SAILLANCE, ZONE D'INTÉRÊT, MISE EN CORRESPONDANCE: TROIS TECHNIQUES POUR LA LOCALISATION ET LA RECONNAISSANCE D'OBJETS EN COULEUR SUR FOND FORTEMENT TEXTURÉ

C. Rauber, C. De Garrini, R. Milanese, S. Startchik, T. Pun

Département d'Informatique, Université de Genève
1211 Genève 4, Suisse
rauber@cui.unige.ch

RÉSUMÉ

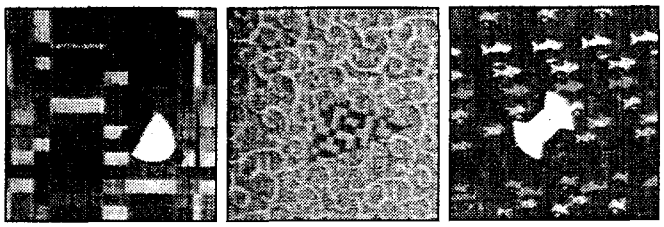
Cet article traite de trois problèmes intervenant lors de la reconnaissance d'objets 2D reposant sur des fonds très texturés: la présence de multiples objets de positions, tailles et orientations inconnues; le fond non uniforme qui génère des primitives irrelevantes pour la reconnaissance et finalement les processus de segmentation qui sont imparfaits. Pour pallier à ces problèmes, une approche basée sur les techniques de saillance, de zone d'intérêt et de mise en correspondance est utilisée et permet la reconnaissance d'objets dans le cas de bases de données contenant un grand nombre d'images.

ABSTRACT

This article is concerned with three commonly occuring problems when trying to recognize 2D objects lying over highly textured backgrounds: presence of multiple objects of unknown positions, sizes and orientations; artifacts occuring due to the background patterns; segmentation errors. An approach based on saliency, attention regions and matching is presented, that allows object recognition with databases containing a large number of images.

1. INTRODUCTION

L'aspect combinatoire de la reconnaissance d'objets basée-modèles représente une limitation majeure en vision par ordinateur [3], en raison principalement des trois difficultés suivantes: l'image d'entrée peut contenir de multiples objets de différentes tailles, positions et orientations, le fond non-uniforme génère des primitives inutiles pour la reconnaissance lors des processus de segmentation et finalement ceux-ci sont eux-mêmes imparfaits. Cet article présente plusieurs stratégies permettant de remédier à chacune de ces difficultés et ainsi de localiser et d'identifier des objets 2D sur des fonds hautement texturés (figure 1).



(a) (b) (c)

Figure 1 : Exemple d'images à analyser (originaux en couleur).

2. SURVOL DU SYSTÈME

Notre système de reconnaissance (figure 2) est spécialisé dans l'identification d'objets en couleur déposés sur un fond fortement texturé (taille de 256x256 pixels, 3x8 bits RGB). Les images comportent des objets à deux dimensions, ou alors des vues stables d'objets tridimensionnels. Malgré cette restriction, ces images sont très difficiles à analyser: le fond non-uniforme produit lors de l'application des processus de segmentation beaucoup de primitives inutiles interférant avec celles de l'objet. De plus, la position, l'orientation ainsi que la taille des objets sont à priori inconnues.

Etant donné ces hypothèses, les difficultés pour la reconnaissance des objets sont principalement de deux types: leur localisation dans l'image est complexe et l'extraction des primitives les plus relevantes pour permettre leur identification est imparfaite.

La localisation des objets est résolue en appliquant un processus appelé *focus d'attention* [4] qui permet d'identifier de manière "bottom-up" les zones les plus intéressantes de l'image. L'algorithme fournit comme résultat un masque binaire entourant chacun des objets. Ceux-ci sont ainsi séparés dans la scène analysée, ce qui empêche le système de reconnaissance de mélanger les informations appartenant à des objets différents.

Afin d'accéder rapidement aux informations les plus pertinentes, une mesure de *relevance,* ou saillance est attribuée à chacune des primitives, permettant ainsi leur classement par ordre d'importance.

Finalement un processus de normalisation et d'indexation des données précédemment calculées permet d'identifier les objets présents dans l'image. Ce processus de reconnaissance fait appel à une base de données (sous la forme de tables de hashing) contenant la description géométrique de tous les modèles à re-

buée. Pour les autres primitives, une valeur représentant la distance moyenne normalisée est calculée:

$$d\left(\tau_i^p\right) = \sum_i^N d\left(P_i, \partial\Re\right) / N$$

où P_i sont les N pixels de la primitive τ_i^p, $\partial\Re$ est le bord de la région d'attention la plus proche et d est la fonction calculant la distance entre le pixel P_i et $\partial\Re$. La mesure de distance d décroît exponentiellement en fonction de l'éloignement en position, ce qui permet d'éliminer sans ambiguïté les primitives les plus éloignés des régions d'attention. Cette mesure de distance est générale et peut donc être utilisée pour toutes les sortes de primitives.

Les mesures de distance et de relevance sont ensuite additionnées pour obtenir la mesure finale de *saillance*:

$$R\left(\tau_i^p\right) = \left(\rho\left(\tau_i^p\right) + d\left(\tau_i^p\right)\right)/2$$

Seules les primitives ayant une valeur $S\left(\tau_i^p\right)$ de saillance plus grande que 0.5 sont conservées (figure 8).

Figure 8 : (a,b,c) Affichage de la relevance finale $R\left(\tau_i^p\right)$ et suppression des primitives les moins significatives ($R\left(\tau_i^p\right)<0.5$) pour les segments, arcs et régions.

8. NORMALISATION ET RECONNAISSANCE

Après avoir obtenu un nombre restreint de primitives décrivant l'objet inconnu (figure 8), il est nécessaire de normaliser ces éléments afin d'effectuer, dans la phase de reconnaissance, une mise en correspondance efficace entre les modèles de la base de données et l'objet. La normalisation doit s'effectuer pour la translation, le changement d'échelle et la rotation.

Pour trouver le facteur de translation de l'objet, le centre de gravité de la région d'attention est utilisé et devient le centre de l'image. Le facteur d'échelle est donné par le rapport entre le plus long axe de l'ellipse approximant la région d'attention et de la taille standard des modèles de 100 pixels (figure 9.b).

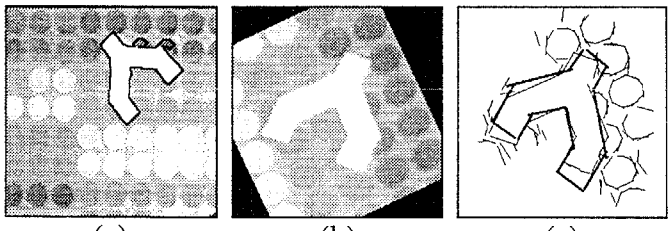


Figure 9 : (a) Image initiale avec indication en noir de la région d'attention. (b) Normalisation en position et en taille. (c) Vérification du modèle sur les données pour toutes les rotations.

La phase d'indexation consiste alors à retrouver le meilleur modèle parmi ceux que contient la base de données [1]. Puisque la position et l'échelle des objets ont été calculées, il ne reste plus qu'à déterminer l'orientation de l'objet par rapport aux modèles. L'angle de rotation est donné en prenant simplement toutes les orientations possibles des objets, de 0 à 360 degrés par pas de 1 degré. A partir de cette normalisation en translation, homothétie et rotation, la position des primitives donne immédiatement le point d'entrée dans la table de hashing. L'objet à identifier correspond au modèle ayant le plus de primitives en commun avec lui pour une orientation donnée (figure 9.c). Le fait d'utiliser une table de hashing permet donc d'obtenir sans calculs supplémentaires l'information souhaitée à partir des quelque primitives extraites de l'image.

9. CONCLUSION

Le système d'indexation a été testé sur 180 images différentes. Malgré la difficulté des images à traiter, la reconnaissance atteint un taux de succès de 80%. Parmi les objets non reconnus, l'erreur provient soit d'une mauvaise localisation de ceux-ci (9 objets), soit d'une normalisation erronée (12 objets), soit d'une très mauvaise segmentation de l'image (3 objets) ou d'un dysfonctionnement du système (9 objets). A travers l'utilisation d'une table de hashing bidimensionnelle indexée par les primitives les plus saillantes, le temps de calcul pour la reconnaissance nécessite environ 5 secondes sur SPARC 10 et ne dépend pas du nombre de modèles.

10. RÉFÉRENCES

[1] R. Bergevin, M.D. Levine, "Generic object recognition: building and matching coarse description from line drawings", IEEE Trans. PAMI 15(1), pp. 19-36, 1993.

[2] J.-M. Bost, R. Milanese and T. Pun, "Temporal precedence in asynchronous visual indexing", Springer-Verlag, Lecture Notes in C.S., D. Chetverikov and W.G. Kropatsch, Eds., 719, 1993, pp. 468-475.

[3] W.E.L Grimson, "The Combinatorics of Object Recognition in Cluttered Environements Using Constrained Search", Artificial Intelligence Journal, vol. 44(1-2), 1989, pp. 121-165.

[4] R. Milanese, H. Wechsler, S. Gil, J.-M. Bost, T. Pun, "Integration of bottom-up and top-down cues for visual attention using non-linear relaxation", IEEE CVPR 94, Seattle, USA, June 20-23, 1994.

[5] A.R. Pope and D. Lowe, "Learning Object recognition Models From Images", in Proc. of the Int. Conf. on Computer Vision, 1993.

[6] T. Pun, C. Rauber, S. Startchik and R. Milanese, "Transforming an image into dataflows of relevant primitives for objects location, reconstruction and indexing", Proc. Vision Interface 95, Quebec City, Canada, May 15-19, 1995.

[7] F. Stein and G. Medioni, "Structural Hashing: Efficient Three Dimensional Object Recognition", in Proc. of the Int. Conf. on Computer Vision and Pattern Recognition, Maui, Hawai, 1991, pp. 244-250.