

Exploration située des scènes naturelles

Joseph Machrouh, Philippe Tarroux
Groupe Perception Située
LIMSI-CNRS, BP 133 F-91403 Orsay FRANCE
{joseph.machrouh, philippe.tarroux}@limsi.fr

Abstract

This paper presents a visual architecture able to extract salient points from a visual scene and to use them to guide an artificial vision system to interesting locations. In order to be able to move in the image, one will distinguish a low-resolution wide field processing and a high resolution focal processing. One thus separates the exploration phase associated with the search of interest points from the exploitation phase associated with recognition. The energy of these interest points according to various orientations and space frequencies is then used to index them. We showed that the low frequency components of this indexed representation are sufficient to bias the saliency of the scene in favour of targets similar to the representations memorized. They are also sufficiently robust to preserve this property in a video sequence subject to strong contrast variations. We showed as well that interest points based on a multi-scale frequency analysis can be used to control exploratory saccades by using a bottom-up mechanism; the low frequency part of such a representation can be used to control the saccades required to attain the target in a top-down way.

1. Introduction

La vision par ordinateur a été influencée par le paradigme de Marr [1]. Ce modèle se fonde sur l'idée que la vision est un processus ascendant qui se caractérise par un traitement linéaire de l'information. Depuis quelques années, ce principe est critiqué par des chercheurs qui suggèrent que des informations descendantes pourraient intervenir pour modifier le traitement de bas niveau afin de mieux prendre en compte le but recherché. Ce processus, nommé vision active [2] [3], permet de replacer la vision dans un contexte comportemental.

Le traitement des informations visuelles en vision robotique doit être effectué en temps réel. Un traitement partiel des scènes visuelles peut améliorer le temps de calcul des systèmes de vision artificiels traditionnels qui traitent de la totalité de la scène. Ce traitement partiel peut être réalisé par l'introduction d'un système sélectif au cours du processus visuel. Les systèmes sélectifs permettent de rechercher des régions saillantes de la scène visuelle en calculant des cartes de saillances [4] [5] [6]. Ces cartes de saillances calculées de différentes façons permettent de sélectionner des régions ou des points de saillance élevée qui guident le champ visuel d'un système de vision et ainsi limitent le traitement uniquement à cette partie de la scène. La sélection de ces points saillants est fondée seulement sur un calcul ascendant de l'énergie et il n'existe aucune interaction avec des informations de haut niveau qui pourraient modifier cette sélection. La sélection de ces points d'intérêt par des informations de haut niveau permet aux systèmes de vision artificielle de focaliser leur attention sur ces régions et ainsi d'atteindre leur but plus rapidement. Les systèmes de vision combinant des informations de haut niveau avec le traitement bas niveau se classent dans la catégorie vision attentionnelle.

Nous présentons dans cet article une architecture d'un système de vision artificielle appartenant à la catégorie vision attentionnelle. Cette architecture s'appuie sur une combinaison entre les informations de haut niveau et des informations de bas niveau. Ces dernières permettent d'obtenir des points saillants qui permettent de guider le champ visuel du système vers des régions intéressantes de la scène visuelle. Les informations de haut niveau issues de la mémoire du système et du but recherché permettent de sélectionner parmi ces points saillants ceux qui sont utiles au but recherché. La particularité de cette architecture réside dans la séparation effectuée entre l'opération d'exploration, qui s'effectue en se basant uniquement sur la basse fréquence spatiale, et l'opération de reconnaissance qui utilise toutes les fréquences spatiales pour confirmer ou non la région focalisée comme étant similaire ou non à la région recherchée dans la scène.

2. Modèle

Les images naturelles que nous explorons tous les jours ont une statistique particulière, elles sont auto-similaires, sont fortement corrélées et redondantes [7] [8] [9] [10]. Barlow [11] a suggéré que la réduction de cette redondance dans le traitement d'information bas niveau améliore le transfert et le traitement de ces informations dans les traitement de plus haut niveau. Cette réduction de redondance peut être effectuée par une

décorrélation du signal d'entrée. Daugman [12] a suggéré qu'un moyen de décorrélérer un signal est de le filtrer par un banc de filtre de Gabor.

$$r(\mathbf{x}, \mathbf{\Omega}_{k,\theta}) = e^{-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}} e^{-i\Omega \mathbf{x}} * I(\mathbf{x})$$

où $I(\mathbf{x})$ est l'image initiale, $r(\mathbf{x}, \mathbf{\Omega}_{k,\theta})$ est l'image filtrée et $e^{-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}} e^{-i\Omega \mathbf{x}}$ est le filtre de Gabor utilisé pour convoluer l'image. $\mathbf{\Omega}_{k,\theta}$ est un vecteur qui définit les orientations préférentielles du filtre telles que $\mathbf{\Omega}_{k,\theta} = \mathbf{\Omega}_k \mathbf{R}_\theta$ est une matrice de rotation et $\mathbf{\Omega}_k = (\omega_k \ 0)$ est la bande de fréquence.

La première partie dans la construction de l'architecture du système présenté a consisté à construire un processus de filtrage ascendant qui utilise un banc de filtre de Gabor dans le but d'extraire des régions d'intérêt qui sont considérées comme des régions d'énergie élevée.

La deuxième étape dans la construction du système a consisté à extraire des caractéristiques de plus haut niveau que celles extraites par les filtres de Gabor. Nous avons décidé que ces caractéristiques de plus haut niveau sont une combinaison linéaire des caractéristiques de bas niveau. La méthode la plus appropriée alors sera l'utilisation d'une approche fondée sur les composantes principales. Une analyse en composantes principales sera alors appliquée sur les caractéristiques extraites par le filtre de Gabor. Celle-ci permet d'obtenir des vecteurs propres qui définissent un espace de projection.

Dans ce travail les orientations utilisées sont $\theta \in \{0, \pi/4, \pi/2, 3\pi/4\}$ et les bandes de fréquence sont $k \in \{1/32, 1/16, 1/8, 1/4 \text{ cyc/pixel}\}$. Seules les normes des filtres de Gabor sont utilisées. On obtient ainsi pour chaque image, 16 images résultantes. Les traitements multifréquences sont réalisés à l'aide d'une pyramide de Burt [13]. Pour la construction du nouvel espace de projection, nous avons extrait 2436 vignettes de tailles 5*5 en basse fréquence (ce qui correspond à des tailles de 20*20 pixels pour la fréquence la plus haute en utilisant 4 fréquences spatiales) d'une façon aléatoire à partir d'un ensemble de 21 images naturelles. Ces vignettes sont caractérisées par des vecteurs dont chaque élément est la moyenne à la sortie des normes de filtres de Gabor à une fréquence et à une orientation donnée. Ces vecteurs sont calculés de la façon suivante :

$$\mathbf{V}_k = (\overline{r_{k,0}}, \dots, \overline{r_{k,\theta}})$$

avec

$$\overline{r_{k,\theta}} = \left\| \overline{r(\mathbf{x}, \mathbf{\Omega}_{k,\theta})} \right\| = \frac{1}{m \times n} \sum_{\mathbf{x}} r(\mathbf{x}, \mathbf{\Omega}_{k,\theta}) \times r^*(\mathbf{x}, \mathbf{\Omega}_{k,\theta})$$

où m et n sont les nombres de lignes et de colonnes des vignettes. L'application d'une analyse en composantes principales sur ces vecteurs permet d'obtenir 4 vecteurs propres pour chaque fréquence spatiale.

L'espace de projection obtenu à partir de ces vecteurs propres est un espace représentatif des régularités statistiques observées dans le sous-ensemble des images naturelles utilisées pour cette construction. Dans cette nouvelle base, les points d'intérêt obtenus diffèrent suivant l'axe considéré [14]

La projection d'une image sur le nouvel espace de projection permet d'extraire des caractéristiques telle que fins de lignes ou des courbures. La figure 1a montre l'extraction des caractéristiques dans une image test après avoir été projetée sur le premier axe de l'ACP correspondant à la fréquence la plus élevée. Ces caractéristiques correspondent à des terminaisons. Ce mode d'extraction est invariant en rotation. La figure 1b représente l'extraction des caractéristiques dans une autre image test projetée sur le premier axe d'ACP correspondant à la fréquence la plus basse. Ces caractéristiques correspondent à des courbures. La figure 1c montre le résultat obtenu sur une image en niveau de gris. Ce résultat est issu de la projection de l'image sur le premier axe d'ACP correspondant à la plus haute fréquence. Les projections sur les autres axes d'ACP ne sont pas mentionnées dans cet article.

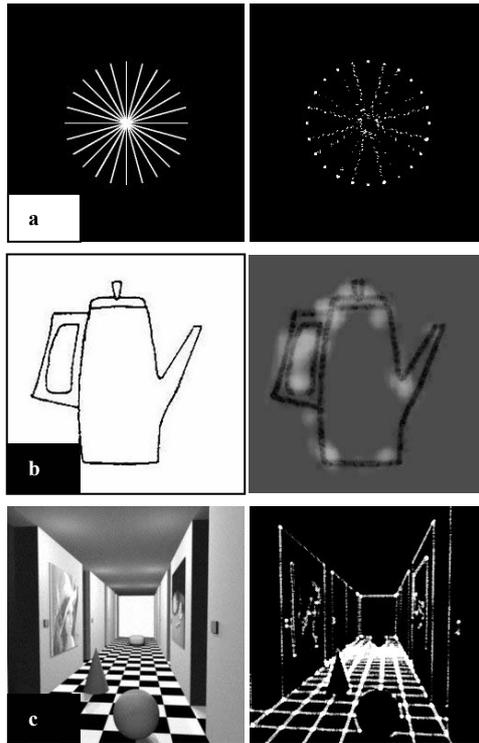


Figure 1. La figure montre la projection des images test sur le nouvel espace de projection obtenu par l'ACP. La figure 1a montre l'extraction de terminaisons obtenue par projection sur le premier axe de l'ACP correspondant à la fréquence la plus élevée. La figure 1b montre l'extraction de courbure obtenue par projection de l'image sur le premier axe de l'ACP correspondant à la fréquence la plus basse. La figure 1c montre le résultat sur une image en niveau de gris issue de la projection sur le premier axe de l'ACP correspondant à la fréquence élevée. Un rehaussement au niveau de luminance et de contraste a été appliqué aux trois figures pour mieux montrer le résultat obtenu.

Ces points caractéristiques extraits par la projection sur le nouvel espace d'ACP, peuvent être exploités pour expliquer quelques illusions optiques. Plusieurs études ont été faites pour modéliser les illusions optiques [15] [16] [17]. En effet plusieurs explications ont été avancées. Une parmi ces explications suggère que les images bidimensionnelles sont interprétées comme étant des perspectives et que le système visuelle interprète ces deux schémas comme ne pas appartenir au même plan [18] [19] [20]. Nous avons testé ce principe d'extraction de caractéristiques sur des figures d'illusions comme celles de Müller-Lyer (figure 2). En effet la figure montre que les points caractéristiques extraits dans les deux figures sont différents. Pour la première figure les points extraits sont à l'intérieur du segment alors que les points extraits dans la deuxième figure sont à l'extérieur. Nous pensons que le système visuel réalise des saccades sur les points saillants et que cette saccade permet d'interpréter que le segment de la première figure serait plus petit que le segment de la deuxième figure.

Cette expérience permet de montrer l'importance de la réalisation des saccades. En effet, ces derniers permettent d'apporter une information supplémentaire qu'on ne pourrait pas acquérir par un traitement d'image classique.

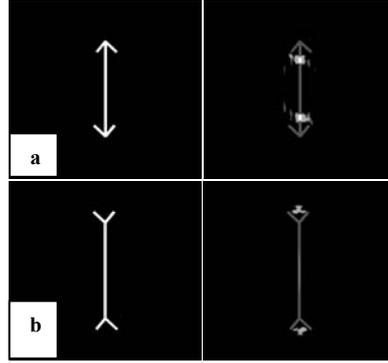


Figure 2. Exemple d'extraction de caractéristiques sur des figures d'illusion optique. La figure montre le résultat de cette extraction sur l'illusion de Müller-Lyer correspondant à la projection sur la quatrième axe de l'ACP correspondant à la fréquence la plus basse. Un rehaussement de luminance et de contraste a été effectué sur l'image résultat.

3. Algorithme d'exploration

La procédure d'exploration s'appuie sur la méthode d'extraction des caractéristiques de haut niveau. Le système explore ces points de deux manières :

- Ascendante : le système est guidé par des saillances issues de la scène visuelle.
- Descendante : le système est un mécanisme attentionnel guidé par une information préalablement mémorisée.

Lorsque l'utilisateur désigne une partie de la scène visuelle, le système se focalise une première fois sur le point désigné, il calcule ensuite les points saillants dans son champ visuel. Il se refocalise une deuxième fois sur le point le plus saillant. Il calcule alors les normes des filtres de Gabor sur les différentes fréquences spatiales et les différentes orientations. Il génère deux vecteurs de taille 12 (3 fréquences spatiales et 4 orientations) : un vecteur correspondant aux énergies des normes de Gabor (VE) et un vecteur image norme (VV). Le vecteur VE est calculé de la façon suivante :

$$\mathbf{V}_E = (\overline{r_{0,0}}, \dots, \overline{r_{0,\theta}}, \overline{r_{1,0}}, \dots, \overline{r_{n,\theta}})$$

avec

$$\overline{r_{l,\theta}} = \frac{\|r(\mathbf{x}, \Omega_{l,\theta})\|}{m \times n} = \frac{1}{m \times n} \sum_{\mathbf{x}} r(\mathbf{x}, \Omega_{l,\theta}) \times r^*(\mathbf{x}, \Omega_{l,\theta})$$

Le vecteur VV est calculé de la façon suivante :

$$\mathbf{V}_V = (\mathbf{N}(\mathbf{x}, \Omega_0), \dots, \mathbf{N}(\mathbf{x}, \Omega_n))$$

avec

$$\mathbf{N}(\mathbf{x}, \Omega_l) = (\mathbf{I}(\mathbf{x}, \Omega_{l,0}), \dots, \mathbf{I}(\mathbf{x}, \Omega_{l,\theta}))$$

$$\mathbf{I}(\mathbf{x}, \Omega_{l,\theta}) = (I_{\Omega_{l,\theta}}(1,1), \dots, I_{\Omega_{l,\theta}}(m,n))$$

où $I_{\Omega_{l,\theta}}(1,1)$ est l'intensité du pixel de l'image norme à la sortie du filtre de Gabor au point (1,1) à la fréquence l et à l'orientation θ .

En mode exploration ascendante, le système analyse chaque point saillant, en comparant le vecteur énergie \mathbf{V}_E de ce point avec le vecteur énergie \mathbf{V}_E correspondant à la région mémorisée. Cette comparaison est effectuée à l'aide d'une fonction à base radiale et permet de calculer un score de similarité

$$s = e^{-\frac{\|\mathbf{V}_E - \mathbf{W}_E\|^2}{2\sigma^2}}$$

En mode exploration descendante, avant que le système ne focalise son attention sur un point saillant, il effectue une opération de comparaison en basse fréquence lui permettant de ne retenir que les points supérieurs à un seuil (voir figure 3). Cette opération s'effectue de deux manières :

- Comparaison à l'aide d'une fonction à base radiale des parties basse fréquence du vecteur énergie VE du point saillant et du vecteur énergie WE de la région mémorisée, qu'on nomme Topdown énergie (TDE)
- Comparaison à l'aide d'une fonction à base radiale des parties basse fréquence du vecteur image normes VV du point saillant et du vecteur WV de la région mémorisée, qu'on appelle topdown vecteur (TDV).

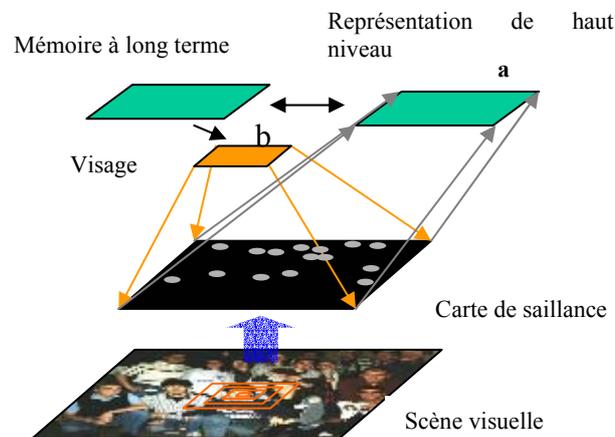


Figure 3. Pour les points saillants, l'information basse fréquence issue de filtres de Gabor est confrontée à une représentation de la cible cherchée également à basse fréquence (rectangle b).

Une fois l'exploration effectuée, le système réalise une opération de comparaison de la région focalisée et de la région mémorisée. Cette opération s'effectue de la même façon pour les différentes méthodes d'exploration. le système compare, à l'aide d'une fonction à base radiale, le vecteur énergie VE de la région focalisée et le vecteur WE de la région mémorisée. Un score de similarité est alors attribué à chaque point visité.

4. Résultats

4.1. Natures des points explorés

Les résultats qui suivent sont issus d'un test du système sur une scène d'extérieur. Le système a mémorisé une zone de l'image représentant un visage, et explore la scène visuelle à la recherche des zones qui sont susceptibles de ressembler à la zone mémorisée. Nous avons testé les trois méthodes d'exploration citées précédemment.

4.1.1. Exploration ascendante (bottom-up)

Le système est guidé par les saillances naturelles classées par ordre décroissant suivant leur degré de saillance. Il se focalise sur chaque point saillant et compare pour ce point, le vecteur énergie VE du point en question avec le vecteur énergie WE de la zone mémorisée. Un score de similarité est alors attribué à chaque point visité. La figure 4 montre le résultat d'une telle exploration.

Les points explorés par ce mode d'exploration ont une grande variabilité. Leur saillance relative varie entre 0,0 à 1,0. Les points les plus explorés ont une saillance relative comprise entre 0,4 et 0,5.

4.1.2. Exploration descendante (top-down) énergie (TDE)

Le système est guidé par une information de haut niveau. Cette information correspond aux énergies à basse fréquence des normes des filtres de Gabor (figure 4).

Nous constatons que la variabilité des points visités est moins grande que celle de l'exploration ascendante. Les saillances relatives varient entre 0,3 et 1,0. Le pourcentage des points ayant une saillance relative comprise entre 0,8 et 0,9 est de 41%. Les points les plus visités ont une saillance relative comprise entre 0,7 et 1,0.

4.1.3. Exploration descendante (top-down) vecteur (TDV)

Pour l'exploration descendante vecteur, la partie basse fréquence du vecteur image norme V_V du point saillant est comparée à la partie basse fréquence du vecteur image norme W_V de la région mémorisée (figure 4).

Nous constatons que la variabilité des points visités diminue par rapport aux deux autres modes d'exploration. Les saillances relatives varient entre 0,6 et 1,0. Le pourcentage des points ayant une saillance relative comprise entre 0,9 et 1,0 est de 26%. Le pourcentage des points ayant une saillance relative comprise entre 0,8 et 0,9 est de 41%. Les points les plus visités ont une saillance relative comprise entre 0,7 et 1,0.

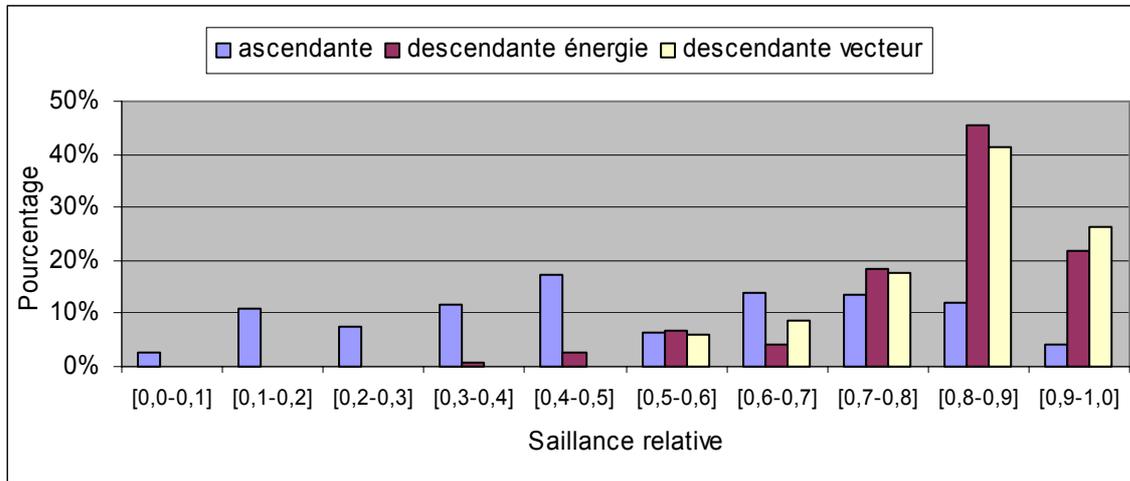


Figure 4. La nature des points explorés suivant les trois modes d'exploration

La sélection des points saillants dans les deux modes d'exploration descendante permet au système de ne visiter que les points intéressants pour l'action qu'il effectue. Cette sélection permet de réduire d'une façon significative le nombre de points visités. Il passe de 222 points dans l'exploration ascendante à 111 points dans l'exploration TDE et à 34 points dans l'exploration TDV (voir figure 5).

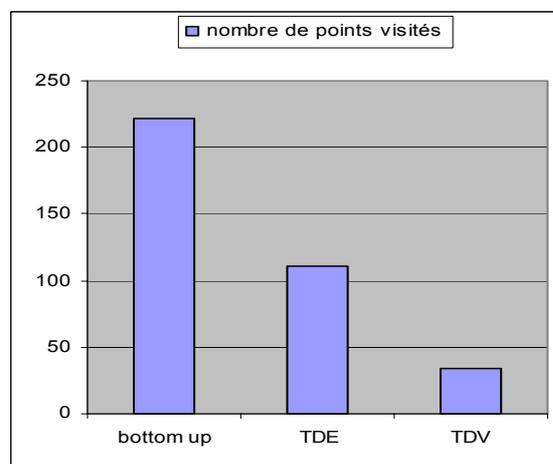


Figure 5. Le nombre de points explorés pour chaque mode d'exploration.

4.2. Résultats d'exploration

Nous avons constaté qu'au cours des différentes expériences à la recherche de visages dans des scènes visuelles effectuées que ces derniers avaient toujours des saillances relatives supérieures à 0,8. Nous avons alors décidé de prendre cette saillance comme seuil de reconnaissance. Cette méthode permet de calculer le taux de reconnaissance qui correspond aux points visités qui ont des scores de reconnaissance supérieurs à 0,8.

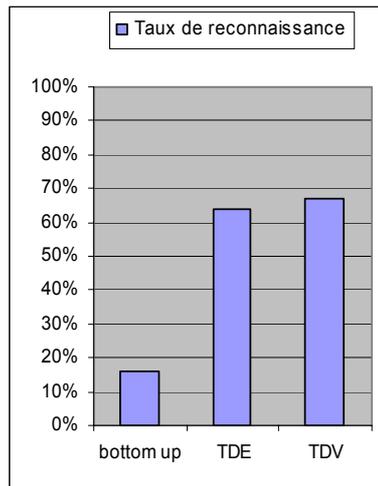


Figure 6. Le taux de reconnaissance obtenus par les trois méthodes d'exploration.

Nous constatons que le taux de reconnaissance en exploration ascendante est de 16% alors que le taux de reconnaissance en exploration descendante énergie est de 64% et en exploration descendante vecteur est de 67% (voir figure 6).

Pour vérifier l'exactitude des taux de reconnaissance obtenus. Une opération consiste à prendre tous les points visités qui ont des taux de reconnaissance supérieurs à 0,8 et vérifier dans l'image d'origine s'ils correspondent bien à des visages. On a appliqué cette opération aux différentes méthodes d'exploration mentionnées précédemment. La figure 7 montre le résultat d'une telle opération.

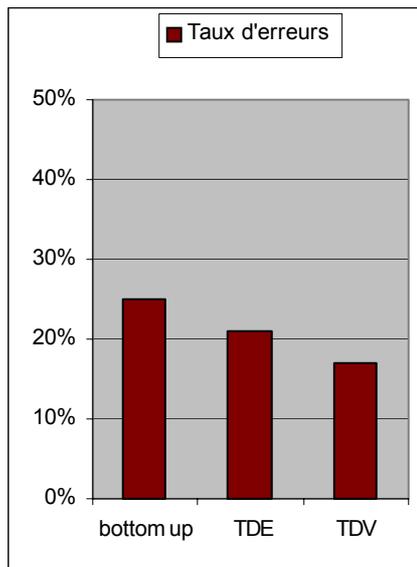


Figure 7. Le taux d'erreurs commis par le système selon les trois méthodes d'exploration.

En exploration ascendante le système a commis 25% des d'erreurs alors qu'il n'a commis que 21% d'erreurs en exploration descendante énergie et 17% en exploration descendante vecteur.

4.3. Robustesse du système vis-à-vis de la luminance des scènes

Les systèmes de vision artificielle doivent reconnaître leur environnement quelles que soient les conditions d'éclairage (robot, séquences vidéo, ..). Une des propriétés essentielles doit donc être sa robustesse vis à vis des conditions d'éclairage et ceci est encore plus important au cours d'une tâche de reconnaissance. Le suivi d'un objet dans une séquence d'images vidéo par exemple ne doit pas être altéré par des modifications de la luminance. Le système décrit dans cet article présente cette propriété.

En effet, l'utilisation des filtres de Gabor permet de munir le système de cette propriété. Ces filtres sont naturellement assez insensibles à la luminance. Ils sont sensibles aux contours, donc aux différences d'éclairage (contraste) et non à la valeur absolue de la luminance.

Nous avons testé notre système pour l'exploration et la recherche d'objets dans des scènes visuelles représentant une scène d'intérieur photographiée dans différentes conditions d'éclairage. La figure 8 montre des images extraites de cette séquence d'images.

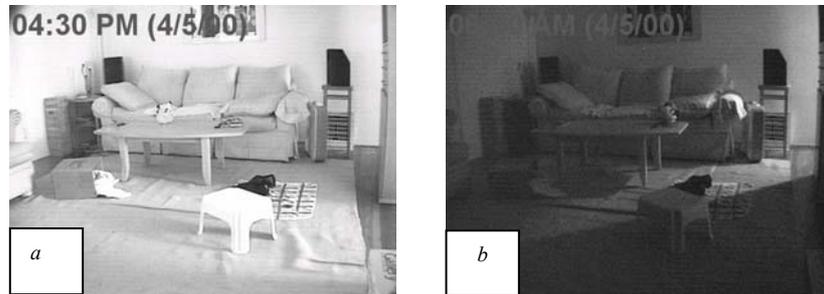


Figure 8. Deux exemples des images utilisées pour tester la robustesse du système par rapport à la luminance de la scène. L'image a représente l'image qui a le plus grand degré de luminance et l'image b celle qui a le plus bas degré de luminance.

Nous avons utilisé le mode d'exploration TDE pour rechercher des zones mémorisées sur une image de luminance moyenne de 151,91 dans des images équivalentes mais dont la luminance varie de 69,24 à 185,69. La figure 9 montre la variation du score de reconnaissance en fonction de la luminance pour les régions similaires et le résultat pour des régions non similaires. Les résultats pour les régions similaires et les régions non similaires ont un faible chevauchement permettant de fixer un seuil de discrimination.

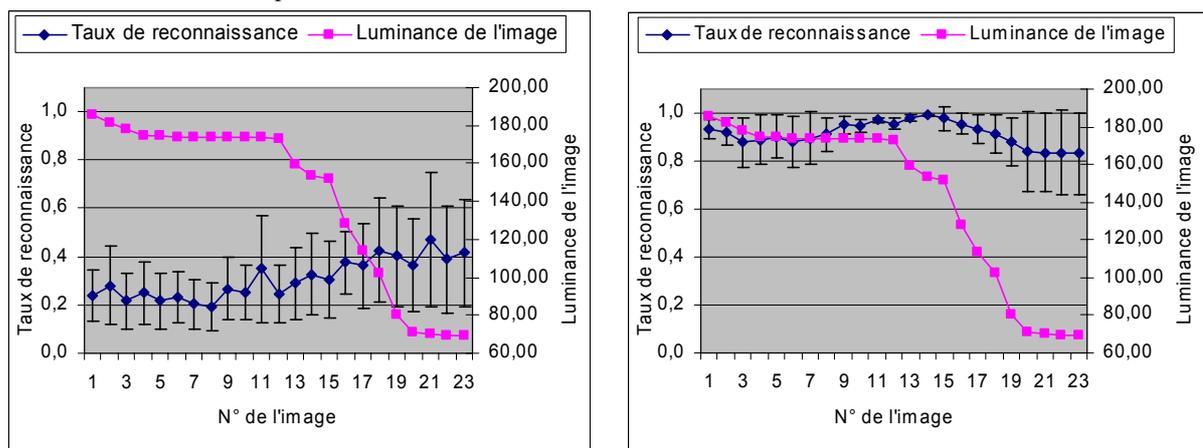


Figure 9. Variation de taux de similitude pour régions non similaires (figure de gauche) et pour régions similaires (figure de droite) par rapport à la variation de luminance.

5. Discussion et conclusion

Le système présenté dans ce travail est fondé sur deux principes :

- La sélection des points saillants d'une scène visuelle qui permet de guider la saccade visuelle du système vers ces points. Cette sélection permet de concentrer le traitement sur ces régions et ainsi réduire le temps de calcul.
- La fusion des informations de bas niveau avec des informations de haut niveau qui permettent de sélectionner parmi ces régions saillantes celles qui sont utiles au but recherché.

Le système présenté montre qu'une architecture fondée sur ces deux principes permet de mieux explorer les scènes visuelles.

L'identification des points saillants dans le système présenté ne se fonde pas sur le calcul d'une carte de saillance de la scène entière comme c'est le cas dans plusieurs travaux [21] [4] [5]. L'identification des points

saillants est effectuée sur des cartes de saillances limitées au champ visuel du système et calculées uniquement en basse fréquence. Ces dernières ne sont pas mémorisées car le système ne mémorise qu'une liste de coordonnées des points qui sont susceptibles de l'intéresser pour accomplir son but. Il ne mémorise aucune autre information. Ces coordonnées vont permettre au système d'aller chercher dans la scène les informations voulues en temps voulu. Il se sert de la scène visuelle comme une mémoire externe [22].

Le système présenté peut effectuer différentes tâches sur différentes scènes visuelles. En effet, l'architecture proposée est indépendante du type des images explorées, des buts recherchés ou de la taille de la scène visuelle.

Notre but a été de réaliser un système de vision exploratoire capable d'effectuer les différentes tâches en temps réel. Ce but a limité le choix des directions préférentielles du banc de filtre de Gabor. Nous avons décidé de prendre quatre directions préférentielles (la verticale, l'horizontale et les deux diagonales), ce choix n'est pas sans conséquences. En effet, nous perdons de l'information sur la scène après le filtrage qui pourrait nous empêcher de reconstruire la scène visuelle. Mais comme notre but n'est pas reconstruire la scène, ceci n'a aucune influence sur le but poursuivi. En effet, on a montré au cours des différentes expériences que les quatre axes suffisent pour le but poursuivi.

Le système décrit dans cet article n'est pas invariant en rotation ou au changement d'échelle. L'opération de reconnaissance est une opération de "pattern-matching" entre la partie mémorisée et la partie explorée. Cette opération ne permet pas de reconnaître des objets à des échelles différentes ou dans des orientations différentes. Une amélioration au niveau du choix des orientations préférentielles du banc de filtres de Gabor pourrait résoudre ce problème d'invariance en rotation. En effet, en sachant le nombre de degré de rotation de l'objet en question, une rotation des directions préférentielles permettrait de résoudre ce problème. Cette solution se limiterait à une orientation bi-dimensionnelle. Une autre approche a été utilisée par Rao et Ballard [23] pour reconnaître des objets dans des images à différentes orientations. Cette architecture utilise un vecteur signature de la région apprise qui se compose de réponse d'un ensemble de filtres de dérivées de gaussienne.

Nous avons montré dans cet article que l'identification des points d'intérêt dans une scène visuelle permet au système de vision artificielle de concentrer les traitements sur ces régions, et ainsi de réduire le temps de calcul. L'utilisation d'une information descendante améliore cette la sélection de ces régions et ainsi permet au système de vision artificielle de mieux explorer les scènes visuelles en fonction du but recherché.

Un pas supplémentaire dans ce travail serait d'équiper le système d'une architecture d'apprentissage par renforcement [24]. Un apprentissage incrémental des différentes parties de l'image pourraient améliorer les taux de reconnaissance obtenus.

6. Références

- [1] Marr D., *Vision: A computational investigation into the human representations and processing of visual information*, Freeman, San Francisco, 1982.
- [2] Y. Aloimonos, I. Weiss and A. Bandyopadhyay, "Active Vision", 1st International Conference on Computer Vision, 1987, pp. 35-54
- [3] R. Bajcsy, "Active Perception", IEEE, 76 (8), 1988, pp. 996-1005
- [4] L. Itti and C. Koch, "A Saliency-based search mechanism for overt and covert shifts of visual attention", *Vision Research*, 40, 2000, pp. 1489-1506
- [5] R. Milanese, "Detecting Salient Regions in an Image from Biological Evidence to Computer Implementation", Department of Computer Science, University of Geneva, Switzerland, 1993, pp. 176
- [6] A. Chauvin, J. Héroult and C. Marendaz, "Modèle de cartes de saillance par filtres passe-bande orientés couplés", ORASIS'99, Aussois, France, 1999
- [7] G.J. Burton and I.R. Moorhead, "Color and spatial structure in natural scenes", *Applied Optics*, 26, 1987, pp. 157-170
- [8] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells", *Journal of the Optical Society of America A*, 4 (12), 1987, pp. 2379-2394
- [9] J.H. Van Hateren, "Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation", *Journal of Comparative Physiology, A* 171, 1992, pp. 157-170
- [10] D.J. Tolhurst, Y. Tadmor and T. Chao, "Amplitude spectra of natural images." *Ophthalmic & Physiological Optics*, 12, 1992, pp. 229-232

- [11] H.B. Barlow, "Sensory mechanisms, the reduction of redundancy, and intelligence", Proceeding of the National Physical Laboratory Symposium on the Mechanisation of thought process, London, 1959, pp. 537-559
- [12] J.G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", J. Opt. Soc. Am., A2 (7), 1985, pp. 1160-1169
- [13] A. Chéhikian, "Algorithmes optimaux pour la génération de pyramide d'images pass-bas et laplaciennes", Traitement du signal, 9 (4), 1992, pp. 297-307
- [14] Y. Machrouh, J.S. Lienard and P. Tarroux, "Multiscale feature extraction from visual environment in an active vision system", International Workshop on Visual Form 4, Springer Verlag, Berlin, Capri, It, 2001, pp. 388-397
- [15] S. Grossberg, "Cortical dynamics of three dimensionnal form, color and brightness perception : monocular theory", Perception and Psychophysics, 41 (2), 1987, pp. 87-116
- [16] M. Seibert and A. Waxman, "Spreading activation layers, visual saccades and invariant representation for neural pattern recognition systems", Neural Networks, 2, 1989, pp. 9-27
- [17] K. Osawa, "Simulation studies on optical illusion based on a position dependant spread function", Pattern Recognition, 23 (12), 1990, pp. 1361-1366
- [18] G.M. Redding and E.A. Hawley, "Length illusion in fractional Müller-Lyer stimuli: An object-perception approach", Perception, 22, 1993, pp. 819-828
- [19] R.L. Gregory, "Visual illusion", Scientific American, 219, 1968, pp. 66-76
- [20] R.L. Gregory and J.P. Harris, "Illusion-destruction by appropriate scaling", Perception, 4, 1975, pp. 203-220
- [21] L. Itti, C. Gold and C. Koch, "Visual attention and target detection in cluttered natural scenes", Optical Engineering, 40 (9), 2001, pp. 1784-1793
- [22] J.K. O'Regan, "Solving the "Real" Mysteries of Visual Perception: The World as an Outside Memory", Canadian Journal of Psychology, 46 (3), 1992, pp. 461-488
- [23] R.P.N. Rao and D.H. Ballard, "An active vision architecture based on iconic representations", Artificial Intelligence Journal, 78, 1995, pp. 461-505
- [24] A.G. Barto, R.S. Sutton and P.S. Brouwer, "Associative search network; a reinforcement learning associative memory", Biological Cybernetics, 40, 1981, pp. 201-211