

Comparaison des algorithmes SPOMF et FMI-SPOMF appliqués à la détection de têtes de robots

Comparison of SPOMF and FMI-SPOMF matching algorithms in natural unconstrained images and application to robot heads detection

M. Nottale

J.C. Baillie

Ecole Nationale Supérieure des Techniques Avancées

matthieu.nottale@ensta.fr, jean-christophe.baillie@ensta.fr

Résumé

L'ancrage des symboles dans la perception est un problème crucial pour l'intelligence artificielle[4]. L'expérience des "Talking Heads"[6] a obtenu des succès significatifs dans ce domaine. Dans cette expérience, deux caméras sont placées face à un environnement constitué de formes géométriques simples sur un tableau blanc. Les agents associés à chaque caméra sont capables de construire un lexique commun, ancré dans leurs perceptions, pour des notions telles que la couleur, la forme ou la position. Des tentatives récentes[1] d'extension de cette expérience à un environnement non-contraint et à des agents ancrés physiquement dans des robots mobiles ont mis en lumière le problème du partage d'attention: les deux robots doivent être capables de se localiser mutuellement et de se déplacer pour se positionner l'un à côté de l'autre. À cette fin, nous considérons dans cet article le problème qui consiste à rechercher une tête de robot dans un flux d'images et à trouver sa distance et son orientation. Nous présentons une méthode utilisant la dérivée temporelle du flux d'images et des mouvements d'oscillation de faible amplitude de la tête du robot pour effectuer un recalage en translation et en échelle, combinée avec un algorithme de reconnaissance des formes. Nous avons testé les algorithmes SPOMF, FMI-SPOMF et la corrélation croisée.

Mots-clef

Reconnaissance des formes, recalage, SPOMF, FMI-SPOMF

Abstract

Grounding symbol in perception is a key issue for artificial intelligence[4]. The "Talking Heads" experiment [6] is a recent and successful attempt in this direction. In this experiment, two cameras are directed at a simple environment made of colored geometric shapes on a white board. Based on these perceptions, the agents associated with each

camera were successful in establishing a common grounded lexicon for simple notions like color, shape or position. Recent attempts [1] to extend this experiment to embodied robots in an unconstrained environment have stressed the problem of sharing attention: two robots must be able to find each other and to move to a side-by-side position. In this paper, we address the problem of finding a robot head as well as its distance and orientation in images acquired by another robot. We present a technique using temporal differential images and slight oscillations of the robot head to perform a scale and translation registration, used in combination with a pattern matching algorithm. We have used cross-correlation, SPOMF or FMI-SPOMF, and we discuss the obtained results.

Keywords

Pattern matching, image registration, SPOMF, FMI-SPOMF

1 Introduction

L'ancrage des symboles dans la perception est un problème crucial pour l'intelligence artificielle. L'expérience des "Talking Heads"[6] a obtenu des succès significatifs dans cette voie. Dans cette expérience, deux caméras sont placées face à un environnement constitué de formes géométriques simples sur un tableau blanc. Les agents associés à chaque caméra sont capables de construire un lexique commun, ancré dans leurs perceptions, pour des notions telles que la couleur, la forme ou la position. L'utilisation de robots autonomes tel que l'ERS-7 de Sony permet de pousser plus avant la recherche sur l'ancrage des symboles en rajoutant les difficultés liées à la robotique située. L'expérience des "Talking Robots" actuellement menée au laboratoire d'électronique et d'informatique de l'ENSTA suit cette voie et vise à reproduire les "Talking Heads" avec des robots autonomes dans un environnement non-contraint.

Pour que cette expérience fonctionne, il est nécessaire que

deux robots soient capable de se positionner côte à côte et de regarder approximativement dans la même direction, afin d'avoir un champ de vision commun, sans aucune information a priori sur la position relative des deux robots. Les deux robots ont des rôles dissymétriques dans cette expérience: l'un d'eux tient le rôle de "speaker" et choisit et nomme un objet, l'autre joue le rôle de "hearer" et essaye de deviner l'objet choisi par le "speaker". Dans ce cadre, il est possible de simplifier le problème de positionnement en décidant que le "speaker" restera immobile, et que le "hearer" se placera à ses côtés. Le "speaker" peut de plus aider le "hearer" à le localiser, à l'aide de mouvements d'oscillation de sa tête, ce qui le rend facile à trouver en utilisant la dérivée temporelle discrète du flux d'images (différence entre les paires d'images successives). Le "hearer" doit donc localiser la tête du "speaker" dans l'image, trouver sa distance et son orientation, et calculer une trajectoire le plaçant à un de ses côtés, de manière itérative jusqu'à atteindre un positionnement correct.

Dans cet article, nous considérons le problème de la détection d'une tête de robot dans un flux d'images et de la recherche de sa distance et de son orientation. C'est un cas particulier du problème de reconnaissance des formes, assez proche du problème de reconnaissance de visages, qui a été très largement étudié [2] [8] [9]. Dans notre cas, nous pouvons utiliser les informations contenues dans la dérivée du flux d'images pour localiser la position approximative de la tête, en supposant que c'est le seul objet mobile de la scène, ce qui évite l'opération coûteuse en temps de calcul qui consiste à appliquer un algorithme de reconnaissance sur de nombreuses fenêtres dans l'image. De plus, ces images dérivées n'ont pas de fond et peuvent être utilisées pour effectuer un recalage en translation et en échelle avant d'appliquer l'algorithme de reconnaissance, ou même être directement utilisées par cet algorithme.

Dans la partie suivante, nous détaillons le processus d'acquisition des motifs et les trois algorithmes de reconnaissance des formes que nous avons testés: corrélation croisée, SPOMF (Simple Phase-Only Matched Filtering) et FMI-SPOMF (SPOMF of the Fourier-Melin invariant). Puis nous exposons les résultats obtenus.

2 Méthode

Dans cette partie, nous commençons par exposer comment nous interagissons avec les robots, en utilisant le langage URBI développé au laboratoire. Puis nous détaillons le mécanisme d'apprentissage des motifs et les algorithmes de reconnaissance.

2.1 Les robots et l'interfaçage: URBI

Les expériences sont réalisées en utilisant des robots ERS-7 de Sony. Nous interagissons avec ces robots en utilisant un protocole client/serveur, URBI (Universal Robotic Body Interface) développé à l'ENSTA. Ce protocole utilise des connexions TCP entre un serveur sur le robot et des clients tournant sur des plate-formes multiples. Les communi-

cations utilisent un langage simple et puissant permettant de commander les moteurs et d'accéder aux senseurs du robot, en particulier sa caméra embarquée. Au prix d'une latence de quelques millisecondes due à la connexion sans-fil (wifi), il est possible de commander les robots depuis un PC classique, ce qui permet d'utiliser des techniques de programmation avancées (threads), et de bénéficier de plus de puissance de calcul. La bibliothèque client C++ encapsulant URBI permet de définir une fonction qui sera appelée à chaque fois qu'une nouvelle image de la caméra est reçue. Les images peuvent être transmises compressées en jpeg et décompressées par la librairie, pour atteindre 30 images par seconde à la meilleure résolution de la caméra, 208x160.

2.2 Apprentissage

Le but de la phase d'apprentissage est d'acquérir une image et une image dérivée représentatives d'une tête de robot oscillante vue sous différentes orientations.

Les deux robots sont placés l'un en face de l'autre à une distance connue (40 centimètres), dans une position fixée et connue. Pour permettre un apprentissage automatisé, le "speaker" tourne sa tête pour que le "hearer" obtienne des images de la tête sous différents angles. Le speaker tourne sa tête du premier angle pour lequel on souhaite obtenir un motif. L'étape suivante consiste à trouver la région de l'image contenant la tête. A chaque image reçue, les N dernières images dérivées sont moyennées. Le résultat est seuillé et passé par une fermeture morphologique. On isole ensuite les régions connexes, puis on calcule les coordonnées de leurs boîtes englobantes, qu'on agrège avec les boîtes englobantes obtenues aux itérations précédentes: quand deux boîtes sont proches, c'est-à-dire quand leurs centres et leurs tailles sont à une distance relative inférieure à un seuil donné, la plus grande est conservée et son score est incrémenté. Dès qu'une boîte dépasse un score fixé, elle est conservée pour la suite comme la région contenant la tête du robot. Cette phase complexe permet de repérer correctement la région de la tête même en présence d'autres objets se déplaçant sporadiquement dans le flux d'images. L'étape suivante consiste à acquérir un ensemble d'images et d'images dérivées dans cette région. À intervalles aléatoires, les N dernières images et images dérivées sont moyennées et la partie correspondant à la région de la tête est conservée, jusqu'à ce que suffisamment d'images aient été acquises (typiquement une vingtaine). Pour extraire une image représentative de chaque ensemble, la matrice des distances entre toutes les paires d'images est calculée en utilisant le rapport signal pic sur bruit (PSNR) du SPOMF comme fonction de distance (cette mesure est détaillée plus bas). L'image la plus proche de toutes les autres est conservée.

Ce processus est répété pour différentes valeurs d'orientation de la tête. On aboutit à un ensemble de paires d'images et d'images dérivées, associées à une valeur d'angle jouant le rôle de descripteur de classe. Le but du filtre moyen tem-

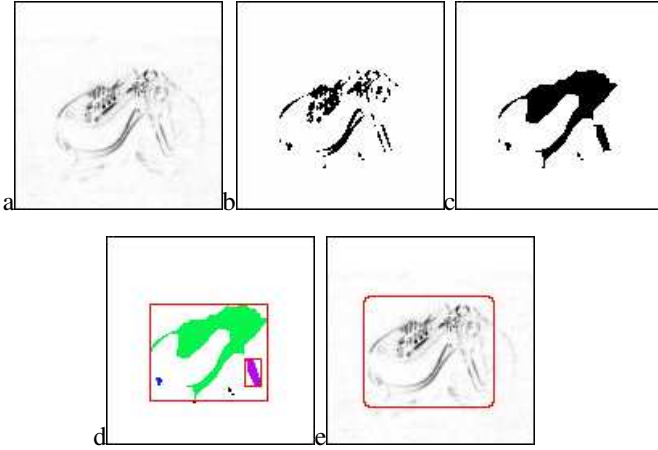


FIG. 1 – Recherche de la région contenant la tête: a: l'image dérivée. b: seuillage. c: fermeture morphologique d: boîtes englobantes. e: fusion des boîtes englobantes

porel est de diminuer la variabilité des images provoquée par les oscillations de la tête. Ce filtre devra être appliqué de manière identique dans le processus de reconnaissance. Il convient de noter que lorsque l'algorithme sera utilisé, le "speaker" aura toujours sa tête orientée droit devant lui, mais sera positionné de manière quelconque par rapport au "hearer" et donc vu sous une orientation quelconque : les images acquises ne correspondent pas exactement à la situation d'utilisation, mais cela n'a pas d'influence car l'axe de rotation de la tête est perpendiculaire au sol.

2.3 Reconnaissance

Le modèle de reconnaissance doit pouvoir, à partir d'un flux d'images, donner à n'importe quel moment et en temps réel la distance, la position et l'angle d'orientation de toutes les têtes de robots trouvées, potentiellement associées avec un facteur de croyance. Les deux robots sont immobiles pendant cette phase, mis à part le mouvement d'oscillation de la tête du "speaker".

La première étape consiste à trouver des régions de l'image pouvant contenir une tête de robot, en utilisant un processus similaire à la phase d'apprentissage. Les N dernières images acquises sont moyennées, le résultat est seuillé et fermé morphologiquement, les régions connexes sont isolées. Chaque boîte englobante d'une région connexe est ensuite testée : la portion correspondante de l'image ou de l'image dérivée est d'abord recalée en translation et en échelle en utilisant l'algorithme détaillé dans la section suivante, puis le résultat est appliqué à l'algorithme de reconnaissance (SPOMF, FMI-SPOMF ou cross-corrélation), qui donne un vecteur de score (un score par motif, soit par angle). Ce vecteur de score est agrégé avec les vecteurs obtenus aux images précédentes pour les régions qui concordent. Quand une région a été considérée un nombre suffisant de fois, elle est marquée comme potentiellement contenant une tête de robot, dont l'orientation est l'angle dont le score est le plus

élevé dans le vecteur de scores.

Nous avons testé trois algorithmes de reconnaissance: corrélation croisée, qui nécessite un recalage en échelle et en translation avant d'être appliqué, SPOMF, qui nécessite un recalage en échelle, et FMI-SPOMF, qui ne nécessite aucun recalage. Nous avons testé ces algorithmes sur les images d'origine et les images dérivées. Dans les parties suivantes, nous détaillons l'algorithme de recalage, puis les algorithmes de reconnaissance et enfin quelques filtres de pré-traitement que nous avons tenté d'appliquer pour améliorer les résultats.

Recalage. L'algorithme de corrélation croisée nécessite un recalage en translation et en échelle, et le SPOMF nécessite un recalage en translation. De nombreux algorithmes ont été proposés pour résoudre ce problème [10]. Nous avons opté pour un algorithme simple basé sur un calcul de moments et utilisant les images dérivées, qui ont la propriété de ne pas avoir de fond.

Étant donnée une image $I(x,y)$, son centre de masse G est donné par:

$$G = \left(\frac{\sum_{x,y} I(x,y) x}{\sum_{x,y} I(x,y)}, \frac{\sum_{x,y} I(x,y) y}{\sum_{x,y} I(x,y)} \right) = (G_x, G_y)$$

et les moments $M_{0,2}$ et $M_{2,0}$ par:

$$M_{2,0} = \frac{\sum_{x,y} I(x,y) |x - G_x|^2}{\sum_{x,y} I(x,y)}$$

$$M_{0,2} = \frac{\sum_{x,y} I(x,y) |y - G_y|^2}{\sum_{x,y} I(x,y)}$$

Deux images de paramètres $(G, M_{0,2}, M_{2,0})$ et $(G', M'_{0,2}, M'_{2,0})$ peuvent être recalées en tradant la deuxième par le vecteur $G' - G$ et en multipliant l'échelle par $\left(\sqrt{\frac{M_{2,0}}{M'_{2,0}}}, \sqrt{\frac{M_{0,2}}{M'_{0,2}}} \right)$. Dans notre cas, les motifs sont tous recalés entre eux dans une phase d'initialisation. Une image ne nécessite donc qu'un seul recalage pour être traitée, au lieu d'un recalage par motif.

Reconnaissance par corrélation croisée. Les deux formules que nous avons utilisées pour le calcul de la corrélation croisée de deux images I et I' sont celles basées sur la norme L_1 :

$$CC_{L1} = \sum_{x,y} |I(x,y) - I'(x,y)|$$

et une version simplifiée de la formule classique:

$$CC = \sum_{x,y} I(x,y) I'(x,y)$$

Cette dernière formule devrait être normalisée par le produit des variances, mais comme dans notre cas les images sont très similaires, cette étape peut être supprimée pour diminuer le temps de calcul. Comme le processus de recalage n'est pas très précis, le calcul de corrélation croisée est effectué en essayant toutes les translations possibles dans une petite fenêtre (5 par 5) et le meilleur score est conservé.

Reconnaissance par le SPOMF. L'algorithme SPOMF (Symmetric Phase-Only Matching Filter) utilise la différence de phase entre les transformées de Fourier de deux images pour les recaler en translation. En utilisant le rapport signal pic sur bruit (Peak Signal to Noise Ratio), il peut être utilisé comme algorithme de reconnaissance. Étant donné une image I et un motif P , on calcule leurs transformées de Fourier F et G , et:

$$Q(u,v) = \frac{F(u,v)G(u,v)^*}{|F(u,v)||G(u,v)|}$$

où le symbole $*$ représente l'opérateur de conjugaison. Le SPOMF S est défini comme la transformée de Fourier inverse de Q . Autrement dit, l'algorithme consiste à calculer la transformée de Fourier de I et P , de calculer leur différence de phase, puis la transformée de Fourier inverse de cette différence. On peut démontrer que si I est une version translaturée de P , le SPOMF est un Dirac aux coordonnées de la translation.

Pour utiliser le SPOMF comme algorithme de reconnaissance, il suffit de calculer le PSNR défini par:

$$PSNR = \frac{(max_{x,y} (S(x,y)) - \bar{S})^2}{\frac{1}{N} \sum_{x,y} (S(x,y) - \bar{S})^2}$$

où N est la taille du SPOMF. Dans nos expériences, un PSNR supérieur à 100 signifie que I est probablement une version translaturée de P . Nous aurions pu directement utiliser la valeur du pic du SPOMF, mais les résultats sont moins marqués. La figure 2 montre un exemple de SPOMF sur des images dérivées d'une tête de robot. Le pic du SPOMF est bien visible autour de $(0,0)$ pour le motif d'angle proche de l'image et s'amenuise au fur et à mesure que l'écart entre l'angle du motif et celui de l'image augmente. Dans notre cas, l'objet comparé au motif par le SPOMF est souvent agrandi par l'algorithme de recalage. Il faut donc éliminer de la transformée de Fourier les composantes de haute fréquence qui n'ont pas de sens pour l'objet. Pour ce faire, la transformée de Fourier est tronquée, en utilisant la taille initiale de l'objet pour déterminer la fréquence la plus élevée qui sera conservée, avant de calculer la différence de phases et la transformée de Fourier inverse de cette dernière. On montre que si l'objet était de taille N , seule la partie centrale de taille $N \times N$ de la transformée de Fourier doit être conservée.

Reconnaissance par le FMI-SPOMF. L'algorithme FMI-SPOMF consiste à appliquer le SPOMF sur l'invariant de Fourier-Melin (FMI) et peut être utilisé pour recaler deux images en rotation, translation et échelle[5].

Le passage en coordonnées log-polaires a la propriété de transformer les rotations et les changements d'échelle en translations [7]. L'invariant de Fourier-Melin d'une image I est défini comme la transformée en coordonnées log-polaires du module de sa transformée de Fourier. Il est invariant en translation, et une rotation ou un changement d'échelle de l'image se traduisent par des translations du

FMI. Appliquer un SPOMF sur le FMI permet donc de recaler en rotation et en échelle deux images. Pour recaler en translation, il suffit d'appliquer les transformations de rotation et d'échelle sur l'une des deux images, et d'appliquer un SPOMF.

Pour utiliser le FMI-SPOMF comme algorithme de reconnaissance, il est possible d'utiliser le PSNR du FMI-SPOMF, ou le PSNR du SPOMF appliqué entre le motif et l'image recalée en rotation et échelle.

Nous avons utilisé les formules suivantes pour la transformée en coordonnées log-polaires:

$$x(r,\theta) = \frac{W}{\sqrt{2}} K^{(r-R)} \sin\left(\theta \frac{\pi}{T}\right)$$

$$y(r,\theta) = \frac{H}{2} + \frac{W}{\sqrt{2}} K^{(r-R)} \cos\left(\theta \frac{\pi}{T}\right)$$

où (W,H) est la taille de l'image d'origine (la transformée de Fourier est donc de taille $(\frac{W}{2},H)$ après suppression de l'information redondante), (R,T) est la taille de l'image transformée, et K un paramètre, qui détermine le plus grand facteur d'échelle que le FMI-SPOMF pourra recaler. Plus K est grand, moins le recalage en échelle sera précis. Nous avons utilisé $K = 1.03$, ce qui permet pour une image de 128 par 128 pixels un recalage d'un facteur 6. Les paramètres de rotation et d'échelle correspondant à une translation (x,y) sont donnés par:

$$echelle = K^x \quad rotation = y \frac{\pi}{T}$$

La figure 3 donne des exemples de FMI-SPOMF, sur les mêmes images que la figure 2. On constate que le pic est beaucoup plus marqué et diminue plus rapidement avec l'écart d'orientation.

De manière similaire au SPOMF, la transformée de Fourier doit être tronquée avant de calculer la différence de phase pour supprimer les composantes sans signification. Mais pour accélérer le calcul, les transformées de Fourier des FMI des motifs sont pré-calculées à l'initialisation. Nous avons donc décidé d'approximer la troncature correcte en coupant la transformée de Fourier du FMI pour $r > R + \frac{\log(\frac{N}{W\sqrt{2}})}{\log(K)}$, ce qui correspond à conserver un cercle de rayon $N\sqrt{2}$ à la place d'un carré de côté N .

Pour éviter les artefacts provoqués par la troncature, un filtre de Hanning est appliqué sur toutes les images avant le calcul de leur transformée de Fourier. On calcule I' pour une image I de taille (w,h) :

$$I'(x,y) = I(x,y) \left(0.5 + 0.5 \cos\left(\frac{2(x-w/2)\pi}{w}\right) \right) \cdot \left(0.5 + 0.5 \cos\left(\frac{2(y-h/2)\pi}{h}\right) \right)$$

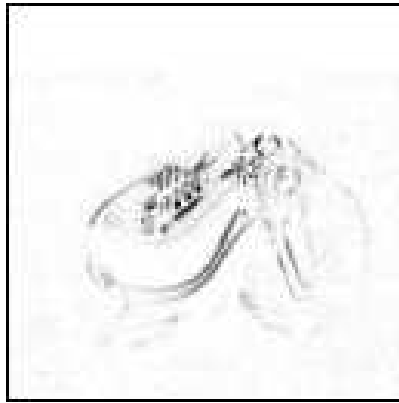


Image d'une tête orientée à -51 degrés.

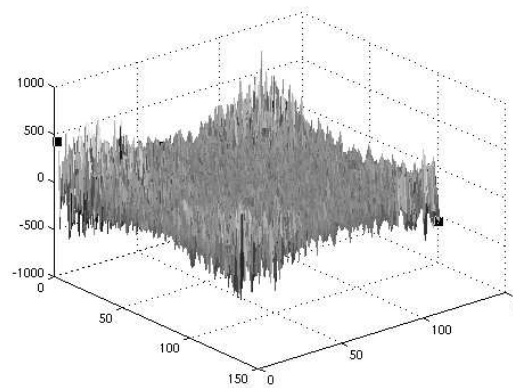
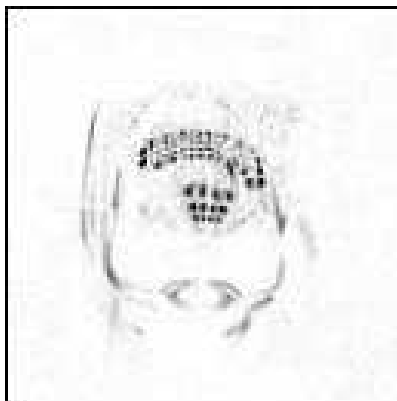
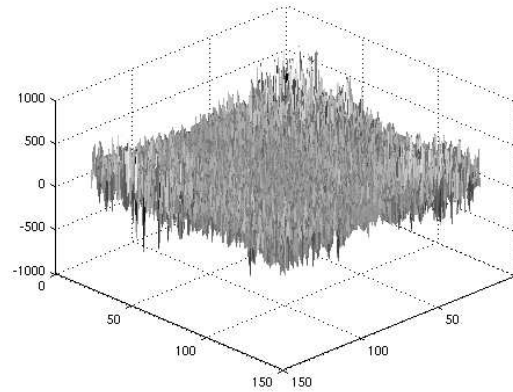
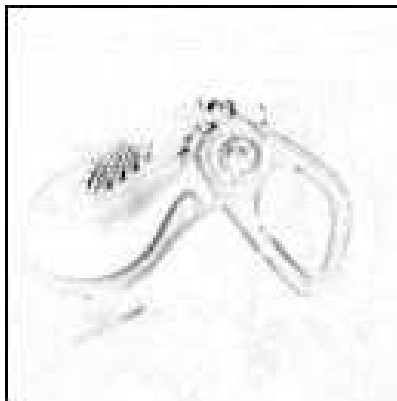
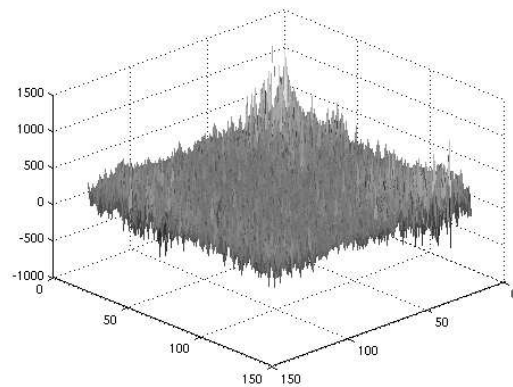
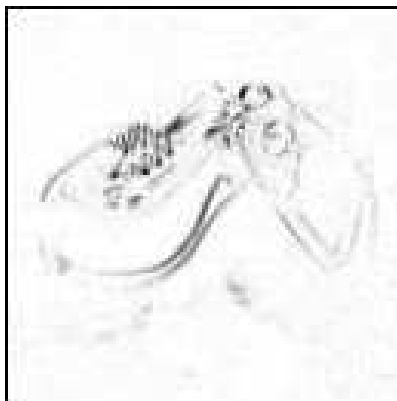
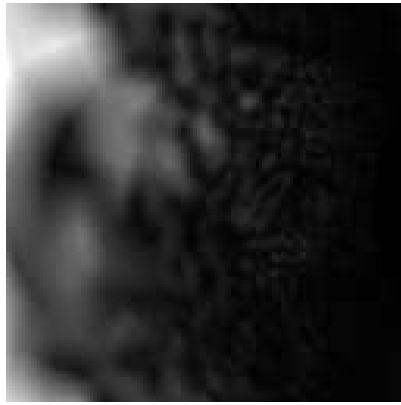


FIG. 2 – Motifs pour les angles -50, -70 et 10, et SPOMF avec l'image à -51°. Les PSNR valent respectivement 267, 94 et 33



FMI d'une tête vue à -51 degrés.

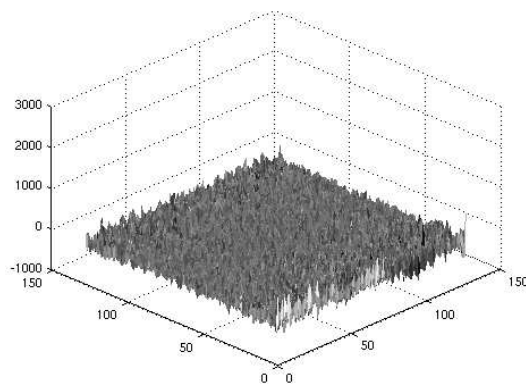
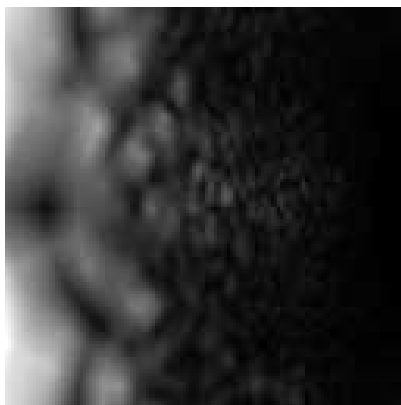
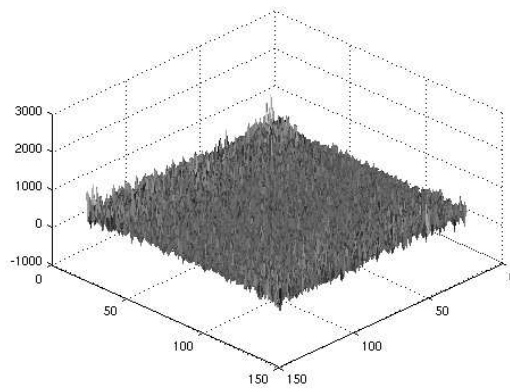
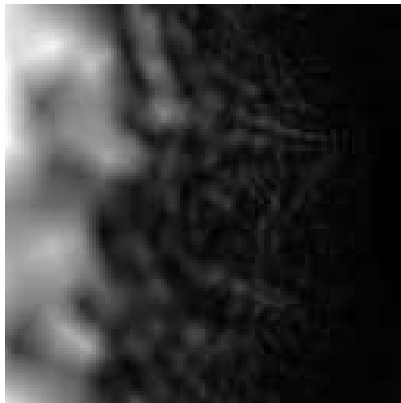
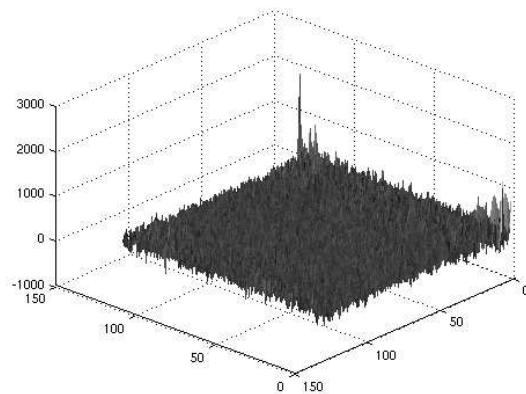
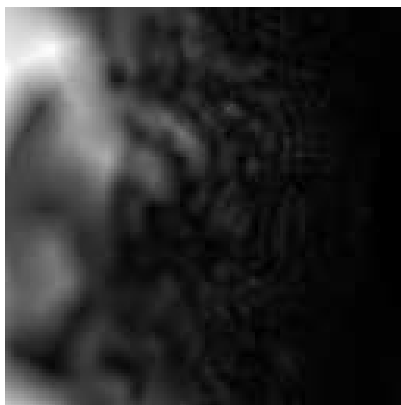


FIG. 3 – FMI pour des motifs d'angles -50, -70, 10 et FMI-SPOMF avec l'objet.

Pré-traitements et autres améliorations. La qualité du recalage est un facteur important pour les performances des algorithmes de reconnaissance. Nous avons essayé d'appliquer divers filtres pour augmenter la stabilité de la forme de la tête sur les images dérivées. La forme elle-même étant plus stable que les intensités relatives des pixels la composant, nous avons tenté d'employer des filtres d'égalisation d'histogramme et de seuillage. Les performances de ces filtres sont détaillées dans la partie résultats.

La présence d'un fond sur les images originales dégrade les performances des algorithmes de reconnaissance. Pour l'éliminer, il est possible d'utiliser les images dérivées pour générer un masque binaire, appliqué sur l'image avant la reconnaissance. Ce masque est créé en seuillant l'image dérivée et en appliquant une fermeture morphologique de taille celle de l'objet, ce qui revient approximativement à prendre l'enveloppe convexe des points de l'objet supérieurs au seuil.

3 Résultats

Dans cette partie, nous détaillons d'abord le protocole et les critères d'évaluation des algorithmes de reconnaissance, ainsi que l'influence de leurs paramètres. Puis nous exposons brièvement pourquoi nous avons abandonné la corrélation croisée. Ensuite nous exposons quelques résultats sur les performances de l'algorithme de recalage et son influence sur les performances globales. Enfin nous évaluons plus en détail les performances du SPOMF et du FMI-SPOMF.

3.1 Protocole expérimental

Pour pouvoir comparer les différents algorithmes, il faut que ceux-ci soient appliqués sur les mêmes données. Pour ce faire, nous avons enregistré des séquences vidéos. Chaque séquence contient:

- 40 secondes d'images d'une tête de robot pour toutes les orientations dans $[-80,80]$ par pas de 2 degrés, à 40 centimètres;
- 10 secondes d'images pour 64 orientations choisies aléatoirement dans $[-80,80]$, à des distances de 40, 80 et 120 centimètres.

La première partie est utilisée pour l'apprentissage, la seconde pour les tests.

Un deuxième jeu de séquences a été enregistré, avec le "speaker" tourné de 180 degrés, afin d'obtenir des images pour tous les angles possibles.

Chaque séquence est enregistrée avec un fond différent et les tests sont effectués en utilisant une séquence différente pour l'apprentissage et les tests, afin d'être dans des conditions les plus proches possibles de l'utilisation de ces algorithmes.

Pour chaque test, nous mesurons les propriétés suivantes:

- temps de calcul moyen par image;

- pour différentes marges d'erreurs admissibles sur la valeur de l'orientation retournée (typiquement 5, 10 et 30 degrés):

- taux de succès moyen: un test pour un angle donné est considéré comme ayant réussi si au bout des 10 secondes, la tête trouvée avec le meilleur score à une orientation correcte,
- nombre moyen d'images nécessaires pour obtenir l'orientation correcte,
- précision moyenne de la mesure de position et de distance,
- taux de faux-positifs.

Ces indicateurs permettent de savoir si l'algorithme se comportera de manière satisfaisante lorsqu'il sera utilisé dans l'expérience des Talking Robots.

3.2 Évaluation de l'algorithme de corrélation croisée.

Des tests préliminaires ont montré que l'algorithme de corrélation croisée ne convenait pas pour plusieurs raisons: le pic obtenu par la mesure de corrélation croisée est généralement très large[3]. Ainsi le score obtenu avec le motif d'orientation correcte, mais légèrement mal recalé est souvent très proche du score obtenu avec un mauvais motif. De plus, du fait de la nécessité d'appliquer la corrélation croisée en testant plusieurs valeurs de translation, cette méthode est plus lente d'un ordre de grandeur que les méthodes basées sur Fourier. Nous avons donc abandonné la corrélation croisée et nous nous concentrerons dans la suite de cet article sur le SPOMF et le FMI-SPOMF.

3.3 Évaluation des performances du recalage et de ses conséquences.

Les performances de l'algorithme de recalage basé sur les moments semblent être en pratique assez mauvaises: de légères distortions ou variations d'intensité de l'image provoquent des différences importantes sur les moments $M_{0,2}$ et $M_{2,0}$, et donc d'importantes erreurs de recalage en échelle. Mais l'impact de ces erreurs sur les performances globales est limité, pour deux raisons:

Tout d'abord, le SPOMF est assez résistant aux erreurs de recalage, comme le montre la figure 4, qui donne le logarithme du score obtenu avec le SPOMF entre une image de tête de robot orientée à -51° zoomée d'un facteur 0.8 1 et 1.2, et 33 motifs dans l'intervalle $[-80,80]$. Dans les trois cas le résultat est correct à 20 degrés près.

De plus, en cas d'erreur de recalage, le score donné par le SPOMF sera le même pour tous les motifs et d'une valeur plus faible qu'en cas de recalage correct. Comme le score est sommé d'image en image, le score final a de bonnes chances d'être correct même si un pourcentage important des images sont mal recalées.

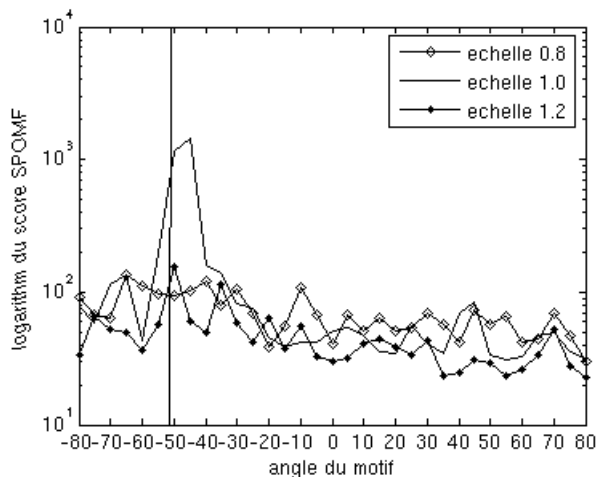


FIG. 4 – Score du SPOMF entre des motifs et un objet à différents zooms.

3.4 SPOMF

Comme dit précédemment, le SPOMF a été testé sur les images d'origines et les images dérivées. Nous présentons les résultats de ces deux tests séparément.

SPOMF sur les images dérivées. En se limitant au domaine $[-80, 80]$, avec un motif tous les 4° et un filtre de normalisation d'histogramme, nous obtenons les résultats suivants:

distance(cm)	40		80		120	
marge (degrés)	15	30	15	30	15	30
taux de succès	0.68	0.73	0.77	0.81	0.53	0.70

Supprimer le filtre dégrade les résultats d'environ 10%. Modifier le nombre d'images moyennées n'a qu'un effet marginal sur les résultats. L'erreur moyenne sur la distance est de 10% et de 5 pixels sur la position, ce qui convient pour notre usage de l'algorithme.

Le temps moyen nécessaire pour trouver l'orientation correcte est de 14 images, soit environ deux secondes.

Le temps de calcul par image est de 110 millisecondes sur un pentium 4 cadencé à 2.8 GHz: l'algorithme peut être utilisé en temps réel.

Quand les tests sont effectués avec des motifs couvrant 360° , les résultats sont convenables à 40 centimètres, mais le taux de succès tombe en dessous de 50% pour toutes les distances supérieures: toutes les images d'orientation dans $[-80, 80]$ sont systématiquement reconnues comme correspondant à des motifs du dos de la tête. Une explication possible est que les informations permettant de discriminer entre le devant et l'arrière de la tête se trouvent dans des hautes fréquences qui ne sont pas disponibles quand l'objet est trop petit. Une étude plus approfondie serait nécessaire pour confirmer cette hypothèse.

SPOMF sur les images d'origine. Un phénomène similaire se produit lorsque le SPOMF est appliqué sur les images d'origine: les résultats sont convenables (0.92, 0.90, 0.70 pour une marge de 30° à respectivement 40., 80 et 120 centimètres) lorsqu'on se restreint à l'intervalle d'orientations $[-80, 80]$, mais tombent en dessous de 50% sur 360° , très certainement pour la même raison.

3.5 FMI-SPOMF

FMI-SPOMF sur les images dérivées. Le FMI-SPOMF appliqué sur des séquences d'images avec des orientations couvrant les 360° donne les résultats suivants:

distance (cm)	40		80		120	
marge (degrés)	15	30	15	30	15	30
taux de succès	1	1	0.75	0.93	0.48	0.65

Le taux de succès est trop bas pour une utilisation pratique. Une grande partie de l'information est perdue en ne considérant que les images dérivées. Comme elle ne contiennent que la forme de la tête, la plus grande partie de l'énergie du spectre de Fourier se trouve dans les hautes fréquences.

FMI-SPOMF sur les images d'origine. Le FMI-SPOMF appliqué sur les images originales avec des orientations couvrant les 360° donne les résultats suivants:

distance (cm)	40		80		120	
marge (degrés)	15	30	15	30	15	30
taux de succès	0.89	0.92	0.84	0.95	0.81	0.92

La mesure de distance est précise à 20% près et la mesure de position à 10 pixels près. Le temps de calcul est de 170 millisecondes par image à 40cm et diminue avec la taille de l'objet. L'intervalle entre les images étant de 160 millisecondes, environ 6% des images sont ignorées, ce qui a pour seul effet d'augmenter le temps de détection, qui est de l'ordre d'une seconde et demi en moyenne pour obtenir une orientation correcte.

Utiliser la technique de masquage décrite dans la partie 2.3.5 diminue le taux de succès de 20%. La principale raison est l'instabilité du masque qui coupe souvent des portions de l'objet.

Augmenter le nombre de motifs à un tous les 2° n'a pas d'effet notable sur le taux de succès. Le diminuer à un tous les 8° les diminue significativement: dans le pire des cas l'objet et le motif le plus proche diffèrent d'une rotation de 4° (rotation dans l'espace et non dans le plan de l'image), ce qui provoque des différences suffisantes pour que le PSNR obtenu soit faible et de l'ordre de grandeur de celui obtenu avec les autres motifs d'orientation incorrecte. Une valeur de 4° entre les motifs semble être un bon compromis entre le temps de calcul et les performances.

4 Conclusion

Nous avons montré que le FMI-SPOMF peut être utilisé avec succès pour localiser une tête de robot et calculer sa

distance et son orientation en temps-réel. De nombreuses voies d'amélioration sont envisageables: Il serait possible d'ajouter des motifs pour différentes orientations verticales de la tête pour pouvoir la reconnaître dans toutes les positions. On pourrait aussi envisager d'appliquer le FMI-SPOMF sur une hiérarchie de fenêtres afin de se passer des images dérivées, mais en gardant la contrainte du temps-réel. Enfin, nous comptons utiliser ces travaux comme base d'une plate-forme plus générale d'apprentissage et de reconnaissance qui sera utilisée par les robots pour apprendre leur environnement visuel.

Références

- [1] Jean-Cristophe Baillie. Grounding symbol in perception with two interacting autonomous robots. *Proceedings of the Fourth international workshop on Epigenetic Robotics*, 2004.
- [2] Eric Hjelmas and Boon Kee Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83:236–274, 2001.
- [3] Thomas M. Lehmann. A two-stage algorithm for model-based registration of medical images. In *Proceedings of the 14th International Conference on Pattern Recognition-Volume 1*, page 344. IEEE Computer Society, 1998.
- [4] John Searle. Minds, brains and programs. *Behavioral and brain sciences*, 3:417–421, 1980.
- [5] Qin sheng Chen, Michel Defrise, and F. Deconinck. Symmetric phase-only matched filtering of fourier-melin transforms for image registration and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(4), 1994.
- [6] Luc Steels. The origin of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103:133–156, 1998.
- [7] G. Wolberg and S. Zokai. Robust image registration using log-polar transform. *IEEE Proceedings of International Conference on Image Processing*, 2000.
- [8] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 2002.
- [9] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.
- [10] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.