Extrapolation d'aspect pour la reconnaissance d'objets sur séquences d'images

Jonathan GUINET[1], Stéphane HERBIN[1], Guy LE BESNERAIS[1], Sylvie PHILIPP-FOLIGUET[2]

[1]ONERA, DTIM/IED
BP-72, 92322 Chatillon Cedex

[2]Equipes Traitement des Images et du Signal (CNRS UMR 8051)
Université de Cergy-Pontoise/ENSEA
6, av. du Ponceau 95014 Cergy-Pontoise Cedex
guinet@onera.fr, herbin@onera.fr, lebesner@onera.fr,philipp@ensea.fr

Résumé – La reconnaissance d'objets 3D est un problème prépondérant en vision par ordinateur, en particulier lorsque les objets présentent de fortes variations d'aspect entre les observations. A partir d'une étude empirique sur la capacité d'extrapolation d'objets 3D, nous proposons une règle de décision originale paramétrée par la pose des objets. Cette règle de décision a été appliquée à un système de reconnaissance, et validée par des expériences sur données réelles. Les résultats montrent une amélioration significative du taux de reconnaissance sur les séquences de test.

Abstract – Tri-dimensional object recognition in video sequences is still a challenging problem in computer vision, in particular when objects have strong aspect variations between observations. Based on an empirical study about the extrapolation ability of 3D object models, we propose a novel decision scheme depending on pose estimation. Our decision rule has been applied to a generic recognition system, and has been validated by experiments on real video sequences. Recognition results improvement show the effectiveness of our approach on real test sequences.

1 Introduction

Nous nous intéressons dans cette étude à l'identification de véhicules mobiles présents dans des séquences vidéos acquises en incidence oblique (caméra sur bâtiment élevé, vidéosurveillance). La fonction d'identification est assimilée au problème de décision suivant : "un véhicule détecté dans une séquence courante a-t-il déjà été observé dans une séquence antérieure", considéré comme un composant élémentaire de fonctionnalités de réacquisition de piste.

Les véhicules présents dans les deux séquences vidéos — courante et antérieure — ne sont en général pas observés dans les mêmes conditions de prise de vue ni avec la même présentation. On suppose cependant qu'un travail préliminaire fournit la pose de l'objet. La principale difficulté à maîtriser est ainsi la variabilité des apparences d'objets produites par les changements d'aspect.

Le point central des travaux présentés ici est l'analyse des capacités d'extrapolation d'aspect d'un modèle d'objet estimé à partir d'une unique séquence vidéo. Le modèle exploité est un polyèdre muni de textures sur chacune de ses faces décrivant de manière simplifiée la surface 3D de l'objet ainsi que sa radiométrie apparente. Bien que schématique, il permet d'extrapoler l'apparence de l'objet sous le même aspect qu'une observation donnée, et de calculer un indice de similarité. Une analyse empirique fine de la capacité d'extrapolation d'aspect du modèle d'objet nous a amené à concevoir une stratégie de décision adaptative améliorant notablement les performances d'identification.

2 Etat de l'art

Le thème de la reconnaissance d'objet a reçu beaucoup d'intérêt depuis les débuts de la vision par ordinateur. On peut distinguer deux types d'approches, les approches basées "apparence", et les approches basées "modèle".

L'approche basée "apparence" s'appuie sur une description formelle de la variabilité des images, en général rendue accessible par une base d'apprentissage représentative [1]. La complexité des images impose en général de travailler dans un espace de primitives ou de caractéristiques censé résumer le contenu informatif utile des données. La définition d'un tel espace est une des activités critique de l'interprétation d'images. Ces approches exploitant des bases d'apprentissage sont en général les plus performantes, cependant elles présentent plusieurs inconvénients. Elles sont limitées par la représentativité de la base exploitée, ce qui implique de sélectionner avec soin la base d'apprentissage.

L'approche basée "modèle" s'appuie sur notre connaissance du monde physique pour définir un modèle de génération de l'image. Elles incluent une description de l'objet, du capteur, parfois de la scène et de ses composants (fond, source d'illumination ou d'occultation ...). Différentes descriptions ont été explorées; il peut s'agir d'un modèle polyédrique 3D projeté dans l'image [2], de modèle de forme 3D [3], de points d'intérêts [4] synthétisés dans un modèle de surface 3D [5]. L'étape de reconnaissance proprement dite consiste à comparer les modèles d'objets.

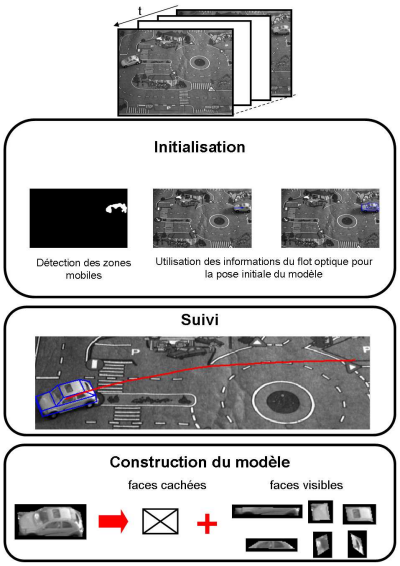

FIG. 1 – L'étape d'extraction de caractéristiques.

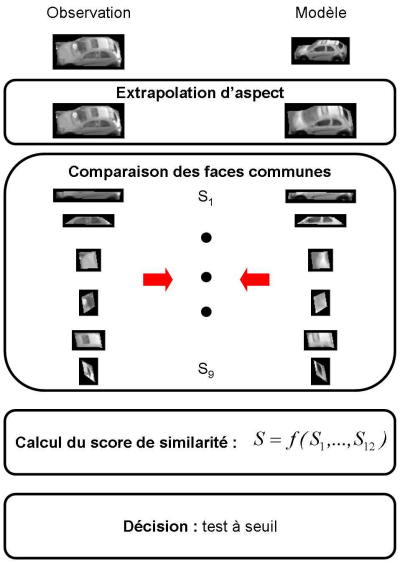

FIG. 2 – L'étape de reconnaissance.

3 Schéma algorithmique générique

L'algorithme utilisé suit le schéma classique des algorithmes de reconnaissance d'objets [6], et peut être scindé en deux étapes : extraction de caractéristiques et reconnaissance.

L'étape d'extraction de caractéristiques (figure 1) s'appuie sur la détection et le suivi de véhicule dans la séquence pour construire un modèle polyédrique [2] texturé à partir d'une estimation de la pose de l'objet. Chaque face est munie d'une description radiométrique locale calculée sur la séquence. Dans les conditions d'observation courantes, le véhicule détecté se déplace en ligne droite, et présente peu de variations d'aspect. Les effets projectifs peuvent limiter ainsi la qualité texturelle de certaines faces du modèle.

L'étape de reconnaissance (figure 2) consiste à comparer deux modèles polyédriques texturés construits à partir de deux séquences vidéos. La connaissance de la pose des véhicules permet de projeter les modèles polyédriques dans une géométrie commune, et de calculer une mesure de similarité sur les apparences de chaque face. L'utilisation d'un modèle tri-dimensionnel permet de gérer les changements d'aspect, en particulier l'apparition/disparition de certaines faces.

La décision finale ("s'agit-il du même véhicule ?") se réduit à un seuillage sur la mesure de similarité.

3.1 Seuillage adaptatif

Lors d'expériences préliminaires nous avons constaté l'influence prédominante des angles de prises de vue sur les valeurs des mesures de similarité et donc sur la fiabilité de la comparaison des modèles polyédriques texturés. La connaissance de l'angle de présentation des véhicules (par flot optique + pistage) nous a amené à étudier une démarche adaptative de seuillage dépendant de ces angles.

Le principe de l'approche est de contrôler la distribution intra-classe des mesures de similarité générée par un modèle construit sur une présentation d'objet donnée. Si $V(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0)$ est l'extrapolation de l'apparence d'un objet à partir d'un modèle de présentation $\boldsymbol{\theta}_0$ dans les conditions de prise de vue décrites par le paramètre $\boldsymbol{\theta}$ (figure 2), l'algorithme de comparaison entre deux modèles V et V' est fonction d'une mesure de similarité S et se réduit au test :

$$S(V(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0), V'(\boldsymbol{\theta} \mid \boldsymbol{\theta}'_0)) < \lambda(\boldsymbol{\theta})$$

où $\lambda(\boldsymbol{\theta})$ est un seuil dépendant de la condition de prise de vue commune $\boldsymbol{\theta}$ sous laquelle est calculée l'indice de similarité. Dans la suite, on prendra comme condition de prise de vue commune $\boldsymbol{\theta}'_0$ la présentation d'apprentissage du modèle V'.

En supposant que l'indice de similarité vérifie une inégalité triangulaire, on a :

$$S(V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}_0), V'(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}'_0)) \leq$$
$$S(V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}_0), V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}'_0)) + S(V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}'_0), V'(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}'_0))$$

Le contrôle de la similarité entre deux véhicules identiques vus sous des poses différentes : $S(V(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0), V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}'_0))$, c'est-à-dire la dispersion intra-classe de l'indice de similarité, permet alors de définir un seuillage de la forme :

$$\lambda(\boldsymbol{\theta}'_0) = S(V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}_0), V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}'_0)) + \text{constante}$$

adaptant le processus de décision aux conditions de prise de vue.

Nous montrons dans la suite comment peut être estimée la dispersion intra-classe

$$D(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}_0) = S(V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}_0), V(\boldsymbol{\theta}'_0 \mid \boldsymbol{\theta}'_0)). \quad (1)$$

3.2 Caractérisation de la dispersion intra classe à partir d'une étude empirique

Nous nous plaçons dans un cadre où les conditions de prises de vues sont maîtrisées pour modéliser empiriquement la dispersion intra-classe (1). Pour cela nous utilisons des séquences d'images prises sur une table tournante (figure 3). La calibration des paramètres extrinsèques de

FIG. 3 – Exemples d'acquisitions de modèles sur la table tournante. Seule l'orientation des véhicules par rapport à la caméra varie dans les séquences.

la caméra est calculée pour chaque image de la séquence par le suivi d'une mire. On obtient ainsi une collection d'images de véhicules à orientations variées et connues. Les conditions d'illumination ne sont pas contraintes. Le score de similarité S (1) utilisé est de la forme : somme de la différence au carré (SSD)

3.3 Estimation de la dispersion D à partir de données de références

Dans cette section, nous supposons qu'une fonction de dispersion D_{ref} a été mesurée pour un véhicule de référence. Le but est alors de déduire D de cette mesure D_{ref} pour les autres véhicules de la base. Les valeurs des indices de dispersion pour divers véhicules et une même présentation d'apprentissage sont représentées sur la figure 4. On note une structure des profils globalement cohérente avec les changements d'aspect - apparition/disparition des faces. Les valeurs numériques des profils dépendent quant à elles de l'apparence des véhicules. On peut néanmoins remarquer qu'une simple pondération de D_{ref} donne une bonne estimation de D. La pondération est obtenue en étudiant l'évolution locale de la courbe autour de la pose $\boldsymbol{\theta}_0$. Si l'observation présente des changements d'aspect suffisamment importants autour de $\boldsymbol{\theta}_0$ (typiquement 10° autour de $\boldsymbol{\theta}_0$) la pondération est calculée par ajustement d'une fonction du second degré aux données. Dans le cas où les changements d'aspect sont trop faibles (par exemple lorsque le véhicule se déplace en ligne droite), la pondération est estimée en simulant un changement d'aspect par un calcul de perturbation sur la projection du modèle texturé autour de $\boldsymbol{\theta}_0$.

3.4 Estimation de la dispersion D à partir d'un ensemble d'informations limité

Dans cette section nous supposons que les séquences présentent de faibles variations d'aspect. Avec cet ensemble d'informations limité il n'est plus possible de créer un profil de référence. Nous proposons alors une approche en trois étapes :

(i) construction et pose du modèle 3d ;

(ii) modélisation du profil de distribution à partir de considérations géométriques sur le modèle projeté ;

(iii) mise à l'échelle de ce profil en utilisant les informations locales de la même manière que dans la section 3.3 ;

Le principal avantage de cette méthode est d'estimer le profil de la dispersion intra-classe sans qu'il soit nécessaire d'effectuer des acquisitions préliminaires. Pour détailler la

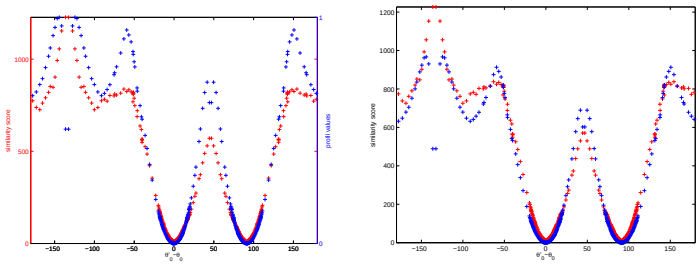

FIG. 5 – *A gauche :* comparaison de la dispersion D pour un modèle ayant pour orientation $\boldsymbol{\theta}_0 = 45°$ (courbe rouge) et le profil normalisé (courbe bleue). *A droite :* comparaison de la dispersion D et du profil mis à l'échelle.

seconde étape, il faut revenir au processus d'extrapolation $V(\boldsymbol{\theta}'_0|\boldsymbol{\theta}_0)$. Pour chaque face i du modèle polyédrique visible dans les poses $\boldsymbol{\theta}_0$ et $\boldsymbol{\theta}'_0$, le calcul de V consiste à projeter tous les pixels de la face dans une pose de référence, puis à calculer la similarité.

Notons $n_i(\boldsymbol{\theta}_0)$ (resp. $n_i(\boldsymbol{\theta}'_0)$) le nombre de pixels de la face i dans la pose $\boldsymbol{\theta}_0$ (resp. $\boldsymbol{\theta}'_0$). Pour minimiser les effets d'aliasing lors de la reprojection, nous choisissons la pose avec la meilleure résolution comme référence (i.e. si $n_i(\boldsymbol{\theta}'_0) > n_i(\boldsymbol{\theta}_0)$, l'extrapolation et le calcul de similarité seront effectués dans la géométrie de la pose $\boldsymbol{\theta}'_0$). Le nombre de pixels intervenant dans le score de similarité est alors $\bar{n}_i = \min(n_i(\boldsymbol{\theta}'_0), n_i(\boldsymbol{\theta}_0))$. Par conséquent l'influence de la similarité d'une face i est mesurée par la proportion de pixels $c_i = \bar{n}_i/\bar{n}$, avec $\bar{n} = \sum_i \bar{n}_i$ étant le nombre total de pixels de comparaison. L'effet de l'erreur d'extrapolation dépend de la différence d'orientation des faces entre les poses $\boldsymbol{\theta}_0$ et $\boldsymbol{\theta}'_0$. Si on note $N_i(\boldsymbol{\theta}_0)$ (resp. $N_i(\boldsymbol{\theta}'_0)$) le vecteur normal à la face i dans la pose $\boldsymbol{\theta}_0$ (resp. $\boldsymbol{\theta}'_0$) l'erreur peut être estimée par une fonction de la forme $(1- < N_i(\boldsymbol{\theta}_0), N_i(\boldsymbol{\theta}'_0) >)$. Ce qui nous donne finalement D en tenant compte de la symétrie que présentent les véhicules :

$$D(\boldsymbol{\theta}'_0, \boldsymbol{\theta}_0) = sum_{i \in faces visibles}(1- < \vec{\mathbf{N}}_i(\boldsymbol{\theta}_0), \vec{\mathbf{N}}_i(\boldsymbol{\theta}'_0) >)c_i \qquad (2)$$

3.5 Résultats expérimentaux

L'algorithme a été évalué sur des séquences prises sur table tournante pour une base de 6 modèles. L'approche de décision par seuillage adaptatif est comparée à une stratégie de seuillage simple (comparaison avec une constante). Chaque séquence contient environ 1000 images. Les performances des algorithmes sont évaluées à partir des courbes P_d/P_{fa} moyennées sur toutes les orientations pour tous les véhicules (à l'exception de celui utilisé pour créer le profil de référence D_{ref}). La figure 6 présente les courbes ROC pour les différentes règles de décision. Les points de fonctionnement $p_d = 1 - p_{fa}$ sont synthétisés dans le tableau ci-dessous. On peut noter que les performance obtenues avec la nouvelle règle de décision surpasse la décision simple, quelle que soit la stratégie choisie. On peut aussi remarquer que la règle de décision obtenue à partir

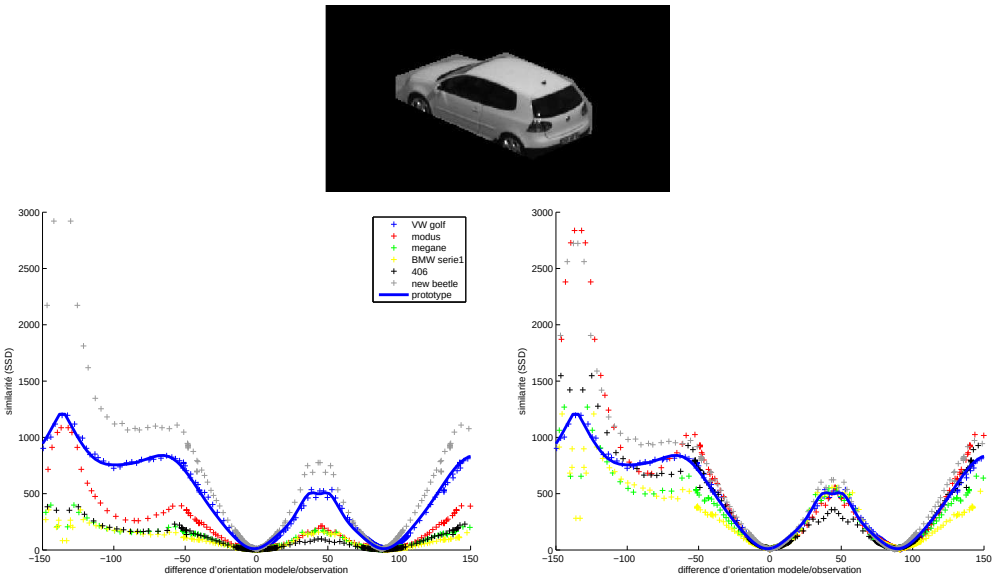

FIG. 4 – Comparaison de la dispersion D pour différents véhicules. Une image du modèle de référence prise à orientation $\boldsymbol{\theta}_0 = 45°$ est présentée sur l'image du haut. La dispersion utilisée comme référence D_{ref} est représentée en traits pleins. Les dispersions sont représentées avant renormalisation (à gauche), et après renormalisation (à droite) par un développement local. Chaque couleur représente un véhicule différent.

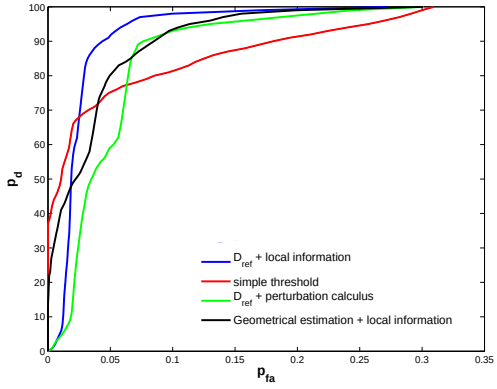

FIG. 6 – Courbes ROC sur les séquences table tournante, la probabilité de bonne détection est représentée en fonction e la probabilité de fausse alarme

de la stratégie purement géométrique donne de meilleurs résultats que celle avec un profil de référence et peu d'information locale dans la séquence requête. Ce qui nous permet d'en déduire que la variation locale de pose dans la séquence a une grande importance dans la description du modèle. Par conséquent nous nous intéresserons au contexte temporel dans la suite de nos travaux.

seuillage simple (courbe rouge)	86.3
D_{ref} + information locale (courbe bleue)	95.1
D_{ref} + calcul de perturbations (courbe verte)	91.3
Estimation géométrique + information locale (courbe noire)	91.1

3.6 Conclusion et perspectives

L'analyse des capacités d'extrapolation d'aspect d'un modèle d'objet nous a permis de définir une règle de seuillage adaptative à partir d'une étude empirique de la variation intra-classe. L'utilisation des courbes de dispersion intra-classe améliore sensiblement les capacités du système de reconnaissance. Dans la suite des travaux nous intégrerons d'autres paramètres de prise de vue (angle d'incidence, résolution) dans la règle de seuillage, et nous intégrerons l'aspect temporel des séquences dans l'algorithme de reconnaissance.

Références

[1] P. Viola et M. Jones, *Rapid object detection using a boosted cascade of simple features*, Computer Vision and Pattern Recognition, 1 (2001), pp. 511–518.

[2] D. Koller, K. Daniilidis et H.-H. Nagel, *Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes*, International Journal of Computer Vision, 10 (3) (1993), pp. 257–281.

[3] V. Blanz et T. Vetter, *A Morphable Model for the Synthesis of 3D Faces*, Dans *Siggraph 1999, Computer Graphics Proceedings*, sous la direction de A. Rockwood, pp. 187–194, Los Angeles, Addison Wesley Longman (1999).

[4] D. Lowe, *Local Feature View Clustering for 3D Object Recognition*, Dans *CVPR01*, pp. I :682–688 (2001).

[5] F. Rothganger, S. Lazebnik, C. Schmid et J. Ponce, *3D Object Modeling and Recognition Using Affine-Invariant Patches and Multi-View Spatial Constraints*, Dans *International Conference on Computer Vision & Pattern Recognition*, vol. 2, pp. 272–277 (2003).

[6] Y. Guo, S. C. Hsu, Y. Shan, H. S. Sawhney et R. Kumar, *Vehicle Fingerprinting for Reacquisition and Tracking in Videos.*, Dans *CVPR (2)*, pp. 761–768 (2005).