

# Détection robuste du plan de la route en milieu urbain

Nicolas Simond, Patrick Rives

► **To cite this version:**

Nicolas Simond, Patrick Rives. Détection robuste du plan de la route en milieu urbain. 14ème Congrès Francophone de Reconnaissance de Formes et Intelligence Artificielle (RFIA'04), Jan 2004, Toulouse, France. inria-00001044

**HAL Id: inria-00001044**

**<https://hal.inria.fr/inria-00001044>**

Submitted on 20 Jan 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Détection robuste du plan de la route en milieu urbain

## Efficient detection of the road plane in urban-like areas

Nicolas Simond, Patrick Rives

Projet ICARE, INRIA

2004 route des Lucioles, 06 902 Sophia-Antipolis Cedex, France

Prénom.Nom@sophia.inria.fr

### Résumé

*Cet article décrit un algorithme de segmentation du plan de la route dans des images de scènes urbaines. Nous supposons que l'environnement contient une majorité de contours alignés selon les 3 directions principales. L'algorithme se décompose en deux étapes. La première détecte conjointement les lignes de fuite qui convergent vers le point de fuite dominant. Les lignes de fuite sont issues d'une détection de contours à laquelle nous appliquons des contraintes spécifiques aux scènes urbaines. Nous recherchons alors des points d'intérêts dans la région de l'image délimitée par les lignes de fuite coplanaires. Lors de la deuxième étape, nous mettons en correspondance l'ensemble des caractéristiques détectées (points d'intérêts et lignes de fuite) en utilisant les propriétés d'invariance projective entre deux prises de vues d'une même scène de façon à estimer l'homographie induite par le plan de la route.*

### Mots Clef

Point de fuite dominant, bi-rapport, homographie, plan de la route, environnement urbain.

### Abstract

*This article describes a vision-based method to segment the road plane in a real road scene. We expect the urban like areas to contain set of edges aligned with the 3 main directions. We assume that a majority of the frontage edges are oriented along the man-made frame directions. We thus detect in a first stage of a two-stage algorithm the vanishing lines and the dominant vanishing point using a contour-based approach and some specific constraints of urban scenes. We detect the feature points in the image area delimited by the coplanar vanishing lines. In the second stage, we both match the feature points and the vanishing lines by introducing the projective invariance properties in order to compute the homography, induced by the road plane.*

### Keywords

Dominant Vanishing Point, Cross-ratio, Homography, Road Plane, Urban Scene.

## 1 Etat de l'art

La navigation autonome d'un véhicule en milieu urbain nécessite un procédé de localisation précis et robuste. Les systèmes GPS ont ainsi été intégré dans la plupart des applications de localisation en extérieur. Cependant, la qualité de la réception dépend de la configuration de satellites visibles durant la mission. Or, les problèmes de masquage dus à l'environnement urbain et les problèmes de trajets multiples rendent de tels systèmes inopérants. Par exemple, Chen [1] indique que la localisation par un système GPS est impossible dans 95% de la métropole de Tokyo. D'autre part, la précision qu'exige la navigation d'un véhicule autonome (quelques cm) ne semble pas actuellement techniquement atteignable avec des coûts compatibles avec une application automobile.

A l'opposé, l'accroissement de la puissance de calcul permet aujourd'hui d'envisager l'utilisation, pour la localisation, de la vision dans un contexte temps réel complémentairement à un système de navigation par GPS. En effet, les structures des façades qui détériorent la qualité de la réception de signaux GPS présentent l'avantage d'être facilement détectables grâce à une caméra. De plus, la structure particulière des scènes urbaines - essentiellement des surfaces planes délimitées par des arêtes orientées selon 3 directions principales orthogonales - pourra être utilisée pour contraindre les algorithmes de vision et accroître leur robustesse.

Cependant malgré leur structuration forte, les scènes urbaines restent des environnements complexes principalement du fait de la présence d'objets ou de personnes en mouvement. Outre la forte dynamique des éléments qui animent les scènes urbaine, l'environnement statique présente aussi de nombreuses difficultés. En ef-

fet, l'environnement varie énormément d'espaces ouverts comme des grands boulevards à d'autres plus restreints comme des ruelles. Une même scène peut contenir des reflets et des jeux d'ombres qui dégradent la qualité des contours des objets structurant les images. C'est pourquoi, il nous a paru opportun de rechercher dans les images des structures caractéristiques sur lesquelles un algorithme de localisation et de navigation pourra s'appuyer. Il est bien évident que la route et, dans une moindre mesure, les façades d'immeubles répondent à ces critères.

Toutes les méthodes de localisation dans des scènes urbaines basées sur la vision reposent sur 2 hypothèses communément admises :

- le sol est localement considéré comme plan,
- l'environnement contient des ensembles de droites ortho-parallèles.

Deux droites parallèles dans l'espace convergent dans une image vers un point appelé point de fuite (Vanishing Point) lorsqu'elles sont vues sous une projection perspective. La détection des points de fuite facilite la compréhension et l'interprétation d'environnements urbains à partir d'une image.

De nombreux auteurs concentrent leurs efforts à améliorer la précision d'estimation des coordonnées des VPs. Il s'agit généralement de méthodes de traitement hors ligne de l'image permettant d'augmenter la précision de localisation des contours. Rother [2] classe les contours détectés par une caméra partiellement calibrée selon 3 directions orthogonales en respectant des contraintes d'orthogonalité et de projection dans l'image des lignes à l'infini. Dans le même esprit, Kosecka and Zhang [3] associent aux techniques robustes de traitement d'images un algorithme EM de façon à calibrer partiellement la caméra et estimer sa position relative par rapport à la scène.

A l'opposé, des auteurs transposent en extérieur les techniques de localisation avec des contraintes temps réel employées dans les applications de robotique mobile d'intérieur. Guerrero and Sagues [4] ont ainsi développé un algorithme de navigation sur route permettant d'estimer l'espace libre devant la caméra en compensant le mouvement de rotation. Lebegue and Aggarwal [5] décrivent un algorithme qui permet de reconstruire automatiquement des environnements de type "couloir".

Snaith et al. [6] travaillent sur un prototype d'aide à la navigation en milieu urbain des personnes mal-voyantes. Ils détectent les encadrements de portes de façon à faciliter leurs franchissements. La direction à suivre est donnée par le point de fuite dominant DVP (Dominant Vanishing Point). Il s'agit de l'intersection de la majorité des droites parallèles horizontales à l'image qui est calculé par le biais d'une transformée de Hough, ap-

pliquée aux contours de l'image. Antone and Teller [7] découplent les mouvements de rotation et de translation entre 2 prises de vues d'une scène urbaine en appliquant un algorithme EM au résultat d'une transformée de Hough appliquée à la segmentation de contours.

Les propriétés géométriques particulières des routes, modélisées par un ruban à bords parallèles, permettent le développement de méthodes spécifiques au niveau de la segmentation d'image. Certains auteurs recherchent les bandes blanches de la signalisation horizontale qui délimitent la chaussée en supposant que celles-ci ne sont pas cachées. Wang et al. [8] détectent les contours naturels de la chaussée en se fondant sur un changement de textures. Sotelo et al. [9] parviennent à isoler la région de la route en identifiant les bords de route grâce à une décomposition HSV de l'image.

Enfin, Okutomi et al. [10] localisent le plan de la route en calculant l'invariant projectif entre 2 images. Cette méthode requiert une connaissance a priori du déplacement réalisé. Hu and Uchimura [11] ont développé un modèle de route structurée à plusieurs voies à partir de clothoïdes de façon à simplifier la corrélation entre la route et sa projection à l'image.

La prochaine section présente les particularités des scènes urbaines pour introduire et justifier la méthode proposée. La section 3 décrit la méthode pour estimer le déplacement de la caméra à partir de 2 vues d'une scène planaire et nous présentons nos résultats dans la section 4. Nous concluons et introduisons nos travaux actuels dans la dernière section.

## 2 Notre approche

Nous développons dans cet article une méthode basée sur la vision pour estimer à terme le déplacement d'un véhicule en milieu urbain. Notre principal objectif est d'extraire un maximum d'informations de la scène à partir de caméras lorsque les données de systèmes utilisant des dispositifs GPS ne sont plus disponibles.

Nous faisons l'hypothèse que la scène observée contient des plans : route, panneaux, murs, façades, etc. Le calcul de l'homographie entre 2 images caractérise alors pleinement le déplacement de la caméra dans l'espace projectif. Nous traitons dans cet article du calcul de telles homographies en considérant le plan privilégié de la route. Nous recherchons pour cela des points et droites coplanaires.

La route qui forme le plan principal des scènes observées apparaît comme une surface homogène où très peu de points d'intérêts (FPs) stables sont détectés quand bien même des défauts du revêtement sont visibles à l'image. De même, les extrémités des lignes blanches du marquage au sol sont aussi peu ou mal détectées. Il n'est donc pas possible d'employer directement une

technique fondée sur les seuls FPs.

Heureusement, les environnements urbains contiennent de nombreuses arêtes 3D. Ces arêtes sont principalement des frontières entre des régions de couleurs, textures, et/ou profondeurs différentes, ainsi que des intersections de surfaces. Les principaux éléments structurant une rue peuvent ainsi être modélisés par un couloir de navigation. Nous portons notre attention sur les éléments du premier plan (gauche, bas et droit de l'image) car leur probabilité d'appartenir au plan de la route ou à des structures qui délimitent ses frontières (trottoir, voiture, façade) est élevée. Les arêtes ayant une extrémité dans l'une de ces régions sont souvent des supports de lignes de fuite VLs (Vanishing Lines) correspondant à des droites parallèles dans la scène.

Nous ferons l'hypothèse que, dans chaque image, les conditions suivantes sont respectées:

- les scènes visualisées contiennent des portions de droites distribuées selon les 3 directions orthogonales principales dont la verticale,
- la route est localement plane, délimitée par des bords parallèles,
- le modèle de caméra est une projection centrale non calibrée,
- la séquence vidéo est enregistrée à cadence élevée, ce qui induit un faible déplacement entre 2 images : la position du DVP évolue lentement.

Cependant, l'ensemble de ces hypothèses restrictives n'est pas toujours vérifié dans le cadre de notre application. Les séquences vidéo enregistrées contiennent une multitude d'objets dynamiques (véhicules, piétons) qui imposent un champ de vision peu ou pas dégagé et introduisent le plus souvent des erreurs de détections.

La projection d'une route dans une image est la plupart du temps caractérisée par une large région plutôt homogène dont les bords sont délimités par une signalisation horizontale et/ou une chaussée. Ces limites sont facilement détectables par leur variations d'intensité (voir Fig.1). Au premier plan, quel que soit le tracé de route (ligne droite ou virage), ces limites peuvent être modélisées par des segments orientés suivant les niveaux de gris. Ces segments se situent aux 2 coins inférieurs de chaque image, ont des orientations opposées et convergent vers le point de fuite dominant (Dominant Vanishing Point), situé dans une région plus haute de l'image.

En milieu urbain, les éléments statiques de la scène sont souvent occultés par des éléments dynamiques. Les segments qui représentent les contours de ces obstacles peuvent avoir une orientation en direction du DVP sans pour autant appartenir à une VL. Nous ne nous contentons donc pas seulement de détecter les bords de la route mais toutes les VLs de l'image qui s'avèrent être des projections de portions de droites 3D parallèles entre

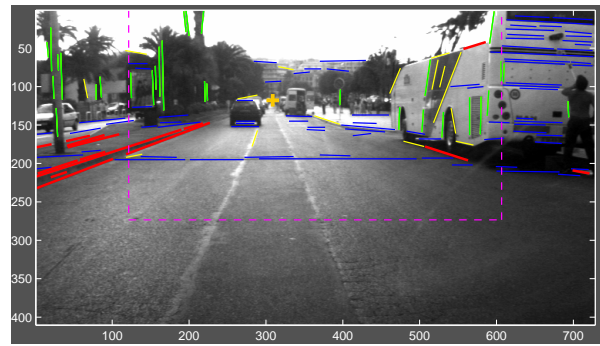


FIG. 1 – Résultats de la détection de contours. Les segments (quasi-)horizontaux et (quasi-)verticaux sont respectivement représentés en bleu et vert. Ils interceptent naturellement tous les autres segments en jaune. Les droites supports des segments en rouge sont assez proches de la position précédente du DVP ('+' orange) : ces segments sont donc considérés comme de potentielles portions de VLs. La région de confiance de l'image, délimitée par des tirets roses, représente le 1<sup>er</sup> plan.

elles. Cela robustifie, d'une part, l'estimation de la localisation du DVP et initie, d'autre part, de prochains travaux de reconstruction de l'environnement.

En tout état de cause, une prise en compte intempes- tive de contours d'un objet dynamique ou d'une ombre en tant que VL n'a que peu de répercussions sur la recherche du DVP tout au long d'une séquence. En effet, soit le contour est dirigé vers le DVP et apparaît comme une portion de ligne de fuite, soit ce n'est pas le cas et l'apparition de tels éléments n'aura lieu d'une part que durant quelques images de la séquence et d'autre part ne peuvent être orientées vers les 2 DVPs de la paire stéréo.

Parallèlement, nous ne tenons pas compte des segments (quasi-)horizontaux car ils ne représentent que très rarement une VL et interceptent naturellement tous les autres segments de l'image, ce qui peut entraîner des erreurs de localisation du DVP. Par contre, 2 cas sont à distinguer pour les segments (quasi-)verticaux. Si ils se situent en dehors des limites du bord de route, il s'agit de projections d'arêtes 3D verticales, sinon il s'agit soit de VLs dans l'axe de visée de la caméra, soit d'arêtes verticales d'un obstacle sur la route.

### 3 Détection et appariement de caractéristiques dans des images de scènes urbaines

#### 3.1 Identification de lignes de fuite

Nous employons un détecteur de Canny afin d'extraire les différents contours de l'image. Ce détecteur a l'avantage de fournir la direction du gradient de niveaux de gris et permet un chainage des contours détectés. Une étape d'approximation polygonale est ensuite réalisée. Le résultat de cette première segmentation est peu fiable. Nos observations montrent que dans la majorité des cas, seul le milieu du segment se situe sur le contour détecté. La précision des coordonnées des extrémités des segments est en effet pixellique. La longueur et l'orientation des segments extraits sont bruitées. Néanmoins, nous remarquons que les segments les plus longs sont les mieux localisés. Nous calculons pour chaque segment  $S_k$  les caractéristiques de sa droite support, paramétrisée par une orientation  $\theta_k$  et une distance à l'origine  $\rho_k$ :

$$\rho_k = \cos\theta_k.u + \sin\theta_k.v$$

où  $(u,v)$  sont les coordonnées en pixel d'un point de l'image.

Le déplacement entre 2 images étant supposé faible, une approximation au premier ordre du déplacement du DVP à l'itération précédente s'avère être une prédiction satisfaisante de la position courante :

$$\widetilde{\mathbf{X}}_{n+1} = \mathbf{X}_n + (\mathbf{X}_{n-1} - \mathbf{X}_{n-2})$$

Nous sélectionnons parmi l'ensemble des contours détectés ceux qui forment les futures VLs. Un premier tri est opéré en ne conservant que les segments dont la distance qui les sépare de la prédiction du DVP est inférieure à un seuil  $d_{\max}$  que nous fixons expérimentalement à 20 pixel. Nous conservons, d'autre part, les segments dont les caractéristiques  $(\rho_k, \theta_k)$  sont proches des VLs retenues pour l'estimation du DVP à l'itération précédente. Nous regroupons récursivement les segments retenus en VLs candidates en appliquant les critères de sélection suivants :

1. les segments les plus longs sont les plus fiables,
2. la direction du gradient des niveau de gris,
3. l'orientation  $\theta_k$ ,
4. la position par rapport à l'horizon,
5. le respect du critère de compatibilité : 2 segments délimitent le même contour à condition qu'il n'y ait pas de recouvrement entre eux (cf Fig.2),
6. la distance entre les extrémités d'un segment et le milieu de celui qui sert de référence est réduite.

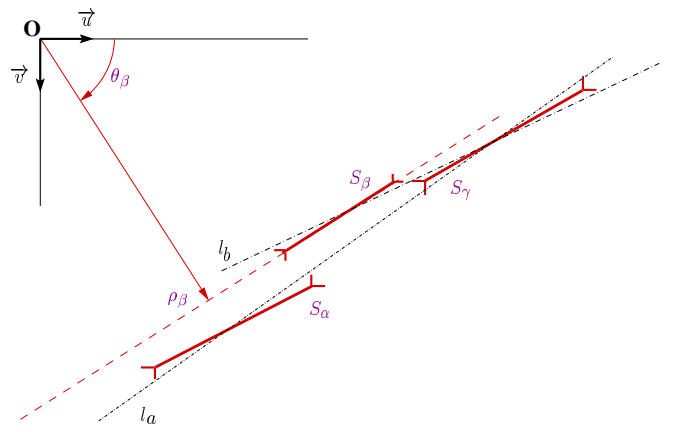


FIG. 2 – Contraintes d'alignement entre segments pour représenter une VL. Les 3 segments  $(S_\alpha, S_\beta, S_\gamma)$  ont des caractéristiques  $(\rho_k, \theta_k)$  similaires, un gradient identique. Les segments  $S_\alpha$  et  $S_\beta$  ayant des coordonnées communes, les 2 VLs candidates sont donc soit  $l_a$  soit  $l_b$ .

Le résultat de cette étape de segmentation des VLs est une liste de  $M$  candidates. Les coordonnées du DVP  $\mathbf{X}(u_X, v_X)$  sont obtenues en résolvant le système d'équations linéaires suivant :

$$\min_x \sum_{m=1}^M (w_m \cdot \mathbf{l}_m^t \cdot \mathbf{X})^2$$

avec  $\mathbf{l}_m^t = [\cos(\theta_m), \sin(\theta_m), -\rho_m]$  et  $\mathbf{X}^t = [u_X, v_X, 1]$

au sens des moindres carrés pondérés, où  $w_m$  est la somme des longueurs des segments.

Après résolution, une étape de vérification est menée sur chaque VLs candidates en calculant le résidu. Seules les candidates ayant un résidu inférieur à un seuil sont conservées pour une nouvelle résolution du système. L'algorithme s'arrête dès que plus aucune VL candidate n'est sélectionnée ou rejetée.

#### 3.2 Faisceaux de lignes de fuite coplanaires

Considérons 2 images  $I_1$  and  $I_2$ , de centre  $C_1$  and  $C_2$ , d'une même scène contenant un ensemble de droites parallèles coplanaires dans la scène. La projection de ces droites sur chacune des 2 images forment un faisceau de VLs convergeant vers un DVP.

Sous l'hypothèse que les droites appartiennent à la partie statique de la scène et que le mouvement du véhicule entre deux images est faible, les caractéristiques des VLs évoluent peu d'une image à l'autre. Nous entreprenons une ébauche de mise en correspondance en comparant entre 2 images les orientations, gradients,

sommes des segments qui les constituent et localisations dans l'image des VLs détectées. Le regroupement de certaines d'entre elles en bandes facilite d'ailleurs cette tâche. Deux VLs consécutives forment une bande si elles sont de gradients opposés et présentent un écart angulaire réduit. Bien que peu précise, cette définition permet d'identifier la plupart des bandes qui forment la signalisation horizontale.

Cependant, peu de séquences à notre disposition présentent des routes où les marquages au sol délimitent de façon sûre les voies de circulation. D'autres droites peuvent être détectées quoique n'appartenant pas au plan de la route (éléments de véhicules, de mobilier urbain, de façades...). Nous identifions les VLs issues du plan de la route en introduisant le bi-rapport (cf [12]). En effet, l'incidence (ordre d'apparition à l'image) et le bi-rapport sont les 2 seuls invariants lors d'une transformation projective. Ainsi, les VLs issues des projections de droites parallèles et coplanaires dans les 2 images  $I_1$  and  $I_2$  sont liées par l'unicité du bi-rapport BR :

$$BR(l_a, l_b, l_c, l_d) = \frac{\sin(\alpha_1)}{\sin(\alpha_2)} / \frac{\sin(\alpha_3)}{\sin(\alpha_4)}$$

où les  $\alpha_i$  sont des angles orientés et signés avec  $i \in [1, 2, 3, 4]$ . La sensibilité au bruit du bi-rapport est généralement un frein à son emploi. En ce qui nous concerne, nous sommes parvenus à réduire cet inconvénient en utilisant non pas les coordonnées des VLs extraites mais les coordonnées "vraies" développées au paragraphe 3.1. Lorsque des VLs mises en correspondance présentent un bi-rapport dont le module est inférieur à  $5 \cdot 10^{-3}$ , nous considérons qu'elles sont coplanaires. Néanmoins, il apparaît parfois des erreurs de sélection : des VLs n'appartenant pas au plan de la route vérifient notre critère alors que des VLs sur la route ne sont pas prises en tant que telles. Nous limitons l'importance accordée à ce critère en vérifiant l'erreur de projection entre les orientations des VLs lorsque l'homographie est estimée. Nous avons remarqué qu'un écart entre 2 VLs en correspondance supérieur à  $5^\circ$  est un critère de non-coplanarité fiable.

### 3.3 Identification du plan de la route

Les VLs appartenant au plan de la route se projettent sur le plan image de la caméra à l'aide d'une transformation particulière, appelée homographie planaire (induite par un plan). Une homographie est définie par une matrice  $H$  [3\*3] et présente 8 entrées : 9-1 de facteur d'échelle. Les coordonnées des objets sont en effet référencées par rapport à une origine arbitraire de chacun des 2 plans. L'homographie contient les mouvements de translation et de rotation au facteur d'échelle près entre 2 plans. Nous pouvons ainsi calculer le déplacement réalisé entre

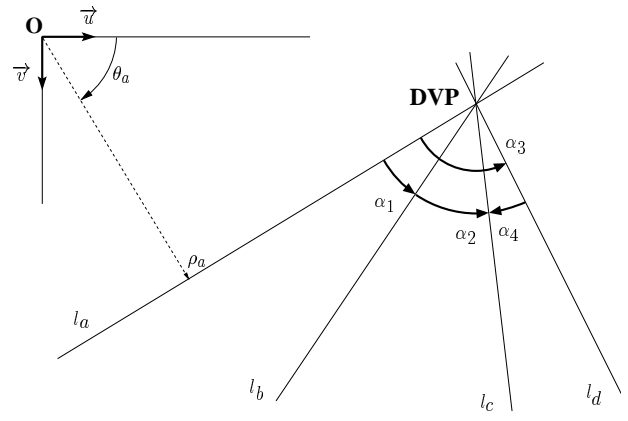


FIG. 3 – Bi-rapport de 4 VLs ( $l_a, l_b, l_c, l_d$ ), projections sur le plan image de 4 droites coplanaires et parallèles entre elles.

$C_1$  et  $C_2$ , centre optiques des images  $I_1$  et  $I_2$ , en introduisant leurs homographies respectives  $H_1$  et  $H_2$  par rapport au plan de référence de la route :

$$l_0 \propto H_1^t \cdot l_1, l_0 \propto H_2^t \cdot l_2 \Rightarrow l_1 \propto [H_1^{-t} \cdot H_2^t] \cdot l_2$$

$$l_1 \propto H^t \cdot l_2 \text{ avec } H \propto H_2 \cdot H_1^{-1}$$

Cependant, l'estimation de l'homographie est impossible à partir d'un faisceau de VLs en raison de leur intersection commune. En effet, les coordonnées des VLs peuvent être paramétrisées à partir d'un couple d'entre elles :  $l_\lambda = l_a + \lambda \cdot l_b$ . Pour mener à terme le calcul, nous devons détecter 2 nouveaux éléments qui appartiennent au plan de la route. En raison du champ de vision limité de la caméra et de l'environnement dynamique, nous ne pouvons détecter de nouvelles VLs. Nous recherchons donc des points d'intérêts (FPs) sur la route qui vérifient les homographies précédentes (voir Fig.4) :

$$p_1 \propto H_1 \cdot p_0, p_2 \propto H_2 \cdot p_0 \Rightarrow p_2 \propto [H_2 \cdot H_1^{-1}] \cdot p_1$$

$$p_2 \propto H \cdot p_1$$

### 3.4 Points d'intérêt coplanaires

Dès que les limites du bord de la route sont détectées, nous segmentons l'image en 2 régions : celle où se situe le plan de la route et le reste. Nous employons la même image de gradients selon les 2 axes, générée par le détecteur de Canny, pour entreprendre la détection de FPs à partir d'un détecteur de Harris, conformément à Schmidt[14]. Seuls les FPs qui ont un score de Harris supérieur à 95% du meilleur score obtenu sur la région de l'image qui nous intéresse sont considérés comme

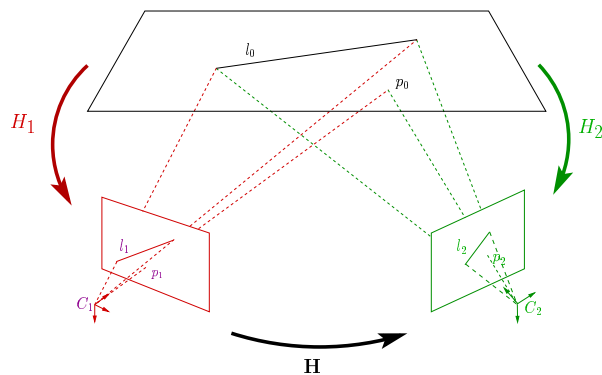


FIG. 4 – L’homographie planaire  $H$  qui lie l’image  $I_1$  à l’image  $I_2$  est obtenue par composition des homographies planaires  $H_1$  et  $H_2$  entre le plan de référence et les plans images. Les droites  $l_1$ ,  $l_2$  et les points  $p_1$ ,  $p_2$  sont les projections de la droite  $l_0$  et du point  $p_0$  sur les plans images des caméras de centre  $C_1$  et  $C_2$ .

correctement détectés. L’étape de mise en correspondance des FPs de 2 images que nous utilisons est fondée sur la décomposition en valeurs propres d’une matrice d’appariement appropriée cf (Pilu[15]). L’estimation de l’homographie entre les images gauche et droite se ramène en effet à la résolution d’un système linéaire entre les coordonnées des FPs et VLs en correspondance et dont l’inconnue est la matrice d’homographie sous forme de colonne. Une normalisation des coordonnées des FPs et des VLs s’impose en raison de la disparité qui existent entre les 2 jeux de caractéristiques. La résolution du système emploie une décomposition en valeurs singulières.

Dès que l’estimation de l’homographie est validée, nous projetons l’ensemble des caractéristiques (DVP, VLs et FPs) de l’image de référence vers la courante de façon à identifier de nouveaux et de faux appariements. Nous calculons pour cela la distance et l’écart angulaire qui séparent respectivement les coordonnées des FPs et des VLs projetés de leurs correspondants.

L’initialisation de l’ensemble de l’algorithme ne nécessite aucune intervention ou connaissance a priori de la scène. Nous avons en effet implémenté une fonction basée sur l’estimateur robuste Ransac conformément à [13]. Cette fonction est appelée chaque fois que l’estimation de l’homographie obtenue à l’itération précédente n’est pas pleinement satisfaisante avec un faible conditionnement de la matrice du système linéaire à résoudre.

Nous supposons en effet que parmi l’ensemble des couples de FPs mis en correspondance entre les VLs coplanaires, la majorité d’entre eux appartiennent effectivement au plan de la route. D’autre part, lorsque l’estimation de

l’homographie à l’itération précédente est satisfaisante, nous l’employons pour accélérer l’étape de mise en correspondance des FPs de l’itération courante. Nous ne retenons dans un premier temps que les couples de FPs compatibles (variations entre ordonnées inférieure à 20 pixels), dont l’erreur de reprojection est minimale. L’estimation de l’homographie entre les images gauche et droite est en effet conjointe à la phase de sélection itérative de couples de FPs que nous agrégeons au fur et à mesure tout en vérifiant que les projections des VLs en correspondance soient vérifiées.

## 4 Résultats

Nous avons enregistré plusieurs séquences vidéo de résolution  $728 \times 440$  pixels, dans les rues du Vieil-Antibes et les abords du port. Nous avons utilisé une paire de caméras stéréo non-calibrées, fixées au chassis du véhicule derrière le pare-brise. Les acquisitions ont été enregistrées à 25 images/s avec une vitesse du véhicule d’environ 10 m/s.

Les expérimentations présentées ont toutes pour origine une séquence test urbaine dans des conditions de trafic favorables: la route est dégagée, très structurée (signalisation horizontale abondante) avec un revêtement homogène. Nous avons pu néanmoins vérifier le bon fonctionnement de l’étape critique de la première partie de l’algorithme proposé qui s’avère la détection de VLs et DVPs sur des séquences plus délicates comme des trajectoires en courbes et/ou la proximité d’obstacles qui réduisent le champ de vision.

Bien que la scène soit une route en ligne droite (Fig.5), la trajectoire suivie par les caméras est représentative des performances de l’algorithme. Le véhicule se déplace en ligne droite, ralentit devant un deuxième véhicule en stationnement pour laisser passer un troisième véhicule qui dépasse l’ensemble, puis finalement redémarre en contournant l’obstacle et poursuit son déplacement.

La Fig.6 présente l’évolution de la position du DVP projeté dans les images gauche (rouge) et droite (verte) tout au long de la séquence. L’évolution des abscisses présentent quelques discontinuités en fin de contournement du véhicule en stationnement car durant cette période la région des images où apparaît la route est très réduite, ce qui nous empêche de détecter suffisamment de limites fiables de la chaussée. Entre les échantillons 940 et 950, seules les VLs représentant la bande blanche de gauche sont extraites, ce qui explique de fortes variations des estimations des DVPs. L’écart angulaire entre les 2 VLs est trop réduit pour estimer correctement les coordonnées des DVPs. Le chronogramme des ordonnées est plus significatif de la difficulté de l’estimation des coordonnées du DVP dans de telles scènes. En effet, le véhi-



FIG. 5 – De haut en bas, de gauche à droite, détail du contournement du véhicule en stationnement (échantillons 870 à 1020 des images de la caméra gauche). Le DVP ('+' orange) est la meilleure estimation au sens des moindres carrés pondérés de l'intersection des VLs représentées en turquoise. Le calcul de l'homographie  $H$  a été entrepris avec les VLs surlignées en rouge et les FPs ('+' jaunes) appariés avec leur correspondants de l'image droite lors d'étapes intermédiaires.

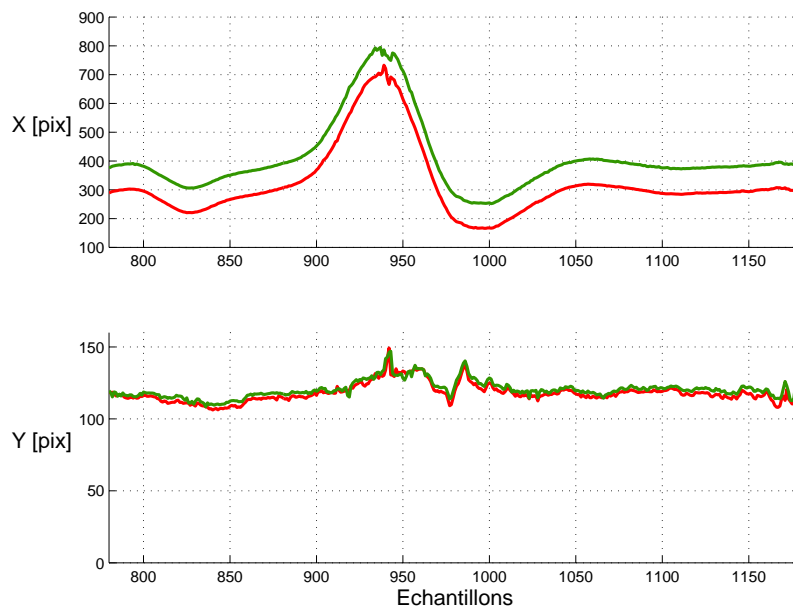


FIG. 6 – Chronogrammes des coordonnées du DVP dans les images gauche (rouge) et droite (vert).



cule se déplaçant sur une route plane horizontale, l'horizon devrait présenter une ordonnée stationnaire. Les variations des ordonnées étant de surcroît corrélées, nous supposons la tendance de ces variations a pour origine le mouvement apparent du plan de la route par rapport au véhicule. Le déplacement vers le haut de l'image des DVPs autour de l'échantillon 975 n'est pas dû à une estimation déficiente mais bien à un mouvement de roulis de la voiture.

Suivant l'algorithme de mise en correspondance des FPs utilisé, les erreurs de reprojections sont limitées à  $\sqrt{5,99} \cdot \sigma$  lorsqu'aucune prédiction de l'homographie n'est disponible,  $\sqrt{2,65} \cdot \sigma$  autrement, avec  $\sigma = 1 \text{ pixel}$  pour les FPs et  $\sigma = 5^\circ$  pour les VLs.

A partir de la calibration hors-ligne de chacune des 2 caméras, nous présentons en Fig.8 les mouvements de rotation et translation estimés entre la caméra gauche à la caméra droite au long de la séquence. Les résultats sont satisfaisants en dehors des échantillons de l'intervalle 930-1030 : les mouvements de rotations et de translation sont en effet quasi-invariants, portés essentiellement par l'axe des abscisses de la caméra gauche, qui sert de référence. Ainsi, la caméra droite s'avère dans le prolongement de la caméra gauche à un peu plus de 50 cm et est légèrement plus orientée vers le sol d'une dizaine de degrés. Les variations de ces transformations observées entre les échantillons 930 et 1030 proviennent d'une mauvaise estimation de l'homographie qui sépare les 2 caméras. La principale cause est le champ de vision plus réduit de la route durant le contournement ce qui induit une quantité plus faible de points d'intérêts détectés. Deuxièmement, la portion de route présente un revêtement plus homogène ce qui induit une répartition des points d'intérêt principalement répartie le long du marquage au sol.

## 5 Conclusion et travaux futurs

Nous avons présenté un algorithme qui extrait le plan de la route d'une paire d'images stéréo en milieu urbain. Nous localisons dans un premier temps le point de fuite dominant de la scène à partir d'une segmentation de contours, résultat d'un filtrage de Canny. Nous recherchons d'abord les lignes de fuite qui s'avèrent les projections dans les plans images de toutes les droites 3D parallèles au bord de la route grâce aux connaissances a priori de la scène. Les images de ces droites ont la particularité de converger vers le point de fuite dominant. Nous extrayons alors parmi le faisceau de lignes de fuite celles qui sont coplanaires en vérifiant l'unicité de leur bi-rapport entre les 2 images.

La région de l'image s'apparentant au plan de la route est ainsi délimitée par les 2 lignes de fuites extrêmes coplanaires. Nous relevons dans cette région des points

d'intérêt à partir d'un détecteur de Harris. Nous apparions les points d'intérêt détectés en employant une méthode basée sur la décomposition en valeurs singulières d'une matrice d'appariement appropriée. Les points d'intérêts coplanaires permettent l'estimation de l'homographie liant les images gauche et droite de la paire stéréo.

La prochaine étape de notre étude est l'exploitation de l'ensemble des caractéristiques coplanaires ainsi détectées pour estimer le déplacement réalisé entre deux prises de vues et retrouver les plans principaux de la scène.

## Références

- [1] T. Chen, *Development of a Vision-based Positioning System for High Density Area*, Proc. of Asian Conference on Remote Sensing, Hong Kong, China, Nov 22-25 1999.
- [2] C. Rother, *A new approach to vanishing point detection in architectural environments*, Proc. of 11th British Machine Vision Conference, pp. 647-655, Bristol, UK, Sept. 11-14, 2000.
- [3] J. Kosecka, W. Zhang, *Efficient Computation of Vanishing Points*, Proc. of IEEE ICRA'02, pp. 3321-3327, Washington DC, May 2002.
- [4] J.J. Guerrero, C. Sagues, *Uncalibrated vision-based on lines for robot navigation*, Elsevier, Mechatronics 11, no. 6, pp. 759-777, Sept. 2001.
- [5] X. Lebegue, J.K. Aggarwal, *Generation of Architectural CAD Models Using a mobile Robot*, Proc. of IEEE ICRA'94, pp. 711-717, San-Diego, CA, USA, 8-13 May 1994.
- [6] M. Snaith, D. Lee, P. Probert, , *Image and Vision Computing Journal*, Volume 16, No 4, pp. 251-263, 1998.
- [7] M. E. Antone, S. Teller, *Automatic Recovery of Relative Camera Rotations for Urban Scenes*, Proc. of IEEE CVPR'00, pp. 282-289, Head Island, SC, USA, 13-15 June 2000.
- [8] R. Wang, Y. Xu, Y. Zhao, *A Vision-based Road Edge Detection Algorithm*, Proc. of IEEE Workshop of Applications of Computer Vision, pp. 237-241, Orlando, Florida, USA, Dec. 3-4, 2002.
- [9] M. A. Sotelo, F.J. Rodriguez, L. Magdalena, *vision-based Navigation System for Autonomous Urban Transport Vehicles in Outdoor environments*, Proc. of IV'02, Versailles, France, June 17-21, 2002.
- [10] M. Okutomi, K. Nakano, J. Maruyama, T. Hara, *Robust Estimation of Planar Regions for Visual Navigation Using sequential Stereo Images*, Proc. of IEEE ICRA'02, pp. 3321-3327, Washington DC, May 2002.

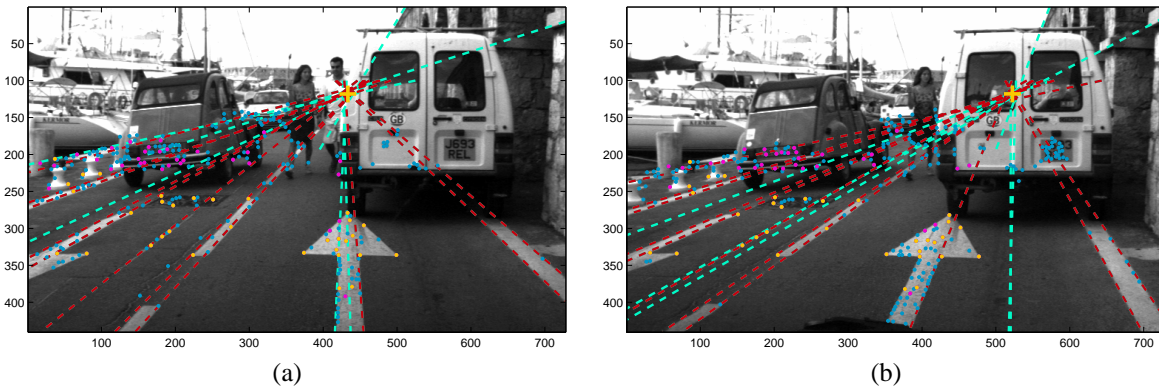


FIG. 7 – Résultats de l'estimation de l'homographie entre les images gauche (a) et droite (b) de la paire stéréo. Les VLs qui ont permis le calcul des coordonnées des DVPs ('+' jaune) sont représentées en turquoise. Les FPs détectés apparaissent en bleu. Ceux que l'algorithme parvient à apparier sont représentés en rose. Nous avons surligné en rouge les VLs et en jaune les FPs qui ont été employés pour estimer l'homographie.

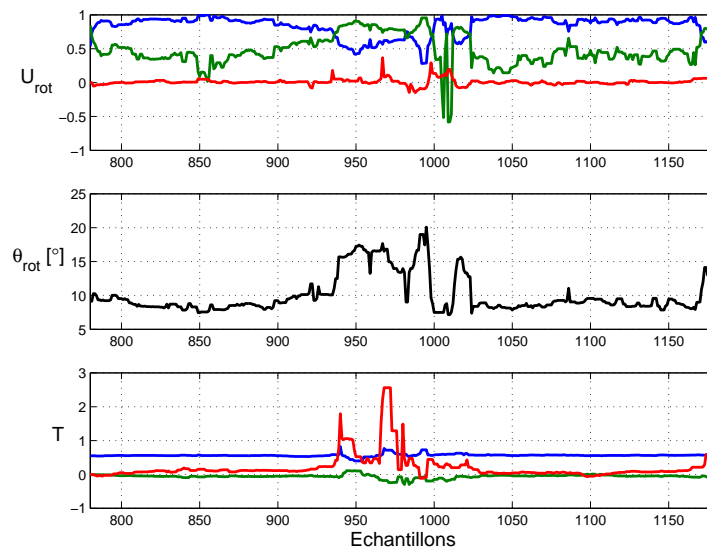


FIG. 8 – Géométrie de la paire stéréo : axes Abs. (bleu), Ord (vert), Opt. (rouge). En haut, les coordonnées du vecteur rotation normalisé, principalement porté par l'axe des abscisses, au milieu, l'angle de rotation exprimé en degrés, en bas, les coordonnées du vecteur translation, exprimées en mètres, essentiellement porté par l'axe des abscisses.

- [11] Z. Hu, K. Uchimura, *Dynamical Road Modeling and Matching for Direct Visual Navigation*, Proc. of IEEE Workshop of Applications of Computer Vision, pp. 237-241, Orlando, Florida, USA, Dec. 3-4, 2002.
- [12] L. Morin, R. Mohr, E. Grosso *Positionnement relatif à partir d'invariants projectifs*, Proc. of RFIA'91, IEEE Reconnaissance de Formes et Intelligence Artificielle, pp. 311-321, Lyon, France, 1991.
- [13] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, pp. 101-112, GB, 2000.
- [14] C. Schmidt, *Appariement d'images par invariants locaux de niveaux de gris*, Thèse, Grenoble, France, Juil 1996.
- [15] M. Pilu, *A direct method for stereo correspondence based on singular value decomposition*, Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 261-266, San Juan, Puerto Rico, June 17-19, 1997.