{m}^{(k)}$
      Calculer interval d'intersection 2D $I_{2D} = l \cap C_j$
      Rétroprojeter interval 2D sur $I_{3D} = I_{3D} \cap \mathsf{P}_j^{-1} I_{2D}$
    end for
    if $I_{3D} \neq \emptyset$ **then**
      $N' = N' + 1$
    end if
  end for
  Retourner $\frac{N'}{N}$

critère de cohérence correspond à la maximisation des longueurs $C_i \cap S_i^{\mathcal{V}}$ et $C_j \cap S_j^{\mathcal{V}}$. Nous nous apercevons que, en-dehors des configurations dégénérées, les deux critères cherchent à minimiser les secteurs définis par les tangentes épipolaires dans une vue et les tangentes épipolaires correspondantes de l'autre vue. Donc, si nous optimisons le critère de cohérence *par paires* de silhouettes, nous obtenons le même comportement qu'avec les méthodes fondées sur les tangentes épipolaires, comme par exemple [16].
Si nous utilisons le critère de cohérence pour $n > 2$, les silhouettes ne sont plus prises *par paires* mais toutes en même temps. Ceci implique que l'information qui est exploitée par le critère de cohérence n'est pas seulement les points de tangence épipolaire, mais **tout** le contour de la silhouette. Ceci implique que, même si nous ne disposons pas de tangentes épipolaires, le critère de cohérence reste toujours valide. Nous montrons deux exemples dans la Section 6 (Fig. 6 a et b) où nous n'avons ni le haut et ni le bas des silhouettes (il n'y a donc pas de tangentes épipolaires extérieures disponibles) mais pour lesquels nous sommes capables d'estimer le mouvement et la distance focale de la caméra avec une très bonne précision.

4.4 Implantation

Nous présentons une implantation rapide du critère de cohérence entre silhouettes $\mathcal{C}^\delta$, obtenue par une discrétisation du contour $C_i \ominus \delta$ en un nombre de points d'échantillonnage équirépartis tout au long du contour. Le terme $(C_i \ominus \delta) \cap S_i^{\mathcal{V}}$ est évalué en testant pour chaque point d'échantillonnage si le rayon optique associé intersecte ou non l'enveloppe visuelle avec une approche par lancé de rayons [24]. Une version simplifiée de cet algorithme a été utilisée en ne prenant pas en compte les contours intérieurs des silhouettes de genre > 0. Nous n'avons pas ainsi calculé tous les intervals de profondeur pour un rayon optique donné mais simplement le minimum et le maximum de l'intersection avec chaque silhouette. Cette mesure est une approximation conservative de la vraie cohérence, la valeur que nous obtenons étant toujours égale ou supérieure à la vraie valeur. En pratique, la différence avec la cohérence calculée avec tous les intervals est petite.

Algorithm 2 Estimation du mouvement et de la focale

Require: Séquence d'images $I_{i=1,\cdots,n}$
  Extraire contours C_i à partir de I_i ([28,29]),
  Initialiser $v = (\theta_{\mathbf{a}}, \phi_{\mathbf{a}}, \alpha_{\mathbf{t}}, \Delta\omega_i, f)$=($\frac{\pi}{2}$, $\frac{\pi}{2}$, 0, $\frac{2\pi}{n}$, f_0),
  Initialiser algorithme de Powell [27]
  repeat {voir [27] pour plus de détails}
    $v' = v$
    v=Powell(v') {une iteration de Powell avec Algorithme 3}
  until $||v - v'|| < \epsilon$

Algorithm 3 Cohérence globale entre silhouettes

Require: Séquence de contours $C_{i=1,\cdots,n}$, paramètres des caméras $(\theta_{\mathbf{a}}, \phi_{\mathbf{a}}, \alpha_{\mathbf{t}}, \Delta\omega_i, f)$
  $\mathbf{a} = (\sin(\theta_{\mathbf{a}})\cos(\phi_{\mathbf{a}}), \sin(\theta_{\mathbf{a}})\sin(\phi_{\mathbf{a}}), \cos(\theta_{\mathbf{a}}))^\top$
  $\mathbf{t} = (\sin(\alpha_{\mathbf{t}}), 0, \cos(\alpha_{\mathbf{t}}))^\top$
  $\omega_1 = 0,\ \omega_j = \omega_{j-1} + \Delta\omega_{j-1},\ 1 < j \leq n$
  $[\mathsf{R}_i|\mathbf{t}_i] = [\mathsf{R}_{\mathbf{a}}(\omega_i)|\mathbf{t}]\ \forall i, K(1,1) = f, K(2,2) = f$
  $sc = 0$
  for all C_i **do**
    $sc = sc + \mathcal{C}^\delta(C_i, C_{j\neq i})$ {Algorithm 1}
  end for
  Retourner $\frac{sc}{n}$

L'algorithme qui décrit la cohérence entre silhouettes $\mathcal{C}^\delta(C_i, C_{j\neq i})$ entre le contour d'une silhouette C_i et les $n-1$ autres contours est détaillé dans l'Algorithme 1.

Si N est le nombre d'échantillons par silhouette, et n est le nombre de silhouettes, la complexité de $\mathcal{C}^\delta(C_i, C_{j\neq i})$ est en $\mathcal{O}(nN\log(N))$ et la complexité de $\mathcal{C}^\delta(C_i, \ldots, C_n)$ dans (10) est en $\mathcal{O}(n^2 N\log(N))$. Comme exemple, le temps de calcul d'une évaluation de (10) avec un processeur Athlon de 1.5 GHz est de 750 ms pour la séquence du Botijo de la Fig. 5 ($n = 18$, $N \approx 6000$).

5 Estimation des caméras

Nous montrons dans cette section comment exploiter la cohérence entre silhouettes pour estimer le mouvement de la caméra et sa distance focale dans le cadre du mouvement circulaire. L'idée est d'utiliser la cohérence comme coût dans une procédure d'optimisation. Pour n vues, nous paramétrons le mouvement circulaire avec $n+3$ paramètres de la façon suivante : les coordonnées sphériques de l'axe de rotation $(\theta_{\mathbf{a}}, \phi_{\mathbf{a}})$ (2 paramètres), la direction de la translation $\alpha_{\mathbf{t}}$ (1 paramètre), les angles entre deux vues consécutives $\Delta\omega_i$ ($n-1$ paramètres) et la distance focale f (1 paramètre). Nous utilisons l'algorithme d'optimisation de Powell [27] pour maximiser (10). Plusieurs centaines d'évaluations sont typiquement nécessaires avant convergence. L'algorithme complet est décrit dans les algorithmes 2 et 3, où f_0 est la distance focale donnée à l'initialisation.

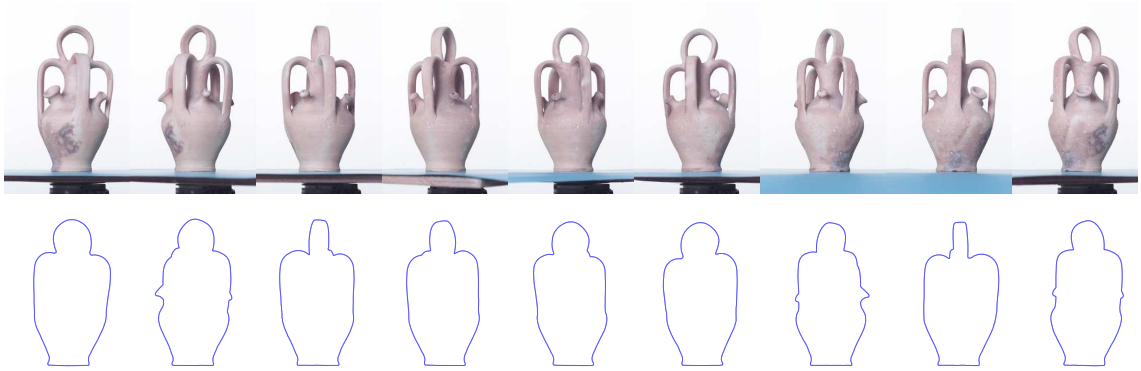
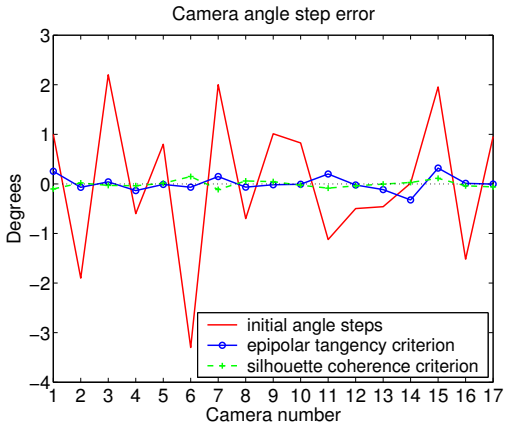


FIG. 5 – Séquence Botijo. Haut : quelques unes des images originales extraites d'une séquence de 18 images. Bas : contours polygonaux extraits à partir des images. Voir matériel supplémentaire pour une vidéo du processus d'optimisation.

Botijo		rotation axis (degrees)		translation (degrees)	focal (pixels)
		$\theta_{\mathbf{a}}$	$\phi_{\mathbf{a}}$	$\alpha_{\mathbf{t}}$	f
initial		90.0000	90.0000	0.0000	5000
calibrated		99.671	90.3431	0.4266	6606
$\mathcal{C}_{et}$	recovered	99.6345	90.3050	0.4314	6576
	error	0.0364	0.0381	0.0049	30
$\mathcal{C}^{\delta}$	recovered	99.6861	90.3419	0.4239	6635
	error	0.0152	0.0011	0.0026	29



TAB. 1 – Estimation du mouvement de la caméra et de sa distance focale pour la séquence Botijo. L'erreur moyen de l'angle entre deux vues consecutives est de 0.11 degrés pour le critère des points de tangence épipolaire ($\mathcal{C}_{et}$) et de seulement 0.06 degrés pour le critère de cohérence entre silhouettes ($\mathcal{C}^{\delta}$).

6 Résultats expérimentaux

Nous présentons un premier exemple avec la séquence Botijo composée de 18 images d'une taille de 2008x3040 pixels acquises avec un plateau tournant contrôlé par ordinateur. Les images ont été segmentées par une procédure automatique [28] (voir Fig. 5). Nous avons aussi acquis une deuxième séquence identique d'un étalon géométrique placé à la place de l'objet pour estimer très précisément (voir [30]) le mouvement de la caméra et ses paramètres intrinsèques afin d'évaluer la précision de la méthode proposée. La valeur de l'offset δ utilisée pour le calcul de la cohérence entre silhouettes est $\delta = 0.25$ pixels.

Nous montrons dans le Tableau 1 les résultats de l'estimation du mouvement (axe de rotation, direction de la translation et angles entre caméras) et de la distance focale. Un total de 21 paramètres sont estimés. Nous comparons la méthode de cohérence entre silhouettes à la méthode des tangentes épipolaires décrite dans [16]. Les résultats sont bons pour les deux critères, la cohérence entre silhouettes étant meilleure que le critère des points de tangence épipolaire pour l'estimation du mouvement de la caméra (voir le tableau 1). Les deux critères estiment la distance focale avec la même précision ($\sim 0.5\%$ d'erreur).

Le même algorithme a été employé sur plus d'une centaine de séquences non-calibrées. Nous illustrons à la Fig. 6 quatre de ces séquences qui sont particulièrement intéressantes. Pour la déesse Hécate (Fig. 6a) et le bronze chinois (Fig. 6b), les deux tangentes épipolaires extérieures ne sont pas disponibles, puisque le haut et le bas des silhouettes ont été coupés. Pour la statue de Millet (Fig. 6c) et le buste de Giganti (Fig. 6d) juste le bas a été coupé. Le problème de la séparation de l'image de l'objet de son plateau tournant est très fréquemment rencontré dans le cadre de la modélisation 3D. En général il est facile d'extraire le haut de l'objet, mais il est beaucoup plus difficile de le séparer du plateau tournant. Pour ces séquences nous validons les résultats de l'estimation du mouvement et de la distance focale par la qualité des reconstructions finales générées par l'algorithme décrit dans [2]. Il est intéressant de noter qu'il serait très difficile d'utiliser des points caractéristiques fiables dans le cas de la sculpture de Giganti (Fig. 6d) (sombre et très brillante), alors que le bronze chinois (Fig. 6b) est vraiment un cas extrême pour les algorithmes fondés sur les tangentes épipolaires.

7 Conclusions et perspectives

Nous avons développé une nouvelle approche pour l'estimation des caméras à partir des silhouettes. Elle est basée sur le concept de cohérence entre silhouettes, défini comme la similarité entre un ensemble de silhouettes et les silhouettes de leur enveloppe visuelle. Cette approche a été testée avec succès pour le calibrage d'une séquence avec mouvement circulaire. La précision des résultats obtenus est due à l'utilisation dans le calcul du contour complet de la silhouette, tandis que les autres méthodes ne prennent en compte que les points de tangence épipolaire. Nous avons validé l'approche proposée à la fois qualitativement et quantitativement.

Une limitation de l'implantation actuelle de la cohérence entre silhouettes est la discrétisation des contours. Afin de s'affranchir de cette source de bruit, une solution possible serait de calculer les silhouettes de l'enveloppe visuelle de façon exacte. Nous pourrions procéder comme dans [23], en utilisant une technique par lancé de rayons.

Enfin, la cohérence entre silhouettes pourrait être employée dans le cadre du mouvement général, mais un soin spécial doit être pris afin d'éviter les minima locaux et les problèmes de convergence, moins critiques dans le cas du mouvement circulaire.

Références

[1] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.

[2] C. Hernández and F. Schmitt, "Silhouette and stereo fusion for 3d object modeling," *Computer Vision and Image Understanding*, vol. 96, no. 3, pp. 367–392, 2004.

[3] G. Vogiatzis, P. Torr, and R. Cipolla, "Multi-view stereo via volumetric graph-cuts," in *CVPR*, 2005.

[4] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN : 0521623049, 2000.

[5] J. H. Rieger, "Three dimensional motion from fixed points of a deforming profile curve," *Optics Letters*, vol. 11, no. 3, pp. 123–125, 1986.

[6] J. Porrill and S. B. Pollard, "Curve matching and stereo calibration," *Image and Vision Computing*, vol. 9, no. 1, pp. 45–50, 1991.

[7] P. Giblin, F. Pollick, and J. Rycroft, "Recovery of an unknown axis of rotation from the profiles of a rotating surface," *J.Optical Soc. America*, vol. 11A, pp. 1976–1984, 1994.

[8] M. Lhuillier and L. Quan, "Surface reconstruction by integrating 3d and 2d data of multiple views," in *Proc. ICCV*, 2003, pp. 1313–1320.

[9] A. Laurentini, "The visual hull concept for silhouette based image understanding," *IEEE Trans. on PAMI*, vol. 16, no. 2, 1994.

[10] A. W. Fitzgibbon, G. Cross, and A. Zisserman, "Automatic 3D model construction for turn-table sequences," in *3D SMILE*, June 1998, pp. 155–170.

[11] G. Jiang, H. Tsui, L. Quan, and A. Zisserman, "Single axis geometry by fitting conics," in *ECCV*, vol. 1, 2002, pp. 537–550.

[12] B. Vijayakumar, D. Kriegman, and J. Ponce, "Structure and motion of curved 3d objects from monocular silhouettes," in *CVPR*, 1996, pp. 327–334.

[13] Y. Furukawa, A. Sethi, J. Ponce, and D. Kriegman, "Structure and motion from images of smooth textureless objects," in *ECCV 2004*, vol. 2, Prague, Czech Republic, May 2004, pp. 287–298.

[14] K. Åström, R. Cipolla, and P. Giblin, "Generalized epipolar constraints," *International Journal on Computer Vision*, vol. 33, no. 1, pp. 51–72, 1999.

[15] P. R. S. Mendonça, K.-Y. K. Wong, and R. Cipolla, "Epipolar geometry from profiles under circular motion," *IEEE Trans. on PAMI*, vol. 23, no. 6, pp. 604–616, June 2001.

[16] K.-Y. K. Wong and R. Cipolla, "Structure and motion from silhouettes," in *8th IEEE International Conference on Computer Vision*, vol. II, Vancouver, Canada, July 2001, pp. 217–222.

[17] S. N. Sinha, M. Pollefeys, and L. McMillan, "Camera network calibration from dynamic silhouettes," in *CVPR*, vol. 1, 2004, pp. 195–202.

[18] S. Sullivan and J. Ponce, "Automatic model construction, pose estimation, and object recognition from photographs using triangular splines," *IEEE Trans. on PAMI*, vol. 20, no. 10, pp. 1091–1096, 1998.

[19] K. Matsushita and T. Kanedo, "Efficient and handy texture mapping on 3d surfaces," *Computer Graphics Forum*, vol. 18, pp. 349–358, 1999.

[20] P. J. Neugebauer and K. Klein, "Texturing 3d models of real world objects from multiple unregistered photographic views," *Computer Graphics Forum*, vol. 18, pp. 245–256, 1999.

[21] H. Lensch, W. Heidrich, and H. P. Seidel, "A silhouette-based algorithm for texture registration and stitching," *J. of Graphical Models*, pp. 245–262, 2001.

[22] M. Li, M. Magnor, and H. Seidel, "Improved hardware-accelerated visual hull rendering," in *VMV*, November 2003, pp. 151–158.

[23] J.-S. Franco and E. Boyer, "Exact polyhedral visual hulls," in *14th British Machine Vision Conference (BMVC)*, September 2003, pp. 329–338.

[24] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan, "Image-based visual hulls," in *SIGGRAPH 2000*, 2000, pp. 369–374.

[25] A. Bottino and A. Laurentini, "Introducing a new problem : Shape-from-silhouette when the relative positions of the viewpoints is unknown," *IEEE Trans. on PAMI*, vol. 25, no. 11, pp. 1484–1493, 2003.

[26] K. Cheung, "Visual hull construction, alignment and refinement for human kinematic modeling, motion tracking and rendering," Ph.D. dissertation, Carnegie Mellon University, 2003.

[27] M. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Computer Journal*, vol. 17, pp. 155–162, 1964.

[28] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, pp. 359–369, 1998.

[29] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut - interactive foreground extraction using iterated graph cuts," *SIGGRAPH*, vol. 23, no. 3, pp. 309–314, 2004.

[30] J. M. Lavest, M. Viala, and M. Dhome, "Do we really need an accurate calibration pattern to achieve a reliable camera calibration ?" in *ECCV*, vol. 1, 1998, pp. 158–174.

FIG. 6 – Reconstructions 3D après estimation du mouvement de la caméra et de sa distance focale en utilisant la cohérence entre silhouettes. (a) Déesse Hécate (36 images de 14 megapixels). (b) Bronze chinois (24 images de 6 megapixels). (c) statue de Jean-François Millet par Henri Chapu (36 images de 6 megapixels). (d) Giganti de Camille Claudel (36 images de 6 megapixels). Gauche bas : silhouettes extraites à partir des images. Centre : visualisation avec ombrage de Gouraud des modèles reconstruits. Droite : modèles texturés.