Reconnaissance d'objets volumiques par mise en correspondance d'indices visuels

Object recognition: solution of the simultaneous pose and correspondence problem

par Frédéric JURIE

LASMEA – UMR 6602 du CNRS, Campus Universitaire des Cézeaux, 63177 Aubière Cedex

résumé et mots clés

Nous nous intéressons à la reconnaissance d'objets volumiques par mise en correspondance d'indices visuels. Nous supposons que les objets à reconnaître sont représentés à l'aide de modèles tridimensionnels, composés d'indices visuels. Reconnaître un objet signifie, dans ce cas, mettre en correspondance les indices du modèle de cet objet avec des indices extraits de l'image, de manière à ce que ces derniers puissent s'expliquer comme une transformation géométrique des indices du modèle. La recherche de la pose (valeur des paramètres de la transformation alignant le modèle sur l'image) et la recherche des correspondances sont ici traitées simultanément. Cela constitue l'originalité et la force de la méthode que nous proposons. Nous présentons de nombreux résultats expérimentaux illustrant l'utilisation de notre approche pour la reconnaissance d'objets.

Reconnaissance d'objets, mise en correspondance

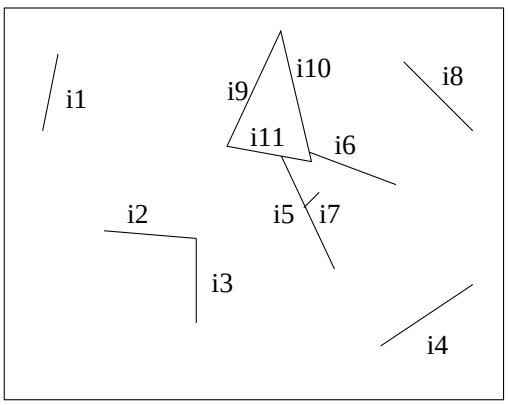
abstract and key words

The use of hypothesis verification is recurrent in the model-based recognition literature. Verification consists in measuring how many model features transformed by a pose coincide with some image features. When data involved in the computation of the pose are noisy, the pose is inaccurate and difficult to verify, especially when the objects are partially occluded. To address this problem, the noise in image features is modeled by a Gaussian distribution. A probabilistic framework allows the evaluation of the probability of a matching, knowing that the pose belongs to a rectangular volume of the pose space. It involves quadratic programming, if the transformation is affine. This matching probability is used in an algorithm computing the best pose. It consists in a recursive multi resolution exploration of the pose space, discarding outliers in the match data while the search is progressing. Numerous experimental results are described. They consist of 2D and 3D recognition experiments using the proposed algorithm.

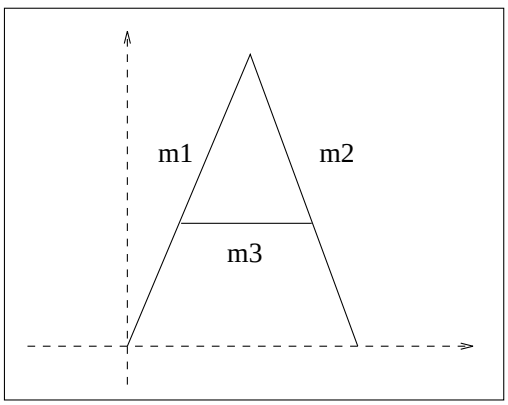
Model-based recognition; pose verification

1. introduction

La reconnaissance d'objets dans une image monoculaire peut être vue comme un problème de mise en correspondance d'*indices visuels*. Supposons que l'image soit segmentée en un ensemble de N *indices visuels* (points, segments, points d'intérêt, etc), et que l'objet à reconnaître soit représenté au moyen de M indices visuels. La reconnaissance de l'objet peut être définie comme la recherche d'un sous-ensemble d'indices de l'image dont l'organisation spatiale est compatible avec l'organisation des indices du modèle. La figure 1 illustre ce problème : comment retrouver, dans l'image, le groupe de segments dont l'organisation est compatible avec celle du modèle ? Le problème se trouve alors repoussé sur la définition de la notion de compatibilité. S'il est possible de définir une transformation géométrique paramétrique du modèle, permettant de transformer les indices du modèle de leur repère au repère de l'image, les indices du modèle sont compatibles avec un sous-ensemble des indices de l'image s'il existe un ensemble de paramètres de la transformation *superposant* ou *alignant* les indices du modèle sur ceux de l'image [25]. La compatibilité peut également s'exprimer intrinsèquement à la forme, sans recourir à l'utilisation d'une transformation géométrique : des propriétés invariantes (présence de textures, de configurations particulières,...) sont déterminées pour le modèle, et recherchées dans l'image. Dans un cas comme dans l'autre, la reconnaissance revient à *mettre en correspondance*, à *apparier*, un sous-ensemble d'indices de l'image avec les indices visuels composant le modèle.



Indices IMAGE



Indices MODELE

Figure 1. – Comment retrouver l'image le groupe de segments dont l'organisation est compatible avec le modèle ?

1.1. combinatoire et stratégies pour la réduire

Ainsi formulée, la reconnaissance d'objets dans des images est un problème combinatoire [1, 3, 11, 8]. Si l'on ne restreint pas le problème, la recherche des sous-ensembles d'indices de l'image compatibles avec ceux du modèle nécessite la génération de tous les appariements possibles entre indices de l'image et indices du modèle, afin de ne garder que ceux qui satisfont un critère de compatibilité (sur lequel nous reviendrons plus tard). Avec les définitions précédentes, $N \times M$ appariements sont possibles, et de l'ordre de $2^{N \times M}$ sous-ensembles doivent être évalués. Ce nombre de sous-ensembles est énorme, même dans des cas simples. La réduction de la combinatoire est donc le problème clé de la reconnaissance d'objets dans des images.

1.1.1. stratégies basées sur la recherche des correspondances

Les stratégies par recherche des meilleures correspondances consistent à créer un arbre de reconnaissance pour lequel chaque feuille constitue un ensemble d'appariements [21, 28]. La figure 2 illustre cette stratégie. L'image comprend N indices, le modèleen comprend M. Le premier indice du modèle peut être mis en

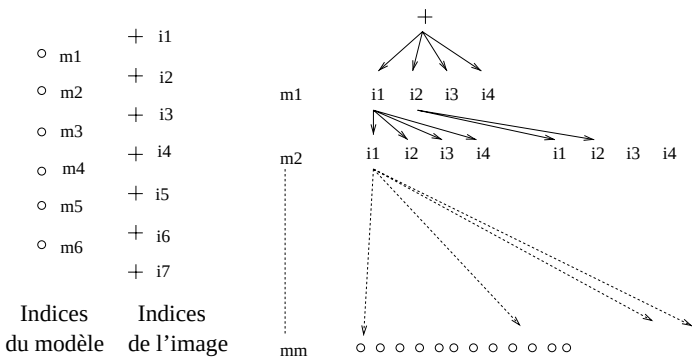


Figure 2. – Reconnaissance par recherche des appariements.

correspondance avec le premier indice de l'image, *etc*. L'ensemble des combinaisons possibles est décrit par un arbre (partie droite de la figure). La combinatoire peut être partiellement réduite en détectant, le plus haut possible dans l'arbre, les groupes d'appariements incohérents, et en coupant la branche. Ces techniques restent cependant coûteuses en temps de calcul.

1.1.2. stratégies basées sur la recherche des meilleures poses

Ces méthodes s'appliquent lorsqu'il existe une transformation géométrique paramétrique, permettant d'expliquer l'apparence visuelle de l'objet dans l'image comme une transformation d'un modèle géométrique. On dénommera, par la suite, les valeurs des paramètres de la transformation alignant le modèle sur l'objet comme la *pose* de l'objet dans l'image. La reconnaissance peut se faire en recherchant les poses alignant le modèle sur des sous-ensembles d'indices de l'image.

Pour cela, chaque appariement d'indice donne une pose ou un ensemble de poses compatibles avec cet appariement. Il est possible, par des techniques d'accumulation [4, 19, 13] (par exemple) de déterminer les poses vérifiant le plus grand nombre d'appariements, et donc alignant au mieux le modèle sur l'image. La figure 3 illustre ce principe : à chaque appariement possible correspond un volume de l'espace des poses.

La technique de prédiction vérification [1] est une alternative intéressante aux techniques d'accumulation. Elle consiste à fixer le nombre minimal d'appariements pour estimer une pose puis à vérifier par transformation du modèle les correspondances des indices restants.

Même si la recherche des meilleures poses est un problème qui est lui aussi combinatoire, l'utilisation de cette stratégie permet d'employer des heuristiques bien plus efficaces que celles applicables aux stratégies de recherche des correspondances.

1.1.3. dualité pose/correspondances - Modèles d'erreur

Nous avons parlé d'alignement des indices du modèle sur ceux de l'image sans réellement définir cette notion d'alignement. Nous supposons que l'alignement est réalisé lorsque la transformation de l'indice du modèle et l'indice de l'image sont compatibles avec un modèle d'erreur [33, 7]. Le modèle d'erreur communément utilisé est le *modèle* d'*erreur borné*, qui impose que la distance entre les deux indices soit inférieure à un seuil. Fixer une pose et un modèle d'erreur revient à fixer un ensemble d'appariements. En fixant un ensemble d'appariements, il est possible d'estimer une pose telle que les appariements vérifient au mieux le modèle d'erreur.

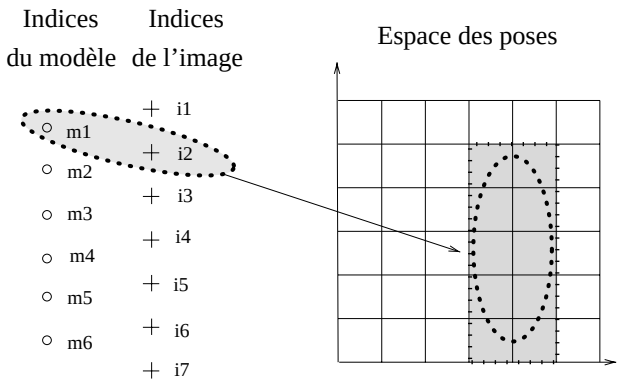


Figure 3. – Reconnaissance par recherche des meilleures poses.

1.2. état de l'art

Malgré de nombreux progrès récents dans le domaine de la vision par ordinateur, la reconnaissance d'objets volumiques dans des scènes complexes demeure un problème clé. L'aspect combinatoire du problème explique en partie la difficulté qu'il y a à progresser.

Cette difficulté peut être dépassée si l'on considère que la positiondes objets dans l'image est connue. Dans ce cas la reconnaissance consiste à comparer efficacement une région de l'image à une collection d'aspects des objets à reconnaître. Des solutions récentes basées sur des analyses en composantes principales [30, 32] ou sur des techniques de *modal matching* [34], ou encore utilisant le *template matching* [9] ont été récemment proposées.

Cependant, comme le souligne Grimson [18] le point difficile de la reconnaissance consiste à séparer, dans l'image, la partie utile (celle qui contient un objet ou des morceaux d'objets) du fond de l'image.

Les différentes composantes d'un système de reconnaissance d'objets ont été l'objet de travaux récents : l'invariance et la pertinence des primitives visuelles utilisées [35], le regroupement d'indices en structures cohérentes [26], l'indexation des modèles [10], l'identification des appariements entre primitives de l'image et primitives des modèles [31], ou les mesures de similarité utilisables pour l'évaluation d'hypothèses [24].

De l'ensemble de ces recherches, on note que la plupart des stratégies performantes peuvent être décrites en terme de stratégie de prédiction-vérification. Dans une phase de pré-traitement les groupes d'indices visuels des modèles ayant des propriétés invariantes sont stockés dans des tables d'indexation. La reconnaissance consiste à regrouper dans un premier temps les indices visuels extraits des images en groupements cohérents. Une phase d'indexation utilisant ces petits groupes de primitives permet ensuite de produire un ensemble d'hypothèses d'appariements, c'est-à-dire un ensemble de mises en correspondance

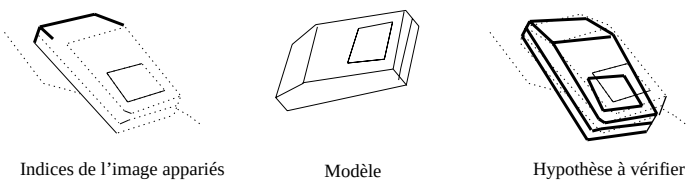


Figure 4. – Vérification d'une pose imprécise.

modèle-image. Les valeurs des paramètres de la transformation géométrique alignant le modèle sur l'image est appelée par la suite *pose* ou *attitude* de l'objet. Les poses obtenues par ces mises en correspondance sont traitées comme des hypothèses qui doivent être vérifiées en appliquant la transformation à l'ensemble des primitives du modèle et en essayant de leur trouver des correspondances dans l'image. Les transformations utilisées sont, dans la plus part des cas, des transformations affines.

Cependant, les erreurs (bruits, imprécisions du détecteur d'indice) introduites dans la mise en correspondance rendent approximatif le calcul de la pose, et la vérification de ces hypothèses devient difficile (comme l'illustre la figure 4).

Comme le soulignent Gandhi et Camps [16], le bruit sur les données se propage sur la pose et diminue la qualité des mises en correspondance produites dans cette étape de vérification.

Cet effet du bruit a largement été étudié, dans le cas de modèles *d'erreurs bornés* (une mise en correspondance est, dans ce cas, considérée valide si la re-projection des indices du modèle se trouve dans un disque autour des indices de l'image). Grimson *et al.* [20] proposent une analyse précise de la mise en correspondance avec incertitude, dans le cas de transformations affines. Ils obtiennent, dans le cas de la mise en correspondance de quatre points l'expression d'invariants affines, dans l'hypothèse où les points appartiennent à des petits disques de l'image. Jacobs [2] a modélisé les erreurs en supposant qu'un point détecté n'est pas à une distance supérieure à ε pixels de l'endroit réel du point. Ainsi, lorsque trois points sont mis en correspondance entre un modèle et une image, ils obtiennent une expression caractérisant l'erreur admissible sur un quatrième point.

Gandhi et Camps [16] ont proposé un algorithme itératif permettant de choisir l'ordre des appariements pour la vérification de l'hypothèse : lorsque la pose à été calculée à partir du nombre minimal de correspondances, les nouveaux appariements permettent de recalculer itérativement la pose. L'ordre d'introduction des appariements est donc primordial. Leur but est de trouver l'ensemble des n primitives (des points dans leur cas) tels que l'effet de l'incertitude sur la pose soit le plus faible possible.

Plutôt que d'utiliser un modèle d'erreur borné, certains auteurs utilisent un modèle d'erreur Gaussien, plus proche de la réalité. En particulier, Sarachik et Grimson [33] estiment la probabilité d'erreur dans la validation d'une mise en correspondance comme une fonction du nombre d'indices (du modèle et de l'image), des occultations, sous l'hypothèse d'un bruit Gaussien sur la position des indices visuels.

Beveridge et Riseman [6] ont proposé un algorithme permettant d'estimer simultanément l'attitude 3D des objets dans l'image, et l'ensemble des mises en correspondances. Leur algorithme est basé sur une exploration aléatoire de l'espace des correspondances, pour trouver l'ensemble de correspondances optimal (celui assurant la meilleur correspondance entre image et projection du modèle 3D dans l'image). L'ensemble des correspondances est itérativement modifié en enlevant ou ajoutant une correspondance.

Malgré l'effort important réalisé par la communauté sur ce sujet, la recherche des meilleurs appariements ou de la meilleure pose demeure un problème difficile.

2. vers une technique efficace pour la mise en correspondance

2.1. introduction

Nous nous intéressons ici à la stratégie de la mise en correspondance d'indices par recherche de la meilleure pose, en raison de sa plus grande efficacité. L'approche que nous proposons ici permet d'obtenir simultanément une attitude précise et un ensemble de mises en correspondance.

Les approches d'alignement mentionnées précédemment consistent à partir d'un ensemble de correspondances initiales et à l'étendre progressivement. L'attitude de l'objet est alors recalculée itérativement en fonction des mises en correspondance ajoutées. Cette solution n'est pas optimale, et nous avons même observé que dans de nombreux cas (particulièrement en cas d'occultation) l'algorithme ne converge pas. Il suffit par exemple d'introduire une mise en correspondance erronée qui perturbe suffisamment l'estimation de l'attitude de l'objet pour que l'ensemble des mises en correspondance suivantes soient également erronées, sans retour en arrière possible.

Nous défendons ici une méthode opposée, que nous avons déjà introduite dans [27]. En supposant que la pose correcte de l'objet appartienne à un domaine connu de l'espace des poses (contenant lapose initiale), l'ensemble des correspondances compatibles avec ce volume est pris en compte. Une correspondance est compatible avec un domaine de l'espace des poses s'il existe une transformation contenue dans ce domaine telle que la projection de l'indice du modèle dans l'image par cette transformation coïncide avec l'indice de l'image. Ce domaine de

l'espace des poses est itérativement réduit, jusqu'à ce qu'il puisse être considéré comme une pose unique. L'algorithme consiste à guider la recherche dans l'espace des poses pour que la pose finale soit optimale au sens des mises en correspondances obtenues. La solution obtenue est bien meilleure que celle obtenue en augmentant progressivement l'ensemble des correspondances, car les influences relatives de chaque mise en correspondance sont prises en compte simultanément.

Cette approche est expérimentalement validée au moyen de deux applications. La première est une application de reconnaissance d'objets basée sur l'apparence : les objets 3D sont représentés par une collection d'aspects 2D. La reconnaissance consiste alors à rechercher des correspondances entre les indices de l'image et les aspects. Une transformation géométrique affine 2D est utilisée dans ce cas. La seconde application consiste à mettre directement en correspondance les indices 3D d'un modèle d'objet avec les indices de l'image. La transformation géométrique utilisée est une projection orthographique avec mise à l'échelle (transformation affine).

2.2. description de l'approche proposée

Comme nous venons de l'indiquer, l'approche proposée consiste à explorer récursivement l'espace des poses, pour trouver celles qui permettent une bonne mise en correspondance des indices d'un modèle avec ceux de l'image.

L'espace des poses initiales peut être tout l'espace des poses possibles, ou peut être restreint si une prédiction de la pose est possible.

L'espace initial est divisé en deux sous-espaces (appelés par la suite *boîtes*). Pour chacune de ces deux boîtes, la probabilité maximale qu'elles puissent contenir une pose telle que le modèle soit vu dans l'image est évaluée. Les boîtes contenant les probabilités les plus hautes sont conservées et divisées à leur tour, récursivement. La division d'une boîte s'arrête lorsque sa taille a atteint un seuil fixé.

Cette probabilité maximale que la boîte puisse contenir une pose telle que le modèle apparaisse dans l'image est vue comme une combinaison de probabilités de mises en correspondances de primitives individuelles (correspondances primitive à primitive), pour une boîte donnée.

La probabilité qu'une primitive du modèle soit vue dans l'image est calculée à partir de la probabilité qu'elle puisse correspondre à une primitive de l'image, pour une pose contenue dans la boîte. De ce calcul de probabilité pour une pose, nous déduisons une probabilité maximale de correspondance pour une boîte donnée. Nous verrons que ces probabilités sont construites à partir d'un modèle d'erreur Gaussien, basé sur une mesure de distance de primitives. L'algorithme peut être décrit par le pseudo-code :

ListeBoites = {BoiteInitiale}
Pour niveau = 0 et tant que (niveau<niveau_max) faire {
Pour chaque B_i **dans ListesBoites faire {**
Diviser B_i **en** B_i^1 **et** B_i^2
$P_i^j = P(M \backslash B_i^j)$ **pour j = [1, 2]**
}
ListeBoites = ensemble des N B_i^j **ayant les** P_i^j **les plus fortes niveau = niveau + 1**
}
où P(M/B) est la probabilité d'avoir le modèle M dans l'image, dans une pose contenue dans la boite B.

La liste suivante résume les différentes notations rencontrées dans l'article, et leur sens.

– $P(C|\mathbf{p})$: probabilité qu'un indice donné du modèle M soit en correspondance avec un indice de l'image, pour une pose $\mathbf{p}$ donnée. Cette probabilité est estimée à partir d'un modèle d'erreur Gaussien portant sur une mesure de distance entre indices.
– $P(C|Box)$: valeur maximale de $P(C|\mathbf{p})$, sachant que $\mathbf{p}$ appartient à la boîte Box. Cette valeur est calculée par une recherche de la valeur maximale de $P(C|\mathbf{p})$ sur Box.
– $P(M|\mathbf{p})$: probabilité d'avoir le modèle M dans l'image sous une pose $\mathbf{p}$. Est calculée en combinant les valeurs de $P(C|\mathbf{p})$ pour chaque indice du modèle.
– $P(C|Box)$: probabilité maximale d'avoir le modèle M dans l'image, en supposant que sa pose soit dans Box. Est calculée en substituant les valeurs de $P(C|\mathbf{p})$ par $P(C|Box)$ dans le calcul de $P(M\mathbf{p})$.

Le plan de cet article est le suivant : la section 3 présente en détail la stratégie d'exploration de l'espace des poses. La section suivante explique le calcul de la probabilité d'avoir un modèle dans l'image pour une pose donnée $P(C|\mathbf{p})$ ainsi que la probabilité $P(C|Box)$, tandis que la section 5 donne le calcul de la probabilité d'une correspondance primitive modèle-primitive image pour une pose donnée $P(C|\mathbf{p})$. La section 6 détaille le calcul de la probabilité maximale d'une correspondance de primitive, sachant que la pose doit appartenir à une boîte de poses donnée $P(C|Box)$.

Ensuite, l'article présente l'utilisation de cet algorithme de recherche de la meilleure pose à la reconnaissance d'objets volumiques modélisés par des collections de vues (section 7) et à la reconnaissance d'objets volumiques modélisés au moyen de modèles géométriques 3D (section 8). Des résultats illustrant le suivi d'objets 3D sont également présentés.

Enfin, la section 9 propose une discussion sur la méthode proposée, en la comparant à différents travaux voisins.

3 exploration récursive de l'espace des poses

L'idée est donc d'explorer l'espace des poses en le subdivisant récursivement en boîtes de plus en plus petites. La recherche est guidée par une mesure indiquant s'il est intéressant ou non d'explorer une boîte donnée. Grâce à cette mesure, seules les régions les plus pertinentes seront explorées.

L'exploration consiste à diviser récursivement la boîte en deux, en alternant les axes de découpe, comme illustré figure 5. Ce procédé peut être perçu comme une recherche arborescente. La racine correspond à la boîte initiale. Les feuilles sont les boîtes dont le volume est suffisamment petit pour qu'elles puissent être considérées comme des poses uniques.

Différentes stratégies sont possibles pour parcourir cet arbre. Breuel [7] propose, par exemple, d'explorer d'abord les branches les plus probables puis de faire des retours en arrière afin de vérifier qu'aucune autre branche ne peut donner de meilleur résultat. Le nombre maximal d'opérations, et en conséquence le temps de calcul ne peuvent être garantis. Dans le pire des cas, l'ensemble de l'espace doit être exploré.

C'est pourquoi nous préférons utiliser un algorithme de type « N-search ». Les N meilleures branches sont explorées simultanément, sans retour en arrière possible. Le nombre maximal de branches explorées est donc de Nh où h représente le nombre de niveaux de l'arbre et où N est le nombre de branches explorées simultanément. En pratique, nous avons constaté que cette valeur, fixée empiriquement, peut être relativement faible. Pour toutes nos expériences, une valeur de $N = 5$ s'est montrée suffisante pour obtenir toutes les poses recherchées.

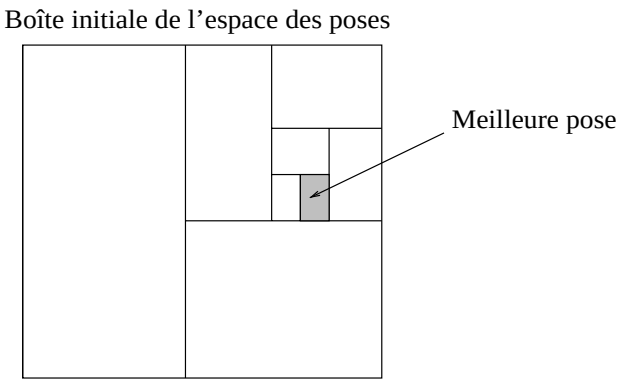


Figure 3. – Division récursive de la boîte initiale (l'espace des poses est représenté, par simplicité, par R^2).

Sélection des boîtes

Les boîtes explorées sont sélectionnées selon la probabilité maximale qu'un modèle soit vu dans l'image, sachant que sa pose doit appartenir à la boîte considérée. Notons $P(M|BOX)$ où M représente le modèle et BOX la boîte considérée. Cette quantité représente une estimation de la valeur valeur maximale que peut atteindre $P(C|_{\mathbf{p}})$, la probabilité d'avoir le modèle avec une pose $\mathbf{p}$ dans l'image, sachant que la pose $\mathbf{p}$ appartient à la boîte de l'espace des poses.

Elle peut être obtenue en évaluant individuellement la probabilité que chaque indice du modèle soit dans l'image, sans garantie qu'une pose unique satisfasse correctement à l'ensemble des indices. Nous avons $\forall \mathbf{p} \in BOX,\ P(M|\mathbf{p}) \leqslant P(M|BOX)$. Il est en effet possible de trouver à l'intérieur de la boîte initiale un ensemble de poses dont chaque pose aligne parfaitement un indice du modèle sur un indice de l'image sans qu'il n'existe aucune pose alignant l'ensemble des indices simultanément.

Nous verrons dans la section suivante que le calcul de la probabilité d'avoir le modèle dans l'image pour une pose donnée ($P(M|\mathbf{p})$) sera obtenu (équation (1)) par combinaison des probabilités de correspondances de chaque primitive du modèle, pour la pose donnée (notées $P(C|\mathbf{p})$). Pour calculer $P(M|BOX)$ nous reprenons le calcul de $P(M|\mathbf{p})$, mais en remplaçant les $P(M|\mathbf{p})$ par des $P(C|BOX) = \max_{\mathbf{p}\in BOX} P(C|\mathbf{p})$.

La valeur $P(M|BOX)$ est riche en information et permet de faire un tri rapide entre l'ensemble des boîtes initiales.

Calcul de la boîte initiale

La boîte initiale doit être suffisamment grande pour compenser les erreurs d'appariement. Si l'on suppose qu'un modèle d'erreur gaussien est utilisé (nous le définirons pas la suite), la boîte initiale devrait avoir une taille infinie.

Pour des raisons pratiques, une définition plus réaliste à été adoptée : la taille de la boîte initiale est telle que il y ait au moins $\mathbf{p}$ chances qu'elle inclue la pose initiale.

Dans toutes nos expérimentations, la boîte initiale, de taille constante est centrée sur la pose initiale.

Si aucune prédiction n'est disponible (cas de la reconnaissance avec modèles d'objets 3D), l'ensemble des poses possibles est pris comme boîte initiale.

4. probabilité qu'un modèle d'objet soit dans l'image, pour une pose donnée, et pour une boîte donnée.

Nous montrons ici comment déduire la probabilité de correspondance d'un modèle d'objet, notée $P(M|\mathbf{p})$, en combinant les probabilités de correspondance de chacun des indices qui le compose.

4.1. calcul de $P(M|\mathbf{p})$

Nous supposons que la probabilité d'occurrence du modèle M dans l'image, connaissant une pose $\mathbf{p}$ (notée $P(M|\mathbf{p})$ ne dépend que de la qualité des appariements des indices qui le composent. Si le modèle comprend $\mathcal{M}$ indices, on dénombre $2^{\mathcal{M}}$ configurations d'appariements possibles, notées γ.

$$P(M|\mathbf{p}) = \sum_{\gamma\in\Gamma} P(M|\mathbf{p},\gamma)P(\gamma|\mathbf{p}).$$

Afin de réduire le nombre de configurations, elles sont regroupées en fonction du nombre d'indices mis en correspondance. Soit $E^k, k \leqslant \mathcal{M}$ le nombre de configurations qui mettent k segments en correspondance. Alors, $E^k = \bigcup_{j=1}^{j<(\mathcal{M})_k} \gamma_j^k$, et $\Gamma = \bigcup_{i=1}^{i\leqslant\mathcal{M}} E^i$ est l'ensemble de toutes les configurations possibles et mutuellement exclusives. $(\mathcal{M})_k$ représente le nombre de k-uplets ordonnés d'indices du modèle. Alors,

$$\begin{aligned} P(M|\mathbf{p}) &= \sum_{\gamma\in\Gamma} P(M|\mathbf{p},\gamma)P(\gamma|_{\mathbf{p}}). \\ &= \sum_{k=1}^{k\leqslant\mathcal{M}} \sum_{j=1}^{j\leqslant(\mathcal{M})_k} P(M|\mathbf{p},\gamma_j^k)P(\gamma_j^k|\mathbf{p}). \end{aligned}$$

Cette formule peut être simplifiée, puisque M et $\mathbf{p}$ sont conditionnellement indépendants pour γ donné :

$$P(M|\mathbf{p}) = \sum_{k=1}^{k\leqslant\mathcal{M}} \sum_{j=1}^{j\leqslant(\mathcal{M})_k} P(M|\gamma_j^k)P(\gamma_j^k|\mathbf{p}).$$

La taille de Γ est trop grande pour que $P(M|\gamma)$ puisse être apprise durant une phase d'apprentissage. Nous avons simplifié cette expression en considérant que le paramètre le plus significatif pour calculer cette probabilité est le nombre d'indices appariés, ainsi que la qualité de ces appariements. Cette hypothèse signifie que le nombre de correspondances prime sur la configuration d'indices mis en correspondance. Elle est communément admise, et utilisée, par exemple, dans toutes les techniques de reconnaissance par accumulation.
Cela revient à dire :

$$\forall k \in [1\ldots\mathcal{M}],\ \forall i \in [1\ldots(\mathcal{M})_k]\ \forall l \in [1\ldots(\mathcal{M})_k]$$
$$P(M|\gamma_i^k) = P(M|\gamma_l^k) = P(M|E^k).$$

La probabilité $P(M|\mathbf{p})$ peut donc être écrite :

$$P(M|\mathbf{p}) = \sum_{k=1}^{k\leqslant\mathcal{M}} \left(P(M|E^k) \sum_{j=1}^{j\leqslant(\mathcal{M})_k} P(\gamma_j^k|\mathbf{p}) \right). \quad (1)$$

$P(M|E^k)$ est la probabilité du modèle M sachant que k de ses indices sont mis en correspondance.
Cette probabilité a été estimée dans une phase d'apprentissage : des modèles d'objets ont été localisés dans différentes images présentant des occultations (à partir d'appariements initiaux générés aléatoirement), et nous avons dénombré le nombre de localisations correctes/incorrectes pour un nombre donné de correspondances. Une table donnant $P(M|E^k)$ en fonction de k a ainsi pu être dressée.
Le calcul de $P(E^k|\mathbf{p}) = \sum_{j=1}^{j\leqslant(\mathcal{M})_k} P(\gamma_j^k|\mathbf{p})$ est plus délicat. L'évènement E^k est fonction de l'union de $(\mathcal{M})_k$ configurations distinctes notées γ_j^k. La probabilité $P(\gamma_j^k|_{\mathbf{p}})$ de chacune de ces configurations peut être écrite comme fonction des appariements,

$$P(\gamma_j^k|\mathbf{p}) = \prod_{i=1}^{i\leqslant\mathcal{M}} P(m_i \xrightarrow{b(i)} |\mathbf{p}),$$

où $b(i)$ est une variable booléenne signifiant que l'indice m_i du modèle est (ou n'est pas) mis en correspondance dans cette combinaison. $m_i \xrightarrow{1}$ signifie m_i est apparié, $m_i \xrightarrow{0}$ signifie m_i n'est pas apparié. Si l'on suppose que $m_i \xrightarrow{b(i)}$, $i \in \{1,\ldots,\mathcal{M}\}$ sont des évènements indépendants, nous avons $P(\gamma_j^k|\mathbf{p}) = \prod_{i=1}^{i\leqslant\mathcal{M}} P(m_i \xrightarrow{b(i)} |\mathbf{p})$.

Dans ce cas, $P(E^k|\mathbf{p})$ est une somme de $2^{\mathcal{M}}$ termes, chaque segment pouvant en effet être considéré comme apparié où non, c'est-à-dire dans deux états possibles. Nous proposons de l'approximer en ne gardant, pour chaque valeur de k, que la combinaison maximale ; cela revient à prendre en compte la combinaison pour laquelle les k meilleurs appariements sont supposés être réalisés, les $\mathcal{M}-k$ autres appariements possibles étant supposés ne pas être réalisés. La somme ne comprend plus 2^k termes mais $\mathcal{M}$ termes (un pour chaque valeur de k).
Les expériences présentées dans cet article, dans lesquelles $\mathcal{M} < 10$ (par exemple cas des modèles 3D) ont été testées avec les valeurs exactes $P(E|_{\mathbf{p}})$ et l'approximation proposée. Nous avons obtenu, dans les deux cas, exactement les mêmes réponses.

Exemple simple. Nous présentons, dans un but d'illustration, un exemple très simple. Supposons qu'un modèle possède quatre indices notés m_1, m_2, m_3, et m_4. Pour la pose $\mathbf{p}$, la probabilité que chacun de ses indices soit apparié est supposée connue. Supposons que les valeurs numériques soient

$P(m_1 \xrightarrow{1} |_{\mathbf{p}}) = .7$ $P(m_2 \xrightarrow{1} |_{\mathbf{p}}) = .3$, $\quad P(m_3 \xrightarrow{1} |_{\mathbf{p}}) = .2$, $P(m_4 \xrightarrow{1} |_{\mathbf{p}}) = .9$.

Un, deux, trois ou quatre segments du modèle peuvent être appariés ; ces quatre cas sont dénommés : E_1, E_2, E_3, et E_4. Si l'on

suppose que

$P(M|E_1) = .1$, $P(M|E_2) = .5$, $P(M|E_3) = .8$, et $P(M|E_4) = 1.$, alors

$$P(M|\mathbf{p}) = \sum_{i=1}^{i\leqslant 4} P(M|E_i)P(E_i|\mathbf{p})$$

$$\begin{aligned}
&\simeq P(M|E_1)P(m_1 \overset{0}{\rightarrow}|\mathbf{p})P(m_2 \overset{0}{\rightarrow}|\mathbf{p})P(m_3 \overset{0}{\rightarrow}|\mathbf{p})P(m_4 \overset{1}{\rightarrow}|\mathbf{p})\\
&+ P(M|E_2)P(m_1 \overset{1}{\rightarrow}|\mathbf{p})P(m_2 \overset{0}{\rightarrow}|\mathbf{p})P(m_3 \overset{0}{\rightarrow}|\mathbf{p})P(m_4 \overset{1}{\rightarrow}|\mathbf{p})\\
&+ P(M|E_3)P(m_1 \overset{1}{\rightarrow}|\mathbf{p})P(m_2 \overset{1}{\rightarrow}|\mathbf{p})P(m_3 \overset{0}{\rightarrow}|\mathbf{p})P(m_4 \overset{1}{\rightarrow}|\mathbf{p})\\
&+ P(M|E_4)P(m_1 \overset{1}{\rightarrow}|\mathbf{p})P(m_2 \overset{1}{\rightarrow}|\mathbf{p})P(m_3 \overset{1}{\rightarrow}|\mathbf{p})P(m_4 \overset{1}{\rightarrow}|\mathbf{p})\\
&\simeq .01 + .17 + .12 + .03 = .34.
\end{aligned}$$

4.2. appariements distincts

Un même indice du modèle peut être mis en correspondance avec plus d'un indice de l'image ; de même, un indice de l'image peut être associé à plus d'un indice du modèle.

Ce problème a été étudié par plusieurs auteurs, tels que Gavrila et Groen [17] ou Huttenlocher et Cass [23] pour le cas de l'utilisation d'un modèle d'erreur borné. Les solutions proposées consistent à évaluer le nombre de correspondances distinctes, ce qui implique une certaine combinatoire (pour le calcul et l'évaluation de l'ensemble des sous-ensembles de correspondances possibles [23]). Ce nombre peut être approché, en calculant le nombre d'indices distincts du modèle appariés, le nombre d'indices images distincts appariés, et en prenant le minimum de ces deux valeurs.

Ces auteurs signalent que ce critère donne une surévaluation du résultat, par rapport au résultat obtenu en utilisant le critère du nombre de correspondances distinctes. Mais ils ajoutent que le gain en calculs compense largement cette surévaluation.

Nous n'utilisons pas ici ce critère, puisque nous ne recherchons pas le nombre de correspondances distinctes mais la probabilité qu'un segment donné du modèle soit apparié.

En notant $P(m \rightarrow |\mathbf{p})$ la probabilité quel'indice $\mathbf{f}$ du modèle soit apparié avec un indice de l'image connaissant la pose $\mathbf{p}$, et $P(_{\mathbf{f}} \rightarrow {}_{\mathbf{f}_d}|\mathbf{p})$ la probabilité que « l'indice modèle $\mathbf{f}$ soit apparié avec l'indice image $\mathbf{f}_d$ connaissant la pose $\mathbf{p}$ », nous avons directement

$$P(\mathbf{f} \rightarrow |_{\mathbf{p}}) = P(\cup_{j=1}^{j<\mathcal{N}} {}_{\mathbf{f}} \rightarrow {}_{\mathbf{f}dj}|\mathbf{p}),$$

où $\mathcal{N}$ est le nombre d'indices de l'image. Cette probabilité est facilement calculable, si l'on suppose que les appariements sont indépendants. Par exemple, si un indice du modèle peut être apparié avec deux segments de l'image avec les probabilités 0.7 et 0.8, la probabilité que cet indice du modèle soit apparié est de $0.7 + 0.8 - 0.7{}^*0.8 = .94$.

4.3. calcul de $P(M|BOX)$

Comme nous l'avons expliqué précédemment, $P(M|BOX)$ (où M représente le modèle et BOX la boîte considérée) représente une estimation de la valeur valeur maximale que peut atteindre $P(M|\mathbf{p})$ pour $\mathbf{p} \in BOX$.

Nous le calculons en calculant la probabilité que chaque indice du modèle soit dans l'image, sans garantie qu'une pose unique satisfasse correctement à l'ensemble des indices. Il ne s'agit que d'une estimation du miximum de probabilité, puisqu'il est possible de trouver à l'intérieur de la boîte initiale un ensemble de poses dont chaque pose aligne parfaitement un indice du modèle sur un indice de l'image sans qu'il n'existe aucune pose alignant l'ensemble des indices simultanément.

Pour calculer $P(M|BOX)$ nous reprenons le calcul de $P(M|\mathbf{p})$ donné équation (1)), mais en remplaçant les $P(C|\mathbf{p})$ par des $P(C|BOX) = \max_{\mathbf{p}\in BOX} P(C|\mathbf{p})$.

Le calcul de $\max_{\mathbf{p}\in BOX} P(C|\mathbf{p})$ est donné section 6.

5. probabilité d'un appariement primitive modèle – primitive image connaissant une pose $\mathbf{p}$ $P(M|\mathbf{p})$

Avant de donner à proprement parler le calcul de $P(M|_{\mathbf{p}})$, nous allons définir certaines notions préalables, en particulier indiquer comment les indices des modèles peuvent être ramenés dans l'image pour une pose donnée (section 5.1). Nous verrons également que la qualité d'une correspondance est évaluée au moyen d'un modèle d'erreur Gaussien, défini section 5.2. Le calcul de $P(M|\mathbf{p})$ est enfin présenté (section 5.3).

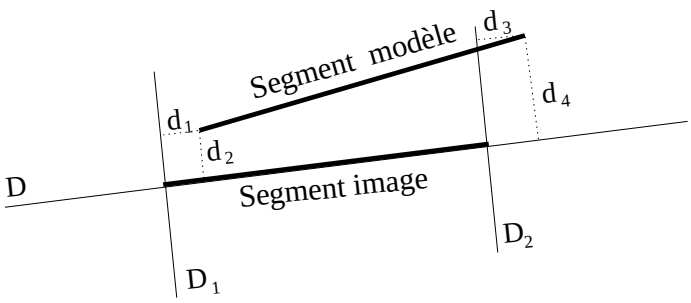


Figure 6. – Exemple de la mesure de distance entre segments.

5.1. transformation d'indices visuels

Notons $\mathbf{f}_d$ un indice extrait de l'image et $\mathbf{f}_t$ la transformation de l'indice $\mathbf{f}$ du modèle par la transformation affine T. La transformation affine de $\mathbf{f}$ peut être mise sous la forme :

$$\mathbf{f}_t = T(\mathbf{f}, \mathbf{p}) = \mathbf{F} \cdot \mathbf{p},$$

où la pose de l'objet est représentée par $\mathbf{p}$ et $\mathbf{F}$ est une matrice construite à partir de $\mathbf{f}$.

Exemple dans le cas bidimensionnel Un indice de type point et la projection du modèle sont définis par :

$$\mathbf{f} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \mathbf{f}_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & x & -y \\ 0 & 1 & y & x \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} t_x \\ t_y \\ sI_x \\ sI_y \end{pmatrix}.$$

Les coordonnées du point modèle (x, y) sont transformées en coordonnées images (x_t, y_t), en utilisant la pose $\mathbf{p}$. Cette transformation réalise une translation planaire t_x, t_y, une rotation planaire θ, et une mise à l'échelle s, où $I_x = cos(\theta)$ et $I_y = sin(\theta)$.

5.2. distribution d'erreur normale

Soit $\delta = \mathbf{f}_d - \mathbf{f}_t$ la différence de position dans l'image entre la primitive image et la transformation de l'indice du modèle.

Les indices extraits des images sont supposés avoir une distribution normale par rapport aux positions réelles des indices. Soit $P(\delta|C)$ la probabilité de δ connaissant C. C signifie « $\mathbf{f}_t$ et $\mathbf{f}_d$ sont en correspondance ». Dans un espace à v-dimensions, la loi de densité décrivant la probabilité s'écrit, en fonction de la matrice de covariance $\mathbf{Q}$ comme

$$P(\delta|C) = (2\pi)^{-\frac{v}{2}} |_{\mathbf{Q}}|^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{f}_t - \mathbf{f}_d)^t \mathbf{Q}^{-1} (\mathbf{f}_t - \mathbf{f}_d))$$

Si les indices sont des points, la dimension de l'espace des indices est 2 et $\mathbf{f} = (x_d, y_d)^t$ sont les coordonnées de l'indice. Dans le cas général, aucune hypothèse n'est faite, ni sur la dimension de l'espace, ni sur la nature des indices visuels. La matrice $\mathbf{Q}$ permet de prendre en compte des incertitudes qui ne sont pas forcements « dirigées » dans le sens des axes des repères. C'est particulièrement intéressant dans le cas de segments de droites, pour lesquels l'incertitude la plus importante, provoquée par le manque de fiabilité de la détection de la terminaison, se trouve dans l'axe du segment. La matrice $\mathbf{Q}$ est calculée en localisant le modèle (au moyen d'appariements manuels) et en mesurant l'erreur entre la projection des primitives et les primitives détectées. La figure 6 illustre la mesure de cette erreur dans le cas de segments.

Dans un but de simplification, les indices ne sont pas représentés dans leur espace de représentation initial. La matrice de covariance $\mathbf{Q}^{-1}$ est décomposée en

$$\mathbf{Q}^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^t,$$

où $\mathbf{U}$ est la matrice orthogonale des vecteurs propres de $\mathbf{Q}$ et $\mathbf{D}$ la matrice diagonale des valeurs propres. Les v valeurs propres sont notées $\lambda_1, \cdots, \lambda_v$.

En représentant les indices dans l'espace propre, l'expression de la distribution normale devient plus simple :

$$P(\Delta|C) = (2\pi)^{-\frac{v}{2}} \prod_{i=1}^{i \leqslant v} \lambda_i^{-\frac{1}{2}} \exp(-\frac{1}{2} \sum_{i=1}^{i \leqslant v} \frac{\Delta_i^2}{\lambda_i})$$

Δ représente la différence $(\mathbf{f}_t - \mathbf{f}_d)$ exprimée dans l'espace propre $(\Delta = \mathbf{U} \cdot \delta)$ liée à l'indice.

5.3. calcul de $P(C|\mathbf{p})$

En notant $\Delta = \mathbf{U}(\mathbf{f}_t - \mathbf{f}_d)$ (où $\mathbf{f}_d$ représente un indice de l'image et $\mathbf{f}_t$ la projection dans l'image de l'indice $\mathbf{f}_d$) du modèle, la probabilité d'une correspondance entre $\mathbf{f}_t$ et $\mathbf{f}_d$ connaissant la pose $\mathbf{p}$ est

$$P(C|\mathbf{p}) = P(C|\Delta) = \frac{P(\Delta|C)P(C)}{P(\Delta)} = \alpha \exp(-\frac{1}{2} \sum_{i=1}^{i<=v} \frac{\Delta_i^2}{\lambda_i}),$$

où

$$\alpha = (2\pi)^{-\frac{v}{2}} \prod_{i=1}^{i \leqslant v} \lambda_i^{-\frac{1}{2}} \frac{P(C)}{P(\Delta)}.$$

$P(\Delta)$ est calculé durant une phase d'apprentissage. Pour cela nous avons localisé différents objets dans différentes images, au moyen d'appariements manuels, et modélisé $P(\Delta)$ par une loi normale. Cette valeur dépend du type d'indices utilisés.

Nous supposons que la probabilité *a priori* de correspondance $P(C)$ est constante, et estimée en fonction du type de scène.

La pose $\mathbf{p}$ est un vecteur de dimension $\mathcal{D}$, où $\mathcal{D}$ représente la dimensionalité de l'espace des poses.

En supposant que la transformation est affine, $\Delta = (\Delta_1, \ldots, \Delta_v)^t = \mathbf{U}(\mathbf{f}_t - \mathbf{f}_d)$ est une combinaison linéaire de vecteurs $\mathbf{f}_d$ et $\mathbf{f}_t$ d'une matrice $\mathbf{U}$. Plus précisément,

$$\Delta = \mathbf{A} \cdot \mathbf{p} + \mathbf{B}, \qquad (2)$$

où la matrice $\mathbf{A}$ et le vecteur $\mathbf{B}$ sont des combinaisons linéaires des vecteurs et $\mathbf{f}$, et de la matrice $\mathbf{U}$. En effet, la transformation affine peut être écrite :

$$\mathbf{f}_t = \mathbf{F} \cdot \mathbf{p},$$

où la matrice $\mathbf{F}$ est composée de valeurs constantes et de valeurs de $\mathbf{f}$. La matrice $\mathbf{F}$ correspondant à une transformation affine 2D est donnée sections 7.2 ; le cas de la projection orthographique avec mise à l'échelle est traité section 8.2.
On peut alors déduire :

$$\begin{aligned}\Delta &= \mathbf{U} \cdot (\mathbf{f}_t - \mathbf{f}_d) = \mathbf{U} \cdot \mathbf{f}_t - \mathbf{U} \cdot \mathbf{f}_d = \mathbf{U} \cdot \mathbf{F} \cdot \mathbf{p} - \mathbf{U} \cdot \mathbf{f}_d \\ &= \mathbf{A} \cdot \mathbf{p} + \mathbf{B}\end{aligned}$$

avec $\mathbf{A} = \mathbf{U} \cdot \mathbf{F}$ et $\mathbf{B} = -\mathbf{U} \cdot \mathbf{f}_d$.
Écrivons

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \cdots \\ \mathbf{A}_i \\ \cdots \\ \mathbf{A}_v \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1j} & \cdots & A_{1\mathcal{D}} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{i1} & \cdots & A_{ij} & \cdots & A_{i\mathcal{D}} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{v1} & \cdots & A_{vj} & \cdots & A_{v\mathcal{D}} \end{pmatrix}$$

$$= \begin{pmatrix} B_1 \\ \cdots \\ B_i \\ \cdots \\ B_v \end{pmatrix}.$$

Soit H_i l'hyperplan de l'espace des transformations défini par

$$H_i = \{\mathbf{p} \ / \ \mathbf{A}_i \cdot \mathbf{p} + B_i = 0\}.$$

Alors $|\Delta_i| \cdot n_i$, avec $n_i = 1/\sqrt{\sum_{k=0}^{k<\mathcal{D}} A_{ik}^2}$,est la distance de la pose $\mathbf{p}$ à l'hyperplan H_i de l'espace des poses. Cela s'écrit : $|\Delta_i| \cdot n_i = D(_{\mathbf{p}}, H_i)$.
Avec ces notations, nous avons :

$$P(C|_{\mathbf{p}}) = \alpha \exp(-\frac{1}{2} \sum_{i=1}^{i<=v} \frac{D^2(\mathbf{p}, H_i)}{\lambda_i n_i^2}). \qquad (3)$$

La probabilité de correspondance est ainsi une fonction de la somme pondérée de distances au carré (distances de la pose $\mathbf{p}$ à l'hyperplan H_i). Cette propriété géométrique est exploitée dans la section suivante.
La section 7.2 donne, en temps qu'illustration, l'expression de H_i pour une transformation affine de segments de droites.

6. probabilité d'une correspondance de primitive image à primitive modèle pour une boîte donnée ($P(C|BOX)$)

La probabilité maximale de correspondance, sachant que la pose doit appartenir à une boîte de l'espace des poses est notée $P(C|BOX)$.
Le calcul de cette probabilité nécessite de maximiser une fonction quadratique (Eq. (3)), sous la contrainte d'inégalités linéaires (l'ensemble des hyperplans de l'espace des poses). Cette maximisation peut être réalisée au moyen de techniques d'optimisation classiques.
Les *multiplicateurs de Lagrange* combinés à des techniques *d'ensembles actifs* sont communément utilisés [15] pour ce genre de problème. Cependant les ensembles actifs sont coûteux en temps de calcul. Par exemple, si l'on travaille dans un espace de dimension 8 comme c'est le cas avec une transformation orthographique avec mise à l'échelle, un système 16×16 doit être résolu à chaque itération. Compte tenu du nombre d'itérations à effectuer (plusieurs dizaines), ce calcul deviendrait pénalisant.
C'est pourquoi nous proposons une technique ne donnant qu'une approximation du résultat, mais avec un nombre d'opérations beaucoup plus restreint.
Nous supposons ici que le nombre d'hyperplan v est inférieur à la dimension de l'espace des pose $\mathcal{D}$. Le cas $v \geqslant \mathcal{D}$ sera traité en fin de section. La pose recherchée est appelée $\mathbf{p}_{min}$.
La technique proposée consiste en deux étapes :

1. soit V l'intersection des v hyperplans H_i correspondants à une paire de primitives appariées (voir la fin de la section précédente pour le détail). V est de dimension $\mathcal{D} - v$. Dans un premier temps, la pose $\mathbf{p}_0 \in V$ est définie par $\forall \mathbf{p} \in V$, $d(\mathbf{c}, \mathbf{p}_0) \leqslant d(\mathbf{c}, \mathbf{p})$ où $\mathbf{c}$ est le centre de la boîte et $d()$ la distance Euclidienne.
   (a) si $\mathbf{p}_0$ est à l'intérieur de la boîte, $\mathbf{p}_{min} = \mathbf{p}_0$.

   Dans ce cas,

   $$D(\mathbf{p}, H) = \sum_{i=1}^{i<=v} \frac{D^2(\mathbf{p}, H_i)}{\lambda_i n_i^2} = 0.$$

   (b) si $\mathbf{p}_0$ n'est pas à l'intérieur de la boîte, alors $\mathbf{p}_{min}$ est l'intersection de la droite $(\mathbf{c}, \mathbf{p}_0)$ avec l'enveloppe convexe de la boîte. Alors, $\mathbf{p}_{min}$ est itérativement amélioré jusqu'à obtenir un minimum pour $D(\mathbf{p}, H)$ avec

   $$D(\mathbf{p}, H) = \sum_{i=1}^{i<=v} \frac{D^2(\mathbf{p}, H_i)}{\lambda_i n_i^2} = \sum_{i=1}^{i<=v} \frac{(\mathbf{A}_i \cdot \mathbf{p} + B_i)^2}{\lambda_j n_i^2}.$$

Ces différentes étapes sont représentées figure 7. Étudions les plus en détail :

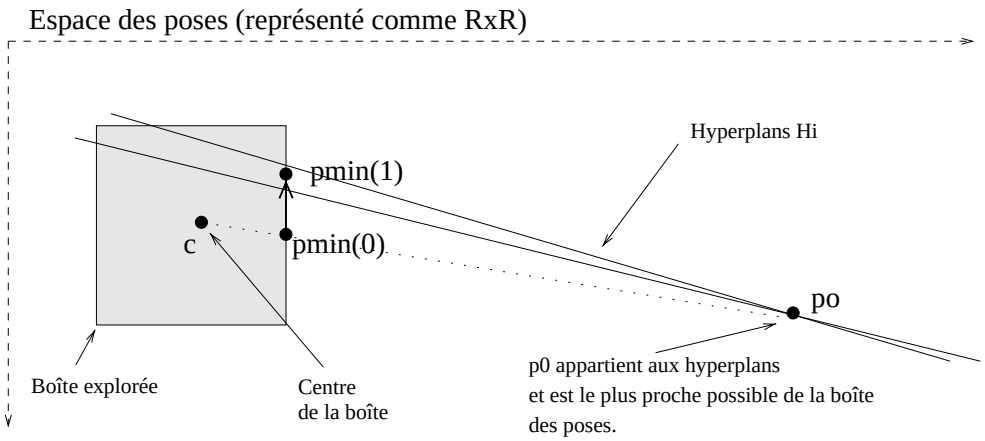


Figure 7. – Minimisation itérative.

Détermination de $\mathbf{p}_0$. La pose $\mathbf{p}_0$, définie par $\mathbf{p}_0 \in V$ telle que $\forall \mathbf{p} \in V, d(\mathbf{c}, \mathbf{p}_0) \leqslant d(\mathbf{c}, \mathbf{p})$ où $\mathbf{c}$ est le centre de la boîte, peut être calculée au moyen des multiplicateurs de Lagrange.

La fonction à minimiser est (avec $\mathbf{p} = (p_1, \ldots, p_{\mathcal{D}})^t$ et $\mathbf{c} = (c_1, \ldots, c_{\mathcal{D}})^t$)

$$f(\mathbf{p}) = \sum_{k=1}^{k \leqslant \mathcal{D}} (p_k - c_k)^2$$

sous la contrainte de l'intersection définie par les v hyperplans $H_i, i \in [1, \ldots, v]$. Le minimum est réellement atteint lorsque $\mathbf{p}$ vérifie :

$$\begin{cases} \nabla(f(\mathbf{p}) - \sum_{i=1}^{i \leqslant v} \ell_i D^2(\mathbf{p}, H_i)) = 0 \\ \sum_{i=1}^{i \leqslant v} D^2(\mathbf{p}, H_i)) = 0 \end{cases}$$

avec ℓ_i les multiplicateurs de Lagrange.

Amélioration de $\mathbf{p}_{min}$. Comme indiqué figure 7, si $\mathbf{p}_0$ n'est-pas dans la boîte des poses à explorer, $\mathbf{p}_{min}$ est d'abord pris comme l'intersection de la droite $(\mathbf{c}, \mathbf{p}_0)$ avec l'enveloppe convexe de la boîte. Ce point n'est pas le minimum de la fonction. Il est possible d'améliorer $\mathbf{p}_{min}$, sachant que le minimum appartient à l'enveloppe de la boîte (sinon $\mathbf{p}_0$ serait dans la boîte). Nous utilisons une stratégie d'alternance des variables, durant laquelle, à chaque itération $k (k \in [1, \ldots, \mathcal{D}])$ la variable p_k est optimisée, en laissant les autres variables inchangées. La direction dans laquelle $\mathbf{p}_{min}$ doit être déplacée est donnée par $\frac{\partial f}{\partial p_k}$. Dans nos expériences $\mathcal{D}$ itérations ont toujours été suffisantes pour assurer une approximation acceptable (toujours à moins de 1 % de la valeur optimale). Cet algorithme est plus de cent fois plus rapide que l'algorithme optimal (utilisant des ensembles actifs).

Probabilité de correspondance. La probabilité de la correspondance est

$$P(C|BOX) = \alpha exp(-\frac{1}{2} \sum_{i=1}^{i<=v} \frac{D^2(\mathbf{p}_{min}, H_i)}{\lambda_i n_i^2}).$$

Remarque. Si la dimension de l'espace de représentation des indices est supérieure ou égale à la dimension de l'espace des poses ($v \geqslant \mathcal{D}$), alors $\mathbf{p}_0$ peut être directement obtenu par une estimation au sens des moindres carrés.

7. application à la reconnaissance d'objets par utilisation de collections de vues 2D

Dans cette section, nous considérons le problème suivant : dix objets différents, présentés figure 8, sont stockés dans une base d'images, sous forme de collections de vues 2D. Le problème et de trouver l'occurrence d'un de ces objets dans une image inconnue, ainsi que son attitude dans l'image.

Comme nous nous intéressons principalement à la recherche des poses-appariements, les autres composantes de l'application sont choisies comme étant aussi simples que possible : les pri-

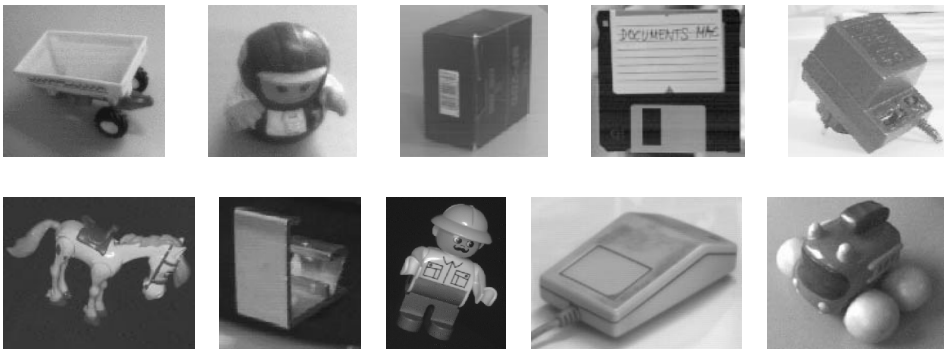


Figure 8. – Les dix objets utilisés pour les expériences présentées.

mitives utilisées sont des segments de droite, les invariants géométriques sont des angles relatifs entre segments. La transformation géométrique utilisée est une isométrie plane.

L'algorithme de reconnaissance se compose des étapes suivantes : les petits groupes de segments ayant des jonctions sont extraits de l'image. Les angles relatifs des segments dans un groupement sont invariants par rapport à la transformation choisie. Cela permet d'indexer une base de vues des objets. Chaque correspondance groupe à groupe permet d'initialiser l'algorithme de recherche de pose.

Les différentes composantes de cette application sont présentées figure 9 et décrites dans les paragraphes suivants.

7.1. base de vues représentant les objets

La description d'objets 3D au moyen de collections de vues d'objet n'est qu'une description « approximative » des apparences possibles des objets, le nombre de vues utilisées étant limité.

Le nombre de vues utilisées peut être optimisé en utilisant l'algorithme proposé par Gavrila et Groen [17]. L'ensemble des points de vue situés à égale distance du modèle est appelé « sphère de vue ». La difficulté est de trouver le maillage de la sphère le plus approprié à l'objet. Observant que l'aspect de la projection de l'objet varie plus dans certaines régions de la sphère que dans d'autres, il semble opportun que la densité des vues ne soit pas la même sur toute la sphère. Les régions où le changement d'aspect varie plus vite doivent être maillées de manière plus dense.

Dans nos expériences, le maillage de la sphère de vues est obtenu en partant d'un maillage régulier grossier, qui est affiné au moyen de la procédure suivante : l'image correspondant au milieu d'une maille est analysée au moyen de l'algorithme de reconnaissance. Si aucune vue de la base ne permet d'expliquer correctement l'aspect de l'objet, cette nouvelle vue est ajoutée à la base, et la maille correspondante est divisée. Ce même traitement est ensuite appliqué récursivement à chacune des sous-mailles produites. Cette approche est voisine de celle proposée par Breuel [8].

7.2. indices et transformation

Un segment du modèle, noté $\mathbf{f}$, est représenté par les coordonnées de ses extrémités, notées $\mathbf{f} = (x_1, x_2, y_1, y_2)^t$.Une pose est un vecteur à quatre dimensions $\mathbf{p} = (tx, ty, sI_x, sI_y)^t$, où tx et ty représentent une translation planaire, et sI_x, sI_y une rotation planaire combinée avec un changement d'échelle. La transformation affine correspondante $\mathbf{f}_t = (x_{1t}, x_{2t}, y_{1t}, y_{2t})^t$, transformation de $\mathbf{f}$ par T est donnée par :

$$\mathbf{f}_t = \mathbf{F} \cdot \mathbf{p}$$

avec

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & x_1 & -y_1 \\ 1 & 0 & x_2 & -y_2 \\ 0 & 1 & y_1 & x_1 \\ 0 & 1 & y_2 & x_2 \end{pmatrix}.$$

Calcul des hyperplans. Les notations utilisées dans ce paragraphe correspondent à celles définies section 5. Dans un but d'illustration, $\mathbf{U}$ est supposé être l'identité (dans les expériences présentées, $\mathbf{U}$ est réellement calculé). Dans ce cas, les hyperplans H_i sont définis par leur normale $(\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \mathbf{N}_4)^t$,

$$\begin{pmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \\ \mathbf{N}_3 \\ \mathbf{N}_4 \end{pmatrix} \cdot \begin{pmatrix} tx \\ ty \\ sI_x \\ sI_y \\ 1 \end{pmatrix} = 0,$$

avec

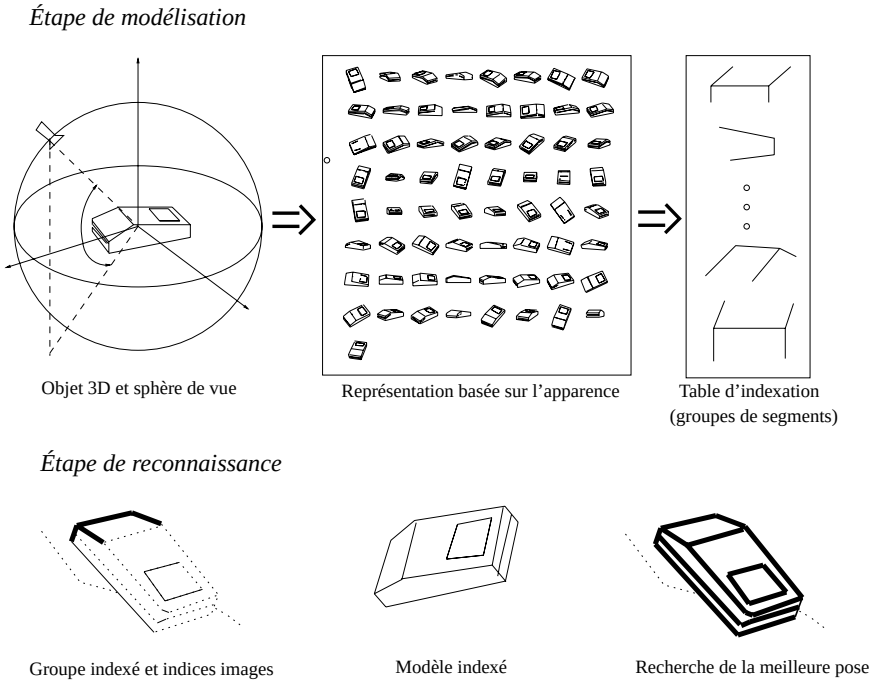


Figure 9. – Principe de la reconnaissance basée sur l'apparence.

$$\begin{pmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \\ \mathbf{N}_3 \\ \mathbf{N}_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_1 & -y_1 & -x_{1d} \\ 1 & 0 & x_2 & -y_2 & -x_{2d} \\ 0 & 1 & y_1 & x_1 & -y_{1d} \\ 0 & 1 & y_2 & x_2 & -y_{2d} \end{pmatrix},$$

où $\mathbf{f}_d = (x_{1d}, x_{2d}, y_{1d}, y_{2d})^t$ représente la paire de segments appariés.

7.3. utilisation de jonctions

Les invariants utilisés sont des angles relatifs entre segments consécutifs ayant des jonctions. D'autres techniques bien plus performantes pourraient être utilisées. Nous rappelons que notre seul but est ici et de valider la technique de recherche de meilleur pose.

L'utilisation de jonctions permet de réduire considérablement le nombre d'indexes à calculer et à stocker pour accéder à la base de vues. De plus les jonctions reflètent la topologie de l'image et ainsi accroissent la probabilité de bonne correspondance.

7.4. indexation de la base de vues

Deux courants d'idées sont présents dans la littérature concernant l'indexation de bases d'images. La première (et aussi la plus populaire) consiste à traiter toutes les hypothèses avec le même poids. Dans ce cas, la difficulté est reportée sur l'étape d'accumulation d'hypothèses.

Une autre voie consiste, comme le proposent Beis et Lowe [5], à utiliser des indices « hautement discriminants ». Il a été montré ([10]) que des tables d'indexation utilisant des indexes de grande dimension sont rapides à accéder, peuvent être appliquées à des bases de grandes dimension et peuvent réduire le nombre d'erreurs d'accès. Cependant, ces tables sont très lourdes à stocker. L'index proposé par ces auteurs est composé de mesures portant sur des chaînes de 4 segments. Le vecteur de mesures est $(\alpha 1, \alpha 2, \alpha 3, l2/l1)$, où $\alpha_i, i \in [1..3]$ sont les angles relatifs des segments successifs et $l2/l1$ le ration des longueurs des segments intérieurs. Nous avons testé cette technique, mais dans de nombreux cas les objets à reconnaître ne comportaient pas de chaînes de 4 segments consécutifs, et dans d'autres cas les appariements se faisaient avec le fond de l'image.

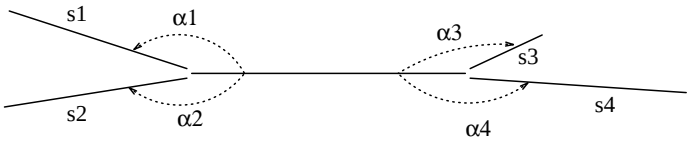


Figure 10. – Chaînes de segments ayant des jonctions.

Tableau 1. – Liste des principales correspondances pour les 3 objets présentés figure 14.

Objet1 P = .91		
Mod.	Im.	Prob.
A	15	.77
O	18	.79
N	19	.74
S	34	.93
L	50	.87
Q	51	.92
T	52	.90
J	6	.78
D	11	.38
C	12	.23

Objet2 P = .95		
Mod.	Im.	Prob.
L	88	.82
P	92	.93
F	95	.87
P	95	.67
M	96	.69
R	97	.90
V	104	.79
Y	104	.81
I	108	.69
]	111	.78
^	111	.61
N	95	.98
S	97	.71
N	100	.97
A	102	.76
Q	103	.90
Z	106	.93
]	106	.62

Objet3 P = .88		
Mod.	Im.	Prob.
I	31	.97
J	32	.82
W	37	.31
C	37	.58
P	48	.56
W	39	.95
I	51	.15
L	29	.90
M	29	.21
Q	30	.76
T	30	.66

C'est pourquoi nous utilisons plutôt le vecteur à 4 dimensions suivant : $(\alpha 1, \alpha 2, \alpha 3, \alpha 4)$, où α_i sont les orientations relatives des segments consécutifs d'une même chaîne, comme représenté figure 10. Comme certains segments peuvent ne pas être détectés dans l'image, des tables d'indexation pour lesquelles des segments sont manquants sont également utilisées.

7.5. expériences

La base de vues contient environ six cents vues 2D de dix objets différents (de 50 à 80 vues par objets). Une vue est un ensemble de segments de droite. Les segments de droite sont obtenus par une approximation polygonale des contours, lesquels sont produits par un extracteur de type « Deriche ».

Les images à analyser contiennent chacune une ou plusieurs difficultés : fond complexe, occultation des objets, fort effet de perspective.

Aucune hypothèse n'est faite, ni sur la pose initiale des objets, ni sur le ou les objets de la base étant présent dans l'image.

Les temps de reconnaissance indiqués (voir figure 15) sont obtenus sur une station de travail HP-700.

Les résultats présentés figure 11 illustrent les différentes étapes de l'algorithme de reconnaissance. Dans un premier temps une approximation polygonale des contours de l'image (a) est calculée (b). 200 segments environs sont extraits. À partir de ces 200 segments, 60 groupes de segments ayant des jonctions communes sont extraits, menant à 800 hypothèses de correspondance groupe à groupe avec l'un des modèles. Chacune de ces hypothèses est vérifiée en calculant la probabilité du modèle dans l'image. La meilleure hypothèse est représentée en (c). Un temps de 4.4 secondes est nécessaire pour tester l'ensemble des hypothèses. Les segments de l'image appariés sont représentés en (d). La meilleure pose (celle donnant la meilleure correspondance entre le modèle et l'image est donnée en (f), tandis que (e) représente le groupe de segments initiaux correspondants à cette hypothèse.

La figure 12 montre les capacité de l'algorithme à traiter des images comportant plusieurs objets, dont certains sont partiellement occultés. L'image initiale est segmentée en une centaine de segments, produisant 40 groupes, qui donnent naissance à 300 hypothèses d'occurrence d'objets. Les trois meilleures hypothèses (celles ayant les plus grandes probabilités) sont représentées sur le bas de la figure (représentation du meilleur aspect et des segments appariés). Le temps de calcul est de 1.4 secondes.

Autres résultats. L'algorithme a été testé sur plusieurs centaines d'images, similaires à celles présentées figure 11. La figure 15 indique le temps de reconnaissance moyen lors de ces expériences, en fonction du nombre d'indices de l'image.

La figure 13 montre, dans plusieurs cas, les meilleures hypothèses obtenues. La figure comporte quatre colonnes. Les deux premières représentent successivement l'image et la segmentation de l'image. La troisième représente la meilleure interprétation obtenue (alignement du modèle sur l'image, pour la

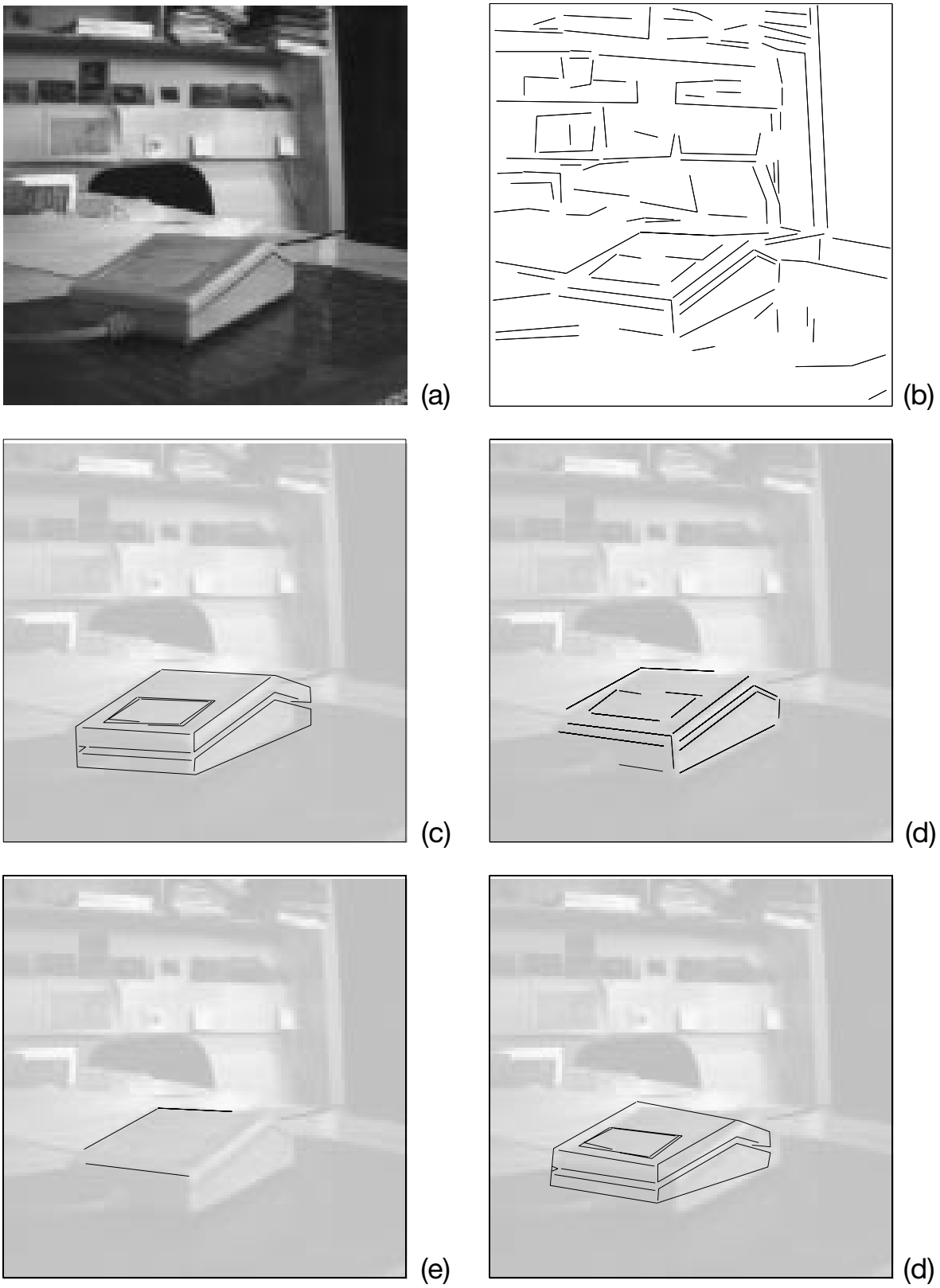


Figure 11. – Reconnaissance 2D-2D : illustration des différentes étapes (voir le texte pour les explications.

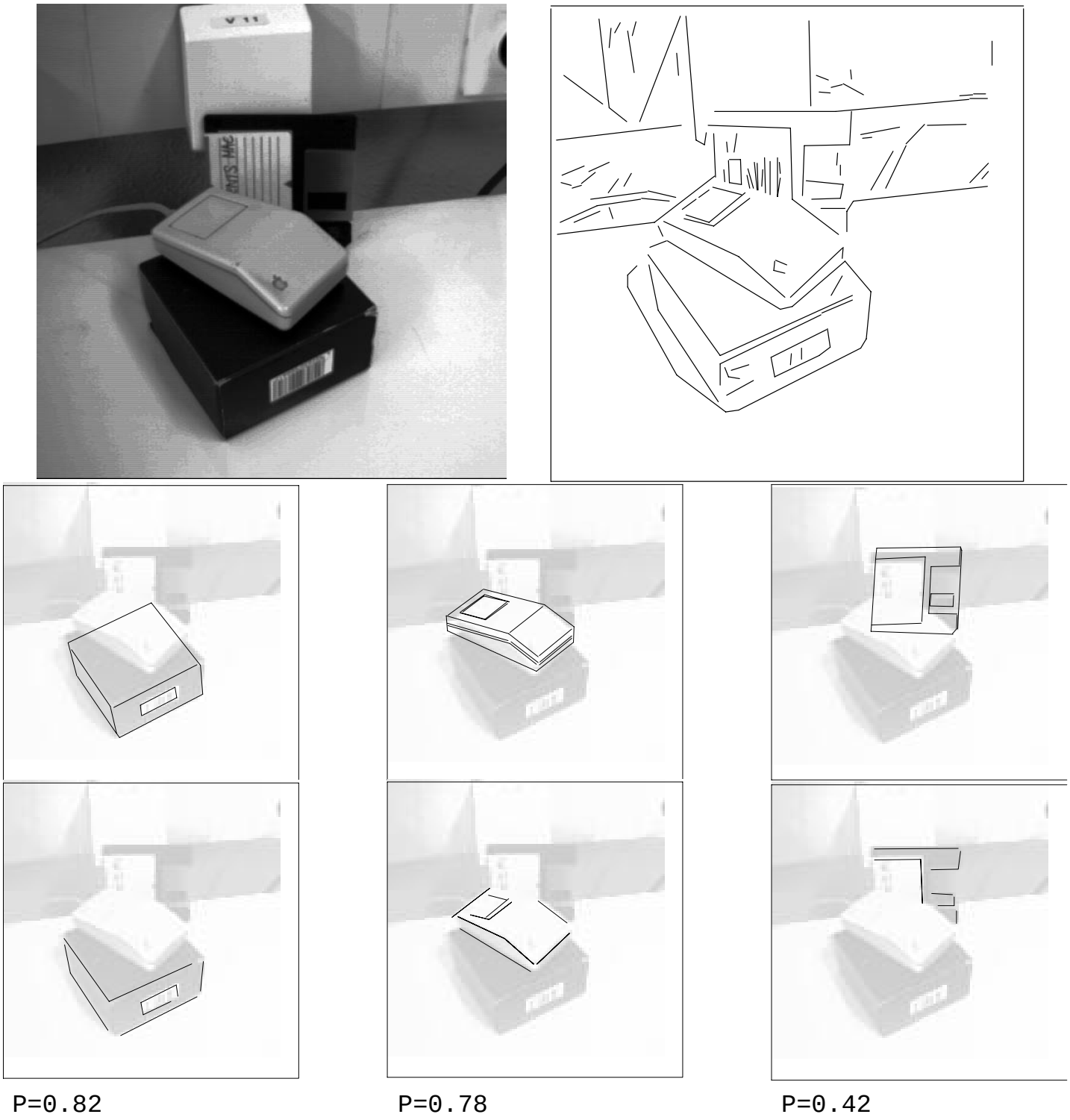


Figure 12.– Reconnaissance 2D-2D : reconnaissance avec occlusions : image, indices visuels, et, pour chaque objet reconnu, recalage du modèle et indices visuels, mis en correspondance.

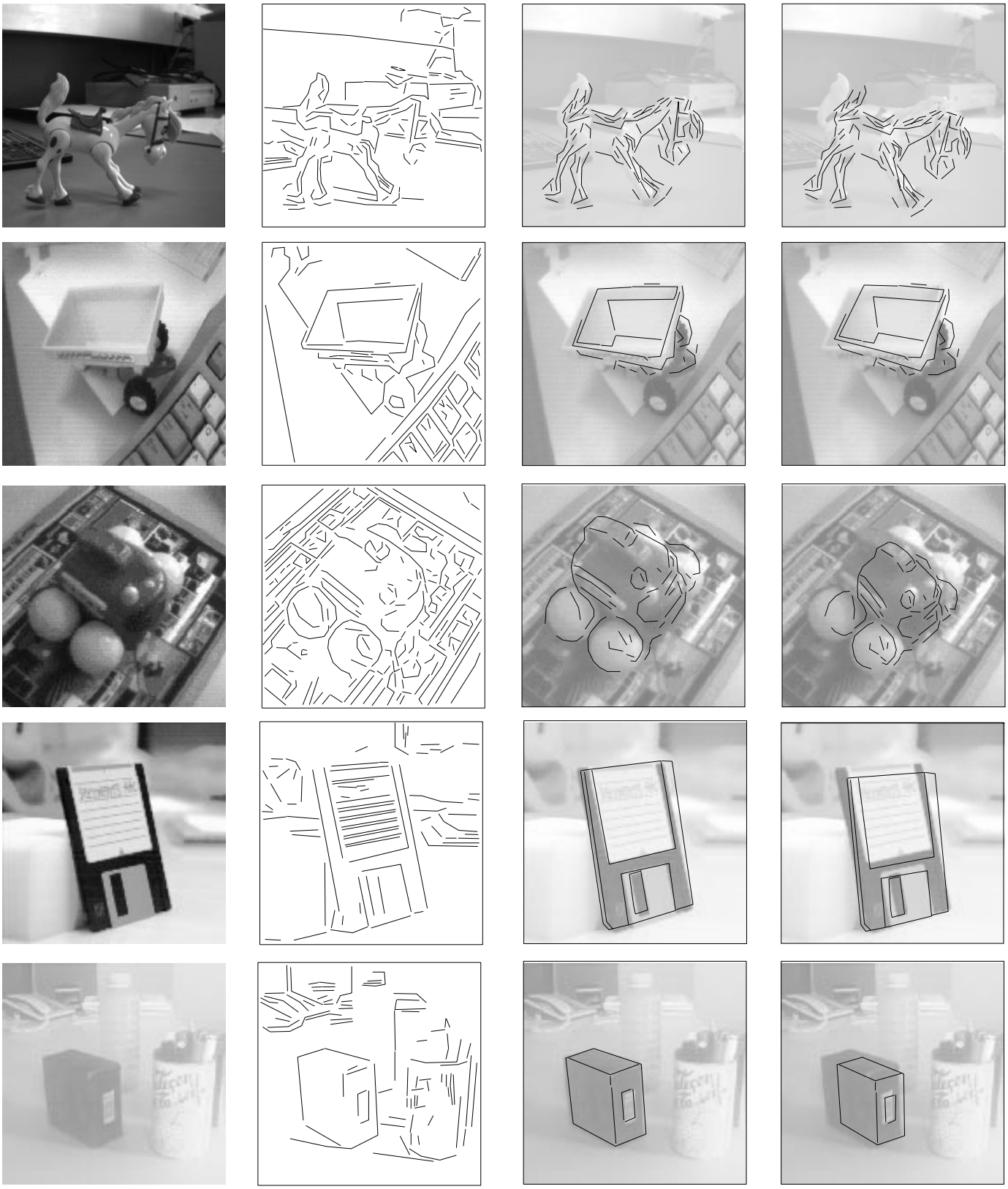


Figure 13.– Reconnaissance 2D-2D : résultats (se reporter au texte pour les explications).

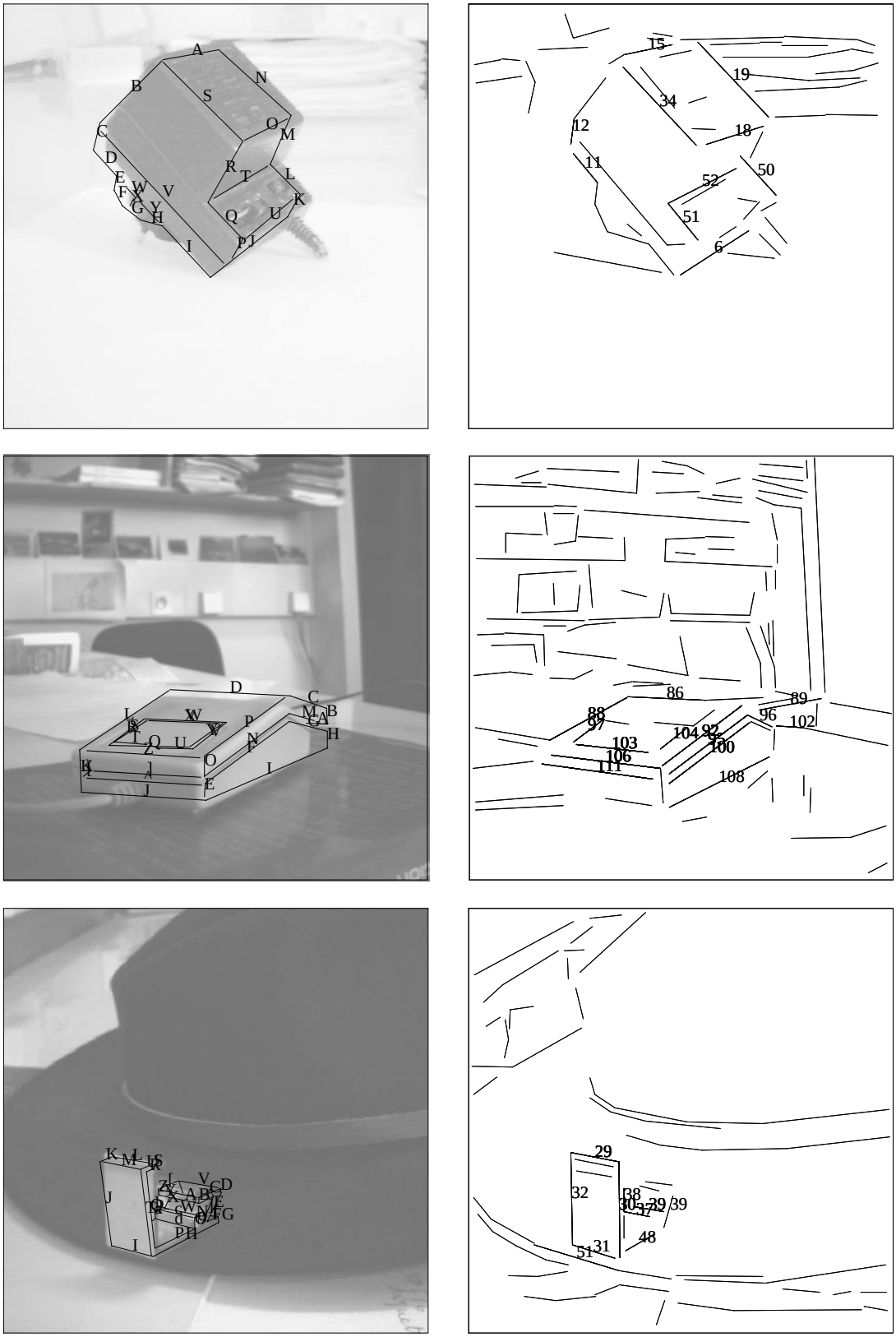


Figure 14.– Reconnaissance 2D-2D : appariements.

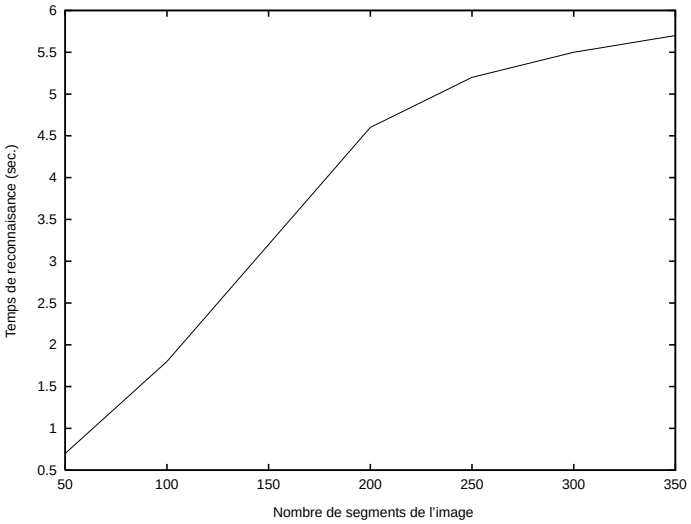


Figure 15. – Temps de reconnaissance, en fonction du nombre de segments de l'image.

meilleure hypothèse). La dernière colonne représente la pose initiale, obtenue par appariement du petit groupe de segments initiaux. Ces résultats montrent les capacité de l'algorithme à vérifier les hypothèses, même lorsque la pose initiale est très imprécise.

La figure 14 et la Table 1 permettent de visualiser les mises en correspondances ainsi que leur probabilité, sur trois autres exemples. La première colonne de la figure représente les images et les meilleurs hypothèses. Les segments du modèle sont numérotés par des lettres. La seconde représente les segments extraits de l'image. Chaque segment est numéroté. La table indique les correspondances, leur probabilité ainsi que la probabilité d'occurrence du modèle.

8. reconnaissance par utilisation de modèles 3D

Dans ce cas, les objets sont représentés au moyen de modèles géométriques 3D, composés de segments de droites. Les objets sont supposés polyédriques, afin de rendre leur modélisation plus facile.

Nous n'utilisons pas ici de techniques d'indexation. D'abord parce qu'en utilisant des modèles 3D plutôt que ces collections de vues, le nombre de modèles est limité. Ensuite, la recherche de propriétés invariantes est beaucoup plus délicate que dans le cas précédent.

Les correspondances initiales sont obtenues en utilisant une technique « d'alignement » [25].

8.1. pose 3D et projection

Les équations de la projection perspective que nous utilisons sont [14]

$$\begin{cases} \mathbf{M}_0\mathbf{M}_i \cdot \mathbf{P}_1 \dfrac{f}{tz} = x_i(1+\epsilon_i) \\ \mathbf{M}_0\mathbf{M}_i \cdot \mathbf{P}_2 \dfrac{f}{tz} = y_i(1+\epsilon_i), \end{cases}$$

où $\varepsilon i = \mathbf{M}_0\mathbf{M}_i(\mathbf{P}_3/t_z - 1)$. $\mathbf{M}_0$ est l'origine du repère objet et $\mathbf{M}_i$ le $i^{\text{ème}}$ point du modèle projeté en $\mathbf{p}_i = (x_i, y_i, 1)$. f est la focale de la caméra et t_z la translation le long de l'axe optique. Nous avons également

$$\begin{pmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{P}_3 \\ \mathbf{P}_4 \end{pmatrix} = \begin{pmatrix} \mathbf{i} & tx \\ \mathbf{j} & ty \\ \mathbf{k} & tz \\ (0,0,0) & 1 \end{pmatrix},$$

où $\mathbf{i}, \mathbf{j}, \mathbf{k}$ représente la rotation et tx, ty, tz la translation.

La projection orthographique avec mise à l'échelle (POE) suppose que les objets appartiennent à un plan passant au centre du repère objetet parallèle au plan image. Cela revient à faire l'approximation $\varepsilon_i = 0$. Dans ces expériences, la projection perspective est dans un premier temps approximée par une POE. ε_i est ensuite calculé itérativement lors de la vérification de la pose [14], rendant l'approximation plus précise.

Comme dans le cas de la reconnaissance 2D-2D, les groupes de segments ayant des jonctions forment les groupements utilisés pour les appariements initiaux. Chaque appariement de groupe permet de calculer une hypothèse de pose.

8.2. indices et transformations

Un segment 3D $\mathbf{f}$ du modèle est représenté par les coordonnées de ses extrémités $\mathbf{f} = (x_1, y_1, x_2, y_2, z_1, z_2)^t$. Une pose est un vecteur de paramètres à huit dimensions $\mathbf{p} = (P_{11}, P_{12}, P_{13}, P_{14}, P_{21}, P_{22}, P_{23}, P_{24})^t$. Avec ces notations, la transformation affine T appliquée à $\mathbf{f}$ donne $\mathbf{f}_t = (x_{1t}, x_{2t}, y_{1t}, y_{2t})^t$. L'équation (4) peut être récrite en

$$\mathbf{f}_t = (\frac{1}{1+\varepsilon_i})\mathbf{F} \cdot \mathbf{p},$$

avec

$$\mathbf{F} = \begin{pmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 \\ x_2 & y_2 & z_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 \\ 0 & 0 & 0 & 0 & x_2 & y_2 & z_2 & 1 \end{pmatrix}.$$

8.4. reconnaissance sans *a priori*

La figure 16 illustre le cas où la pose de l'objet est inconnue. La première image représente la scène, la seconde montre la pose initiale correspondant à la meilleure hypothèse. Les deux images suivantes correspondent respectivement à la meilleure transformation de type projection orthographique avec mise à l'échelle, et meilleure projection perspective, à la fin de la l'étape de vérification [14].

8.5. reconnaissance dynamique

L'algorithme de reconnaissance proposé se prête très bien à des applications de suivi d'objets. Dans ce cas, l'hypothèse initiale faîte sur la boîte contenant la meilleure pose est obtenue simplement à partir de la pose estimée dans l'image précédente.

La figure 17 montre des résultats obtenus sur une séquence d'une centaine d'images. Durant cette séquence, une souris d'ordinateur se déplace au milieu d'objets. Elle est d'abord déplacée sur la droite, puis, lorsqu'elle est complètement occultée par la boîte en carton, elle repart dans l'autre sens. Le mouvement n'est pas prévisible.

Un algorithme de suivi ne peut faire face à ce scénario, l'objet changeant de direction alors qu'il est complètement occulté.

Les résultats présentés ont été obtenus en initialisant la recherche de la meilleure pose avec la pose obtenue sur l'image précédente. La taille de la boîte est dépendante de la qualité de la mise en correspondance obtenue avec l'image précédente.

L'algorithme est appliqué plusieurs fois sur la première image afin d'obtenir une approximation correcte des ε_i. Le temps de traitement est d'environ 200 ms par image sur une station de travail HP-700 (hors temps de segmentation).

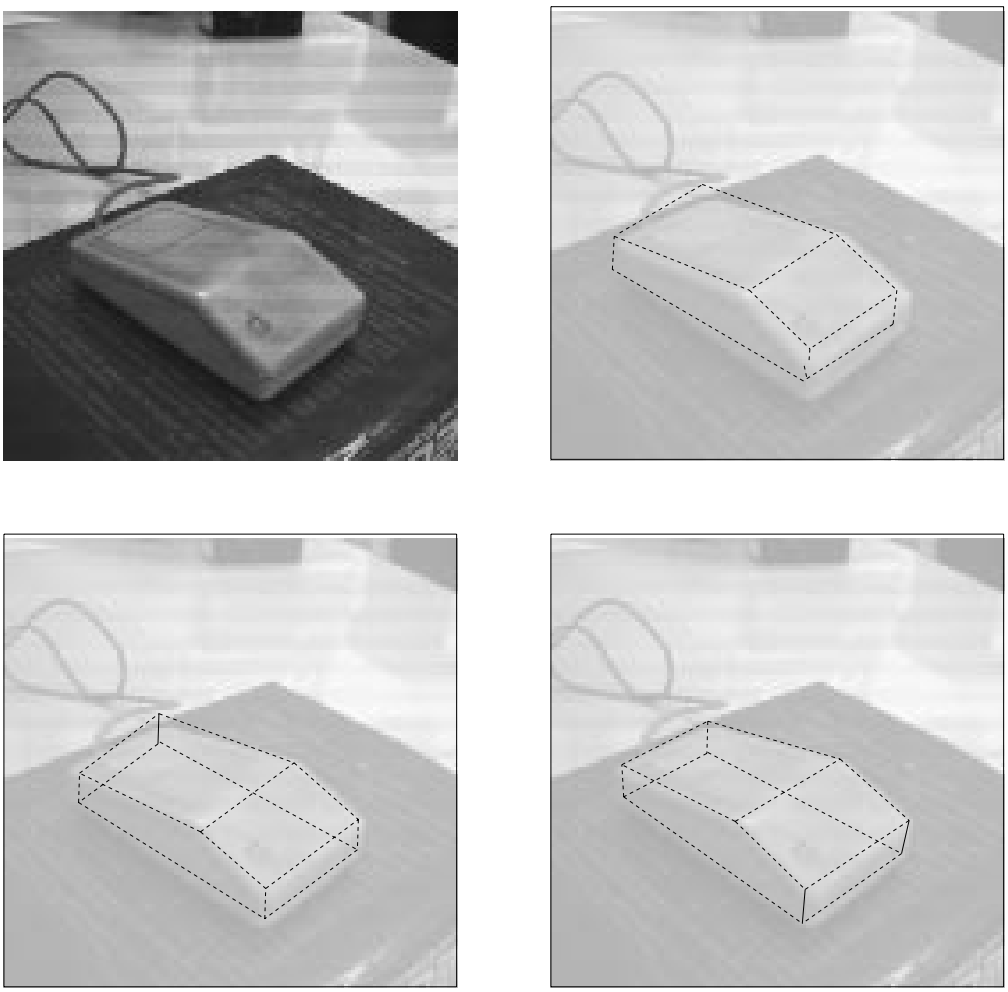


Figure 16.– Reconnaissance 3D-2D : image, pose initiale, meilleure projection orthographique avec mise à l'échelle, meilleure projection perspective.

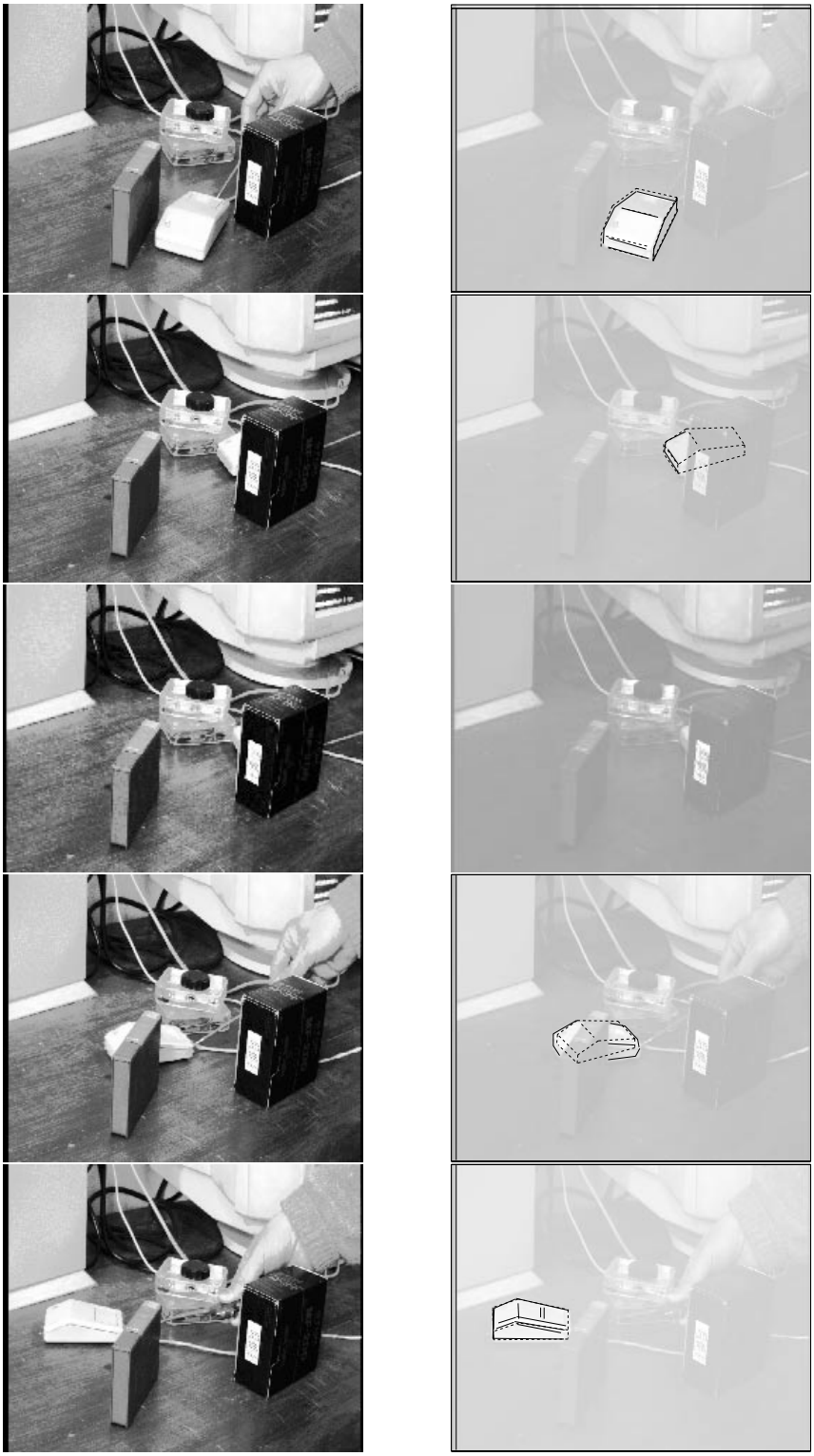


Figure 17.– Reconnaissance 3D-2D : meilleure pose (pointillés) et segments appariés (lignes) sur une séquence d'image.

Nous remarquons que la pose obtenue pour la quatrième image de la figure 17 n'est pas très précise. Cela est dû au mouvement de rotation que subit l'objet lorsqu'il est partiellement occulté. Ce mouvement est difficilement observable en raison des occultations.

9. comparaison avec des travaux antérieurs

Le problème de la reconnaissance d'objets par vérification d'hypothèses présente des liens avec des sujets connexes tels que l'estimation de pose, l'estimation robuste, le suivi robuste d'objets, l'usage des modèles d'erreur. Regardons les avantages de la méthode proposée par rapport aux travaux relatifs à ces domaines.

Hel-Or et Werman [22] proposent une méthode de fusion de données 2D (vues comme des données 3D avec une incertitude infinie sur une coordonnée). Elle requière la connaissant d'une estimée de la pose initiale, laquelle peut être obtenue par la connaissance d'un petit nombre d'appariements. Malgré ses capacités à traiter des données bruitées, cette méthode ne tolère pas de points aberrants et nécessite des mises en correspondances initiales. Elle n'est donc pas directement applicable à notre problème de reconnaissance.

Lorsque l'on s'intéresse aux techniques permettant de prendre en compte les appariement erronés (outliers), deux types d'approches sont possibles. La première consiste à détecter puis éliminer les données erronées en même temps que l'estimation robuste est calculée. La technique utilisée par Kumar et Hanson [28] permettant d'estimer une pose lorsque des appariements sont erronés s'apparente à cette approche. Ils ont comparé l'utilisation de M-estimateurs (de type maximum de vraisemblance) et de L-estimateurs. Ils notent que les M-estimateurs sont très sensibles aux estimations initiales (hypothèses). Cela est dû à la présence de multiples minima locaux dans la fonction à minimiser. Les L-estimateurs (moindres carrés médians) n'ont pas cet inconvénient mais impliquent une phase combinatoire et sont donc beaucoup plus lourds en calcul. La méthode que nous proposons est plus efficace, puisqu'elle peut être vue comme un M-estimateur opérant simultanément autour de différents minima locaux. Ces minima locaux sont le résultats de l'exploration récursive de l'espace des poses. L'algorithme proposé ne garantie pas d'obtenir le minimum global puisqu'il n'explore qu'un nombre fini de branches. Cependant, les résultats expérimentaux que nous avons obtenus montrent que même dans des conditions difficiles, des solutions satisfaisantes ont été obtenues.

L'autre approche pour la prise en compte des appariement erronés consiste à tenter de les détecter avant de procéder à l'estimation. Par exemple Lowe [29] propose d'utiliser une sélection probabiliste itérative des appariements. Cette procédure consiste à utiliser les appariements les plus fiables pour augmenter la probabilité de mettre correctement en correspondance les autres primitives. La propagation du bruit dans l'estimation d'une pose a également été étudiée par Gandhi et Camps [16]. Nous avons observé, en accord avec Kumar et Hanson [28], que les techniques tentant de retirer les appariements erronés avant l'estimation échouent souvent lorsque le nombre d'indices parasites est important.

Plutôt que de tenter de trouver la pose minimisant une fonction d'erreur, Cass [12] propose de générer toutes les interprétations possibles des données dans un temps polynomial. Il avance qu'il n'y a qu'un nombre fini de cas à prendre en compte, en utilisant un modèle d'erreur borné. Pour le cas d'objets 3D planaires soumis à des rotations et translations 3D, il y a $k^6M^6N^6$ classes d'équivalences à évaluer, où M est le nombre d'indices du modèle et N le nombre d'indices de l'image, et k la dimension de l'espace des poses. Dans notre cas les valeurs typiques sont $k = 8, M = 30$, et $N = 100$ ce qui amène à $8^6 30^6 100^6$, ce qui est beaucoup trop important pour être exploré exhaustivement. Cass indique aussi que les méthodes à erreur bornées peuvent être utilisées pour approximer un modèle d'erreur gaussien. L'inconvénient est que cela augmente k et par conséquent le nombre de combinaisons à évaluer.

L'opportunité de l'usage du modèle d'erreur gaussien a fait l'objet de plusieurs travaux [33, 29]. Wells [36] développe une formulation du maximum de vraisemblance dont notre formulation est assez proche. La différence est que nous formulons la fonction à minimiser dans l'espace des poses et non dans l'image. De cette façon cette fonction devient une forme quadratique qui peut être aisément utilisée pour calculer la meilleure pose.

10. conclusions

Nous venons de proposer une technique robuste répondant au problème de la recherche d'appariements, dans un contexte de reconnaissance d'objets. Elle répond au besoin récurrent de valider des hypothèses de pose calculées à partir d'un très petit nombre de correspondances.

La méthode consiste à explorer récursivement l'espace des poses possibles. À la fin de cette exploration, la pose maximisant la probabilité d'occurrence de l'objet dans l'image est obtenue. Durant cette exploration, l'ensemble des appariements compatibles avec cette pose est également établi. L'exploration est réalisée par des divisions successives d'une boîte initiale de l'espace des poses, obtenue par l'appariement d'un petit nombre d'indices. Une recherche hiérarchique de l'espace des poses est ainsi réalisée. Nous avons montré que la recherche de la meilleure pose implique une étape de minimisation quadratique avec contrainte, en supposant que le modèle d'erreur utilisé est

un modèle Gaussien, et en supposant que la transformation géométrique est affine. L'algorithme explore simultanément plusieurs branches de façon à ne pas être stoppé par des minima locaux. L'implémentation de cet algorithme est simple et directe.

Cet algorithme a été intégré dans deux applications de reconnaissance différentes : une application 2D-2D (utilisation de bases d'aspects) et une application 3D-2D (modèle géométrique d'objets 3D), en utilisant des segments de droites. Il peut également être utilisé avec des indices visuels de nature différente ou avec des stratégies de reconnaissance différentes.

Nous avons expérimentalement montré l'efficacité et la robustesse de l'algorithme, même lorsque le bruit et les occlusions sont fréquents. La robustesse est due à la formulation probabiliste utilisée pour caractériser les appariements modèle-image ainsi qu'à la stratégie pyramidale utilisée pour l'exploration.

Nous avons également montré que cet algorithme, utilisé dans un contexte de suivi d'objets, donne des résultats fiables, même en cas de fortes occultations.

BIBLIOGRAPHIE

[1] T.D. Alter, and W.E.L. Grimson, Fast and robust 3d recognition by alignment. In *Proc. IEEE International Conference on Computer vision*, pp. 113-120, Berlin, Germany, 1993.

[2] T.D. Alter and D.W. Jacobs, Error propagation in full 3d from 2d object recognition. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 892-898, Seatle, Washington, 1994.

[3] N. Ayache, and O.D. Faugeras, A new approach for the recognition and positioning of two-Dimensional objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(1): pp. 44-54, January 1986.

[4] DH. Ballard, Generalized hough transform to detect arbitrary shapes, *Pattern Recognition*, 13(2): pp. 111-122, 1982.

[5] J.S. Beis, and D.G. Lowe, Learning indexing functions for 3-d model-based object recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 275-280, Seatle, Washington, 1994.

[6] J.R. Beveridge, and E.M. Riseman. Optimal geometric model matching under full 3d perspective. *Computer Vision and Image Understanding*, 61(3): pp. 351-364, 1995.

[7] T.M. Breuel, Fast recognition using adaptive subdivisions of transformation Space,In *Proc. IEEE International Conference on Computer vision and Pattern Recognition*, pp. 445-451, Champain, Illinois, 1992.

[8] T.M. Breuel, *Geometric Aspects of Visual Object Recognition*. PhD thesis, M.I.T., May 1992.

[9] R. Brunelli, and D. Falavigna, Person identification using multiple cues. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(10): pp. 955-966, October 1995.

[10] A.C. Califano, and R. Mohan, Multidimensional Indexing for Recognizing Visual Shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(4): pp. 373-392, April 1994.

[11] T.A. Cass, Feature matching for object localization in the presence of uncertainty. In *Proc. IEEE International Conference on Computer Vision*, pp. 360-364, Osaka, Japan, 1990.

[12] T.A. Cass, Polynomial-time geometric matching for object recognition. *International Journal of Computer Vision*, 21(1/2): pp. 37-61, 1997.

[13] L.S. Davis, Hierarchical generalized hough transforms and line segment based generalized hough transforms. *Pattern Recognition*, 15(4): pp. 277-285, 1982.

[14] D.F. DeMenthon, and L.S. Davis, Recognition and tracking of 3d objects by 1d search. In *ARPA Image Understanding Workshop*, pp. 653-659, Washington, DC, 1993.

[15] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, New York, wiley-interscience publications edition, 1987.

[16] T.L. Gandhi, and O.I. Camps, Robust feature selection for object recognition using uncertain 2d Image Data. In *Proc. IEEE Conference on Computer Vision and pattern Recognition*, pp. 281-287, Seatle, Washington, 1994.

[17] D.M. Gavrila, and F.C.A. Groen, 3d object recognition from 2d images using geometric hashing. *Pattern Recognition Letters*, 13: pp. 263-278, 1992.

[18] W.E.L. Grimson, The combinatorics of heuristic search term for object recognition in cluttered environment, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13 (9): pp. 920-935, 1991.

[19] W.E.L. Grimson, and D.P. Huttenlocher, On the sensitivity of the hough transform for object recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(3) pp. 255-274, March, 1990.

[20] W.E.L. Grimson, and D.P. Huttenlocher, and D.W. Jacobs, A study of affine matching with bounded sensor error. In *Proc. European Conference on Computer Vision*, pp. 291-306, Santa Margherita Ligure, Italy, 1992.

[21] W.E.L. Grimson, and T. Lozano-Perez, Localizing overlapping parts by searching the interpretation tree. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(4): pp. 469-482, July, 1987.

[22] Y. Hel-Or, and M. Werman, Absolute orientation from uncertain point data : A unified approach. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 77-82, Champain, Illinois, 1992.

[23] D.P. Huttenlocher, and T.A. Cass, Measuring the quality of hypotheses in model-based recognition. In *Proc. European Conference on Computer Vision*, pp. 773-777, Santa Margherita Ligure, Italy, 1992.

[24] D.P. Huttenlocher, and G.A. Klanderman, and W.J. Rucklidge, Comparing images using the hausdorff Distance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(9): pp. 850-863, 1993.

[25] D.P. Huttenlocher, and S. Ullman, Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(2): pp. 195-212, November 1990.

[26] D.W. Jacobs, Robust and efficient detection of salient convex groups. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(1): pp. 541-548, 1996.

[27] F. Jurie, Solution of the simultaneous pose and correspondence problem using gaussian error model. *Computer Vision and Image Understanding*, 73(3): pp. 357-373, 1999.

[28] R. Kumar, and A.R. Hanson, Robust methods for estimating pose and a sensitivity analysis. *Computer Vision and Graphics Image Processing*, 60 (3): pp. 313-342, 1994.

[29] D.G. Lowe, Robust model-based motion tracking through the integration of search and estimation. *International Journal of Computer Vision*, 8(2): pp. 113-122, 1992.

[30] H. Murase and S.K. Nayar. Visual learning and recognition od 3d object from appearance. *International Journal of Computer Vision*, 18 (14): pp. 5-24, 1995.

[31] C.F. Olson, Time and space efficient pose clustering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 251-258, Seatle, Washington, 1994.

[32] A.P. Pentland, and B. Moghadam, and T. Starner, View-based and modular eigenspaces for face recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-91, Seatle, Washington, 1994.

[33] K.B. Sarachik, and W.E.L. Grimson, Gaussian error models for object recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 400-406, New York, 1993.

[34] S. Sclaroff, and A.P. Pentland, Modal matching for correspondence and recognition. *IEEE Pattern Analysis and Machine Intelligence*, 17: pp. 545-561, June 1995.

[35] B.M. ter Haar Romeny, and L.M.J. Florack, and A.H. Salden, and M.A. Viergever, Higher order differential structure of images. *Image and Vision Computing*, 12(6): pp. 317-325, 1994.

[36] W.M. Wells. Statistical approaches to feature-based object recognition. *International Journal of Computer Vision*, 21(1/2): pp. 63-98, 1997.

Manuscrit reçu le 2 octobre 2000

L' AUTEUR

Frédéric JURIE

Actuellement Chargé de Recherche au CNRS au sein du département STIC (Sciences et Technologies de l'Information et de la Communication). Il mêne ses activités de recherche au Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique (LASMEA), UMR 6602 du CNRS, à l'Université Blaise Pascal de Clermont-Ferrand. Ses activités de recherche concernent la reconnaissance et le suivi d'objets dans des séquences d'images vidéo.