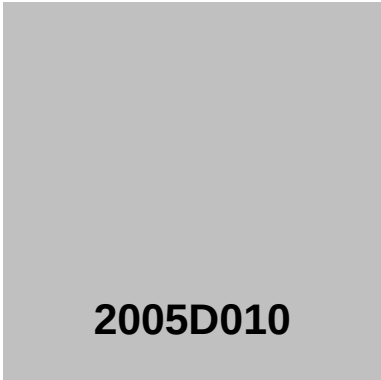




Développement d'algorithmes de reconnaissance de rond-points dans des images satellitaires et évaluation de leurs performances

Jean-Baptiste Bordes



2005D010

octobre 2005

Département Traitement du Signal et des Images
Groupe Traitement et Interprétation des Images

Développement d'algorithmes de reconnaissance de rond-points dans des images satellitaires et évaluation de leurs performances

Jean-Baptiste Bordes

24 octobre 2005

Table des matières

1	Introduction	3
2	Travaux antérieurs sur le sujet	3
2.1	Reconnaissance d'objets à l'aide de moments invariants.	3
2.2	Indexation à l'aide de points d'intérêts	4
2.3	Construction d'un graphe avec attributs	5
3	Travail effectué	6
3.1	Descripteur utilisé	6
3.1.1	Présentation du descripteur	6
3.1.2	Avantages et inconvénients du descripteur	9
3.2	Utilisation des moments de Zernike	9
3.3	Evaluation des descripteurs	10
3.3.1	Base de données utilisée	10
3.3.2	Application des descripteurs sur la base de données	10
4	Mise en oeuvre du descripteur sur la base de données	13
4.1	Résultats obtenus	13
4.2	Analyse des résultats	13
5	Conclusion, améliorations possibles et perspectives	18
A	Moments de Hilbert	20
B	Moments de Zernike	21
C	Autre résultat de détection	24

RÉSUMÉ :

Dans cet article, nous présentons un nouvel algorithme de détection d'objets dans des images satellitaires. Notre méthode nécessite un apprentissage des structures à détecter à partir d'images exemples, mais ne nécessite aucune connaissance à priori sur ces structures. Il est donc aisé pour l'utilisateur de rajouter au fur et à mesure différents types d'objets à détecter. L'algorithme employé utilise un descripteur radial et un descripteur angulaire invariants par rotation. Les caractéristiques angulaires sont extraites en chaque point de l'image en calculant les premiers coefficients de Fourier sur les signaux unidimensionnels définis par les niveaux de gris d'une série de N cercles de rayons inférieurs à un rayon maximal R à fixer, ces cercles ayant pour centre le point où l'on fait le calcul. Seuls les modules des coefficients sont conservés pour que ces caractéristiques soient invariantes par rotation. Les caractéristiques radiales sont extraites en calculant en chaque point les premiers coefficients de Fourier sur une série de M signaux unidimensionnels définis par les niveaux de gris sur les rayons d'angles équirépartis du cercle de rayon R . Contrairement aux caractéristiques précédentes, la phase est un élément important qu'il est intéressant de conserver. De plus, pour avoir une relative invariance par rotation, il convient, pour comparer deux objets, de calculer le meilleur alignement entre les différents segments. Cette méthode nécessite que l'on définisse un centre pour chaque objet et que les paramètres (nombre de cercles, nombre de rayons, nombre de coefficients de Fourier conservés pour chaque cercle et pour chaque rayon, et surtout taille du disque d'analyse R) soient correctement définis. Pour chaque imagerie exemple, dont on s'assure que le centre coïncide avec celui de l'objet, on calcule les caractéristiques au centre de l'imagerie. Pour l'image dans laquelle on souhaite détecter les objets, les caractéristiques sont calculées en chaque point et sont comparées par une distance euclidienne aux caractéristiques calculées pour les images exemples. L'efficacité de ces caractéristiques est évaluée sur une base d'images de rond-points, et est comparée à celle des moments de Zernicke, qui sont des caractéristiques assez souvent utilisées pour la détection d'objets. Les rond-points sont en effet des objets ayant une variabilité modérée et une occurrence relativement élevée dans le paysage européen, ils sont donc bien adaptés pour une première évaluation. Les rond-points retenus pour l'évaluation proviennent d'images de Stockholm, Toulouse et Bagdad. Il apparaît que les performances de ces nouvelles caractéristiques surpassent celles des moments de Zernicke, mais que les performances sont très dépendantes du "bruit" entourant l'objet et du bon ajustement des paramètres.

1 Introduction

L'indexation d'images satellitaires est un domaine en expansion. En effet, la quantité d'images qui nous parvient des différents satellites en orbite autour de la Terre augmente plus vite que notre capacité à les traiter. Il est donc crucial de parvenir à élaborer des algorithmes robustes permettant d'en extraire l'information et de l'indexer. L'indexation est l'opération qui consiste à extraire une signature numérique ou textuelle qui décrit le contenu de manière précise et concise afin de permettre une recherche efficace dans une base de données. Le but du stage est de proposer des algorithmes génériques d'indexation des images permettant de détecter, après apprentissage, des objets spécifiques. Pour cela, nous allons présenter dans ce rapport un nouveau descripteur et comparer ses performances par rapport à un autre descripteur classique. Les images utilisées pour l'évaluation des algorithmes sont des images SPOT5 d'une résolution de 2,5 mètres par pixel, et les objets que nous avons employés sont principalement des ronds-points car ils ont une variabilité modérée et une occurrence assez élevée dans le paysage européen.

2 Travaux antérieurs sur le sujet

Très peu de travaux ont été menés dans le domaine de l'indexation d'images satellitaires à cette résolution. La plupart de travaux qui ont été effectués jusqu'à présent en reconnaissance d'objets dans des images satellitaires utilisent de l'information a priori et ne nécessitent pas d'apprentissage. Mon travail de bibliographie s'est donc essentiellement concentré sur des méthodes très générales de reconnaissance d'objets dans des images. Ceux-ci ont été introduits pour répondre à des tâches très diverses comme la classification de chromosomes, l'identification de cibles ou l'analyse de scènes. Le but de ces méthodes est de définir un lot de caractéristiques pour représenter l'image tout en réduisant sa dimension [10]. Après que les caractéristiques aient été calculées, elles sont utilisées par une règle de classification qui assigne un label à l'image. Suivant les applications, on exigera ou non de ces caractéristiques certaines propriétés d'invariances par translation, rotation, changement d'échelle ou par d'autres transformations. Voici ici un bref descriptif des principales approches qui semblent pertinentes pour notre problème :

2.1 Reconnaissance d'objets à l'aide de moments invariants.

Les invariants algébriques sont obtenus à partir de quotients ou de puissances de moments, ils sont utilisés comme caractéristiques pour faire de la reconnaissance d'objets.

Moments de Hilbert Soit $I(x, y)$ le niveau de gris d'un pixel de l'image I , on définit le moment d'ordre $p + q$ ($p, q > 0$) de l'image I par :

$$m_{pq} = \int_{R^2} x^p y^q I(x, y) dx dy$$

A partir de ces valeurs, on calcule les 7 invariants de Hu [2], qui sont invariants par translation, rotation et changement d'échelle. Les formules détaillées sont données en annexe 1. Notons que le moment m_{pq} correspond à la projection de la fonction $I(x, y)$ sur la base de polynômes $x^p y^q$. Mais cette base n'est pas orthogonale. Par conséquent, reconstruire une image à partir des moments est compliqué et coûteux en temps de calcul. De plus, cela implique que l'information contenue par les quantités m_{pq} est quelque peu redondante. Nous allons voir à présent les polynômes de Zernike, qui présentent l'avantage, par rapport au cas présent, d'être orthogonaux.

Moments de Zernike Les moments de Zernike sont construits en utilisant un ensemble de polynômes complexes qui forment un ensemble orthogonal complet défini sur le disque unité [2], [5].

Les polynômes de Zernike sont donnés en coordonnées polaires par :

$$V_{mn}(r, \theta) = R_{mn} \exp(jm\theta)$$

où n est un entier positif ou nul, et m est un entier positif ou négatif inférieur en valeur absolue à n . $R_{nm}(\rho)$, le polynôme radial, est défini par :

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)!}{s! (\frac{n+|m|}{2} - s)! (\frac{n-|m|}{2} - s)!} \rho^{n-2s}$$

Notons que $R_{n-m}(\rho) = R_{nm}(\rho)$.

Les moments de Zernike A_{mn} sont les projections de l'image sur ces polynômes. On calcule qu'une rotation d'angle α de l'image entraîne une multiplication de A_{mn} par $\exp(-jm\alpha)$. Ainsi, $|A_{mn}|$ peut-être pris comme caractéristique invariante par rotation.

2.2 Indexation à l'aide de points d'intérêts

Cette méthode consiste à extraire des points d'intérêts (avec par exemple un détecteur de Harris), puis à calculer des invariants à la rotation en ces points [8], [9], [3], [7], [6]. Pour caractériser un point d'intérêt, on utilise des dérivées de niveaux de gris calculés jusqu'à l'ordre 3. Afin d'obtenir des dérivées indépendantes de la rotation, la direction de calcul des dérivées est rapportée à celle du gradient. On obtient ainsi un vecteur d'invariants. Pour détecter un objet, on effectue des appariements entre les points de l'image de test avec les objets de l'apprentissage.

Le point fort de cette technique est d'être robuste aux changements de fond et aux occultations dans les petites régions autour des points d'intérêt. Cependant, l'information fournie par les descripteurs est très locale et ne semble pas en

mesure d'apporter suffisamment d'information pour reconnaître une structure comme un rond-point.

2.3 Construction d'un graphe avec attributs

Cette méthode consiste à réduire l'image à un ensemble de primitives (segments et arcs de cercle par exemple). A partir de cette représentation simplifiée de l'image, on construit un graphe dont les nœuds sont des primitives et les arcs sont des relations entre ces primitives (angle d'intersection, distance etc.). Une fonction de similarité entre deux images est calculée en comparant les deux graphes avec une métrique bien définie. A partir d'images d'apprentissage d'une structure telle qu'un rond-point, on construit les graphes correspondants et on cherche le plus grand sous-graphe commun qui sera utilisé pour la détection dans une plus grande image. Cette méthode commence à être adoptée pour la reconnaissance de rond-points et de ponts [4] [1]. Elle semble très prometteuse mais elle est cependant très lourde à mettre en œuvre.

Conclusion de cette partie : Au vu de ces trois différentes approches pour aborder le problème, la première semble être la meilleure façon d'obtenir des résultats satisfaisants dans les délais impartis par le stage. En conservant l'idée d'extraire dans un vecteur de caractéristiques des informations géométriques sur l'image, j'ai souhaité créer un nouveau descripteur en m'affranchissant de la contrainte de prendre les coefficients provenant de la décomposition de la fonction image sur une base. J'ai souhaité aussi évaluer son efficacité sur une base de données de ronds-points que j'ai créée à partir des images urbaines à 2,5 mètres de résolution qui étaient disponibles.

3 Travail effectué

L'essentiel du travail effectué consiste en la création et la mise en œuvre d'un nouveau descripteur appliqué ensuite sur une base de données que j'ai créée à partir des images disponibles et contenant essentiellement des ronds-points. Nous exigeons de ce descripteur une invariance par rotation, et nous souhaitons qu'il soit aussi peu dépendant que possible de la radiométrie en capturant essentiellement les caractéristiques géométriques de l'objet. J'ai aussi fait une mise en œuvre de la méthode utilisant les polynômes de Zernike, ce qui permet de faire une comparaison par rapport au nouveau descripteur. Nous allons commencer par une présentation de ce descripteur, puis nous verrons sa mise en œuvre.

3.1 Descripteur utilisé

Le descripteur utilisé dans ce stage consiste en un algorithme très simple mais efficace d'extraction de caractéristiques pour faire de la reconnaissance d'objets quelconques dans des images.

3.1.1 Présentation du descripteur

L'idée principale de ce descripteur réside dans une séparation des composantes radiales et angulaires de l'image. Ainsi, il est composé en fait de deux sous-descripteurs : un descripteur radial et un descripteur angulaire.

Descripteur angulaire Soit une image contenant un objet de centre de gravité O. On trace N cercles de rayons $\frac{kR}{N}$ de centre O et on considère les N signaux unidimensionnels formés par la valeur des pixels situés sur les cercles, l'origine du signal étant choisie arbitrairement. Ensuite, on extrait un vecteur de caractéristiques pour chacun de ces signaux unidimensionnels, notons les $V_1, V_2, \dots, V_N$, où $V_i = (c_1^i, c_2^i, \dots, c_n^i)$. Il existe beaucoup de possibilités dans le choix de ces caractéristiques, mais après beaucoup d'essais, le plus robuste semble être le module des premiers coefficients de Fourier normalisés par la quantité D définie par : $D = \sqrt{\sum_{i=1}^N \sum_{j=1}^N (c_i^j)^2}$. Cela permet en effet d'obtenir des caractéristiques invariantes par rotation car on ne garde pas la phase de la transformée de Fourier.

Pour comparer deux objets entre eux, on peut tout simplement calculer la somme des distances euclidiennes des vecteurs correspondant à chaque cercle, mais on peut également calculer le meilleur alignement avec un algorithme de Viterbi, ce qui permet d'obtenir une bonne robustesse en ce qui concerne les changements de taille des objets, et on peut ainsi avoir une relative invariance par changement d'échelle. Pour obtenir davantage de robustesse, on peut également, au lieu de prendre directement la valeur d'un pixel situé sur un cercle de rayon r et d'angle θ , faire une moyenne des pixels d'angle θ et de rayon compris entre $[r-R/2N, r+R/2N]$. Les paramètres à fixer sont : le nombre de cercles, le rayon

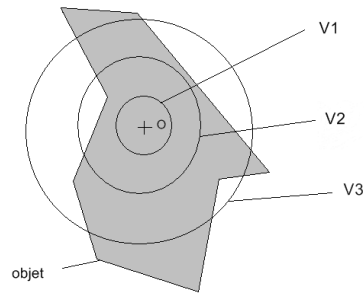


FIG. 1 – schéma explicatif du descripteur angulaire. v_1 , v_2 , v_3 sont les vecteurs de caractéristiques extraits des signaux unidimensionnels définits par les cercles.

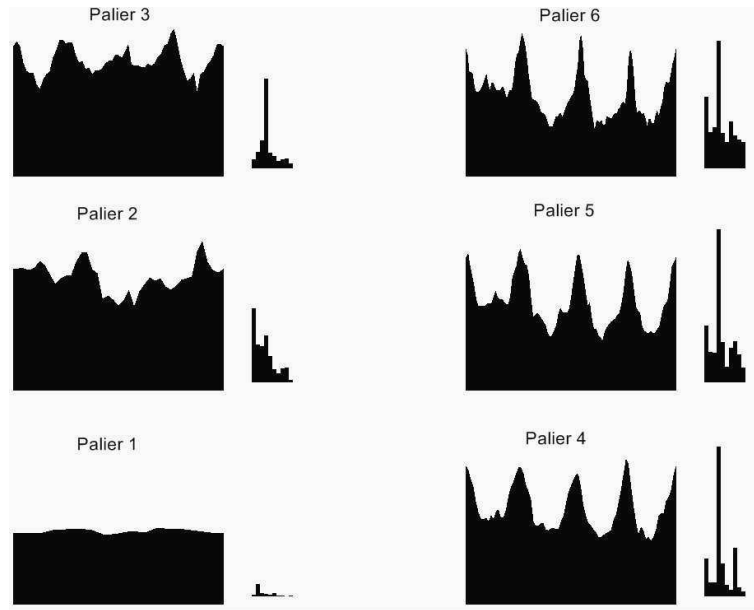


FIG. 2 – De haut en bas et de gauche à droite, signaux unidimensionnels obtenus sur les cercles de rayons croissants interpolés pour avoir tous une longueur 128. L'image sur laquelle a été faite le calcul est l'image du rond-point de gauche de la figure 8. À côté de chaque signal, dix premiers coefficients de Fourier normalisés.

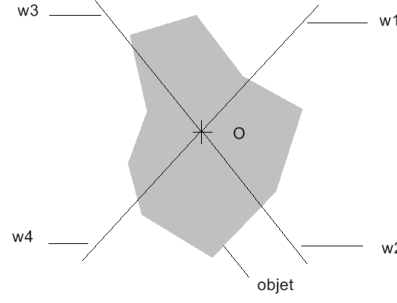


FIG. 3 – schéma explicatif du descripteur radial. w1, w2, w3 et w4 sont les vecteurs de caractéristiques extraits à partir des signaux unidimensionnels commençant en O et se dirigeant vers l'extérieur. On prend dans ce travail en général 8 signaux unidimensionnels radiaux, mais seulement 4 sont représentés ici pour plus de clarté.

du cercle le plus grand, le nombre de coefficients de Fourier à conserver. Le bon choix de ces paramètres est crucial pour l'efficacité de ce descripteur.

Descripteur radial Le descripteur précédent ne calcule des caractéristiques géométriques que sur les cercles concentriques et n'apporte donc pas d'informations sur les relations entre des pixels de distances au centre différentes.

Pour remédier à cette insuffisance, on introduit un autre descripteur de caractéristiques de la façon suivante : On considère une image contenant un objet de centre de gravité O. On trace M segments de longueur R , d'angles $\frac{k2\pi}{R}$ et ayant une de leur extrémité en O, on considère les M signaux unidimensionnels formés par la valeur des pixels situés sur ces lignes et allant du centre vers l'extérieur. On extrait un vecteur de caractéristiques pour chacun de ces signaux unidimensionnels, notons les $W_1, W_2, \dots, W_M$, où $W_i = (d_1^i, d_2^i, \dots, d_n^i)$. Il existe, tout comme le précédent descripteur, un choix infini de caractéristiques qui peuvent être choisies. Après beaucoup d'essais, le plus robuste semble être les premiers coefficients de Fourier. Contrairement au descripteur angulaire, la phase semble être un élément important qu'il convient de conserver, on ne prend donc pas le module. On normalise ces coefficients par la quantité D' définie par :

$$D' = \sqrt{\sum_{i=1}^N \sum_{j=1}^N (d_i^j)^2}.$$

Pour comparer deux objets, on calcule la somme des distances euclidiennes des vecteurs correspondant à chaque signal. Il convient cependant, pour obtenir autant d'invariance par rotation que possible, de chercher le meilleur résultat parmi les M alignements possibles entre les vecteurs. Pour obtenir davantage de robustesse, on peut également, pour obtenir la valeur d'un pixel situé sur un segment d'angle θ , faire une moyenne des pixels de rayon r et d'angle compris entre $\theta - \frac{\pi}{N}$ et $\theta + \frac{\pi}{N}$. Les paramètres à fixer sont : le nombre de segments, la taille des segments (que l'on prend en général égal à la taille du cercle le

plus grand, pour plus de cohérence) et le nombre de coefficients de Fourier à conserver.

3.1.2 Avantages et inconvénients du descripteur

Avantages Ce descripteur est invariant à la luminosité et par multiplication de la fonction de niveau de gris $I(x, y)$ par un facteur constant λ , on voit donc que ce sont véritablement des informations sur la géométrie de l'objet qui sont calculées par ce descripteur, qui est par conséquent peu perturbé par la radiométrie. Ce descripteur est également invariant par rotation, ce qui est indispensable dans une image satellitaire. Le fait de ne retenir que les premiers coefficients de Fourier fait que ce descripteur est relativement peu sensible au bruit, qui est contenu essentiellement dans les derniers coefficients.

Inconvénients Le nombre de coefficients nécessaires pour ces deux descripteurs est relativement élevé. Dans le cas où l'on prend 4 cercles et où l'on retient 5 coefficients de Fourier (c'est un minimum) pour chaque cercle, on obtient 20 coefficients pour le descripteur angulaire. Pour le descripteur radial, si l'on prend 8 segments et que l'on prend 4 coefficients pour chacun, on a 32 coefficients pour ce descripteur soit 52 coefficients pour les deux descripteurs.

La rayon du cercle le plus grand est un paramètre très sensible, en effet, si l'objet est par exemple plus petit, le cercle va intersecter d'autres choses sans rapport avec l'objet lui-même, ce qui va générer du bruit. De plus, si l'objet n'est pas de forme ronde, d'autres objets peuvent s'insérer dans la zone d'analyse du descripteur et vont altérer les caractéristiques.

Il est de plus nécessaire de connaître le centre de l'objet avec une certaine précision.

3.2 Utilisation des moments de Zernike

L'avantage d'utiliser comme caractéristiques les coefficients provenant de la décomposition de la fonction image sur une base est que l'on peut évaluer, en gardant un certain nombre de caractéristiques, l'erreur de reconstruction qui est commise. Cela permet d'évaluer assez précisément le nombre de caractéristiques à garder, ce nombre étant suffisant quand l'erreur de reconstruction est inférieure à un seuil prédéfini. Si $f_{rn}(x, y)$ est la fonction image reconstruite avec n moments, et que $f(x, y)$ est la fonction image d'origine, dont les valeurs sont comprises entre 0 et 255, on définit le taux d'erreur de reconstruction T_n par $T = \frac{\sum_{i=0}^n |f_{rn}(x, y) - f(x, y)|}{\sum_{i=0}^n f(x, y)}$. Après avoir calculé ce taux pour différentes valeurs de n sur un lot d'images de rond-points, on vérifie que pour une valeur de n supérieure à 65, ce taux est inférieure à 10%, ce qui assure une bonne reconstruction.

Sur la figure 4 est représentée la reconstruction d'une image de rond-points avec un nombre croissant de caractéristiques.

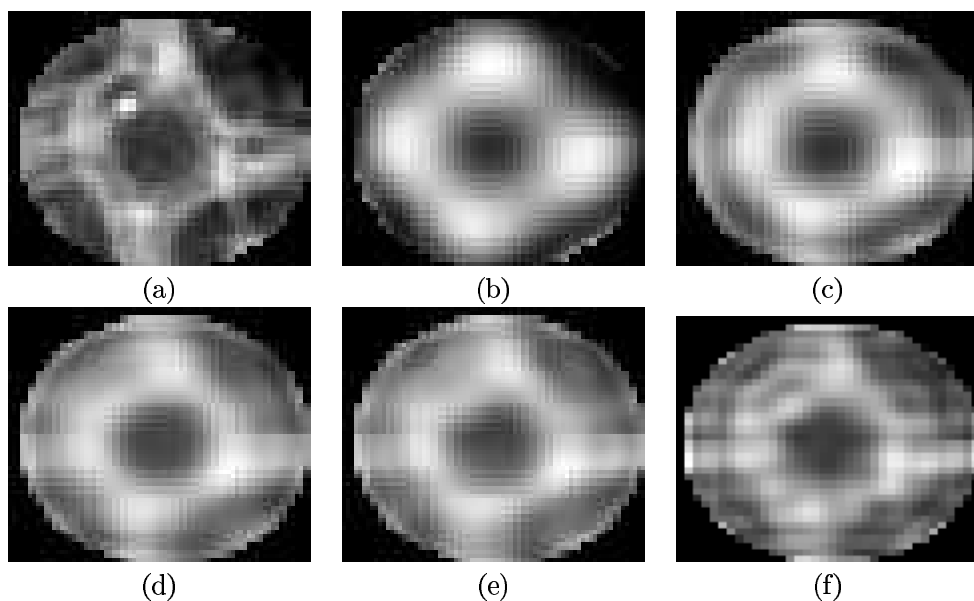


FIG. 4 – (a) : image de rond-point. Les images (b),(c),(d),(e),(f) correspondent à la reconstruction de la partie centrale de l'image avec respectivement 10, 30, 50, 70 puis 90 caractéristiques.

3.3 Evaluation des descripteurs

3.3.1 Base de données utilisée

Une base de données a été constituée au cours du stage pour permettre l'évaluation des descripteurs à partir d'images de résolution 2,5 mètres de Toulouse, Tunis, Stockholm et Bagdad. Les rond-points sont des objets pertinents pour faire cette évaluation car ils ont une variabilité modérée et une occurrence relativement élevée dans ces images. La base de données est constituée d'images de taille 128*128, 85 rondpoints ont ainsi été "découpés", mais également des stades, des échangeurs autoroutiers, des gros carrefours, des terrains de sport, des ponts, des bâtiments isolés et des ports (voir images 5,6 et 7).

La difficulté des données est très variable au sein même de la base de données. Pour ce qui est par exemple des rond-points, certains rond-points sont très nets tandis que d'autres sont beaucoup plus bruités (voir figure 8), avec, par exemple, d'autres bâtiments présents sur les côtés.

3.3.2 Application des descripteurs sur la base de données

Détection dans des images Une première tâche qui a été réalisée a été de prendre un certain nombre d'images de rond-points pour l'apprentissage et de calculer par les k plus proches voisins (on prend en général k entre 3 et 5)

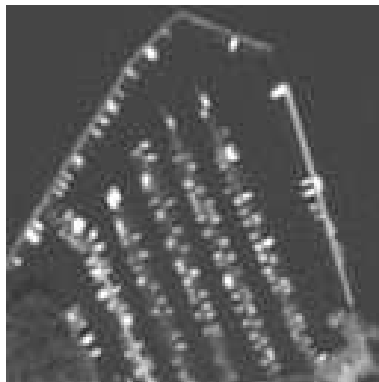


FIG. 5 – Imagerie de port.



FIG. 6 – imagerie de stade.



FIG. 7 – imagerie de rond-point.

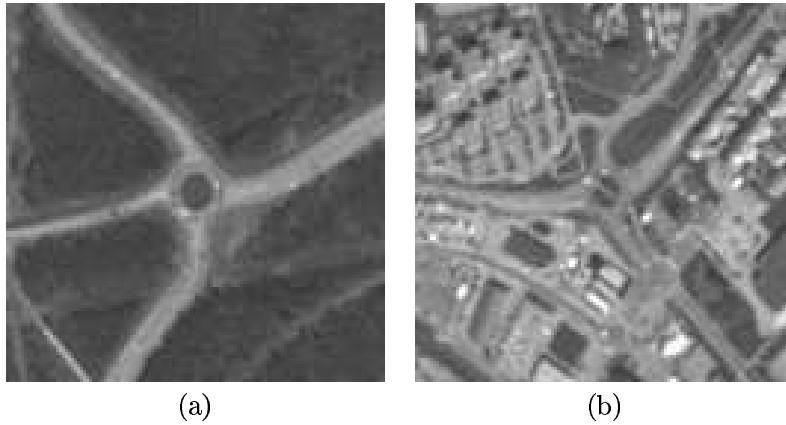


FIG. 8 – A gauche, image de rond-point très net. A droite, image de rond-point très bruité

la distance à ces images d'apprentissage en tout point d'une grande image. On génère ainsi des images dont les pixels sont d'autant plus noirs que la distance aux images d'apprentissage est petite. Ceci permet de bien visualiser les performances du descripteur et de voir si un simple seuillage suffit à détecter les rond-points.

Mais si cette tâche permet de bien visualiser l'efficacité du descripteur, la définition d'une métrique et la mise en place d'une série de tests pour obtenir une mesure précise des performances de cette façon est trop lourde. Une mesure précise des performances à partir d'images 128*128 semble plus appropriée.

Classification d'images Etant donné que l'on peut s'attendre à ce que les résultats dépendent très étroitement de la difficulté des données de test, on a fait deux lots d'images de rond-points de difficultés "moyenne" et "forte". Chacun de ces lots est lui-même séparé en un lot d'apprentissage et un lot de test. On calcule la distance aux images d'apprentissage par les k plus proches voisins et on classe l'image en comparant cette distance à un seuil. On fait varier ce seuil pour obtenir la courbe de faux rejets en fonction des fausses alarmes. Le lot de difficulté "moyenne" comprend une base d'apprentissage de 25 rond-points et une base de test de 80 images contenant 10 rond-points. Le lot de difficulté "forte" comprend une base d'apprentissage de 55 rond-points et une base de test de 95 images contenant 25 rond-points. Le lot de rond-points de difficulté "forte" contient celui de difficulté "moyenne" mais a en plus d'autres rond-points plus bruités. Pour chaque image, un vecteur de caractéristiques est calculé au pixel central (64,64), on a pris soin en effet de placer le centre du rond-point au centre de l'image.

4 Mise en oeuvre du descripteur sur la base de données

4.1 Résultats obtenus

Détection dans des images Les résultats de détection de rond-points dans les images dépendent beaucoup de la difficulté des rond-points à détecter, du bon réglage des paramètres et de la base d'apprentissage utilisée. Une base d'apprentissage trop éloignée de ce que l'on souhaite détecter, un nombre de caractéristiques insuffisant, un rond-point trop petit par rapport au rayon R choisi ou trop bruité peuvent dégrader les résultats. Mais, on remarque que si toutes les conditions sont réunies, la détection est tout à fait satisfaisante, comme le montrent les figures 9 et 10. Cependant, le fait de devoir calculer le descripteur en tout point de l'image rend le calcul assez long (10 minutes environ pour une image 512×512 et un processeur d'une puissance de 1,6 GHz). De plus, sur des images contenant des rond-points plus bruités, ces derniers sont souvent très mal détectés.

Classification d'imagettes Les figures 11 et 12 présentent un comparaisn des performances sur les deux bases de rond-points entre le nouveau descripteur et les moments de Zernike, ces résultats sont analysés dans la section suivante.

4.2 Analyse des résultats

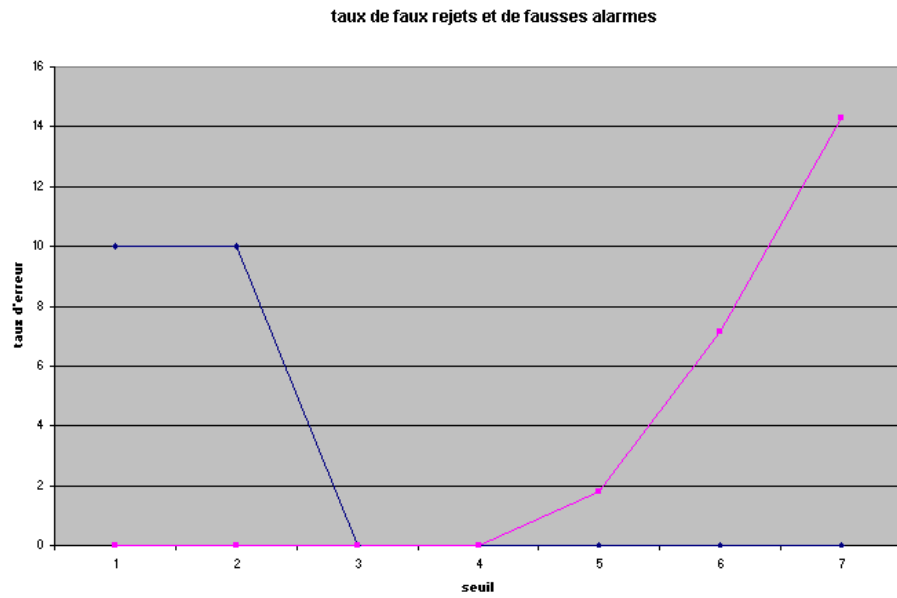
On voit, comme on pouvait s'y attendre, que les performances dépendent essentiellement de la "difficulté" des données sur lesquelles on travaille et de la proximité visuelle des données d'apprentissage aux données de test. Il est important d'avoir donc en apprentissage un échantillon représentatif de ce que l'on souhaite détecter et un bon réglage des paramètres. On remarque que le nouveau descripteur a des performances nettement supérieures aux moments de Zernike, pour un nombre de caractéristiques employé équivalent (respectivement 64 et 70). En effet, sur le lot d'images de difficulté forte, le nouveau descripteur a un taux d'égale erreur de 8% tandis que les moments de Zernike ont un taux d'égale erreur de 40%. Pour avoir un taux de faux rejet nul, il faut tolérer un taux de fausses alarmes de 25% avec le nouveau descripteur et de 70% avec les moments de Zernike. Sur le lot d'images de difficulté moyenne, le nouveau descripteur a un taux d'égale erreur nul, c'est à dire que les images de rond-points et les images ne contenant pas de rond-points sont bien séparées. Avec les moments de Zernike, on a par contre un taux d'égale erreur de 20% environ. Cependant, une sélection de caractéristiques permettrait sans doute d'améliorer les performances du descripteur contenant les moments de Zernike pour supprimer les composantes inutiles ou apportant du bruit.



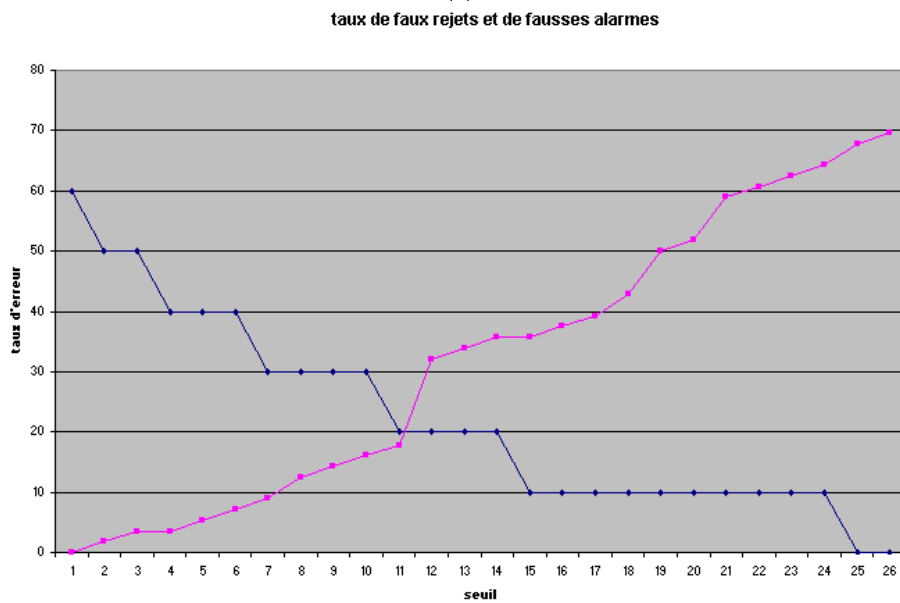
FIG. 9 – image 512*512 de stockholm contenant trois rond-points.



FIG. 10 – fonction de coût du descripteur associée à la figure 9, la couleur blanche correspond à un coût élevé et la couleur noire correspond à un coût faible. On voit donc que les trois rond-points sont ici bien localisés et qu'un simple seuillage suffit à les détecter.

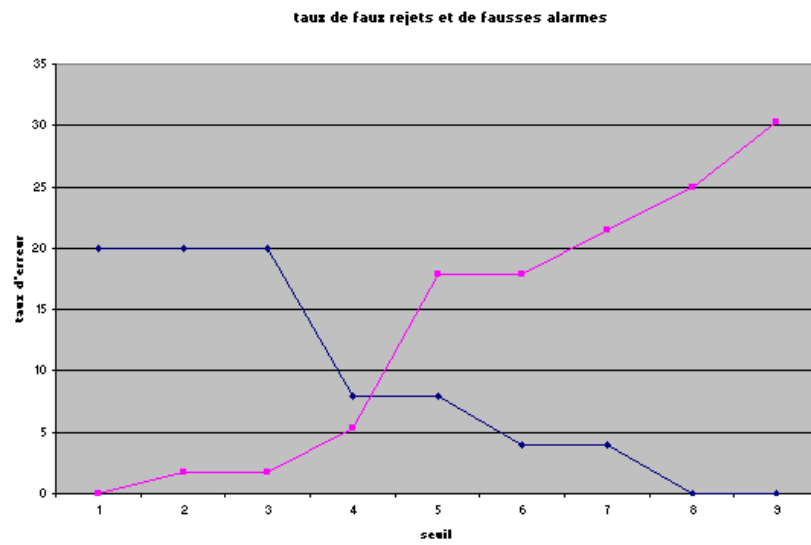


(a)

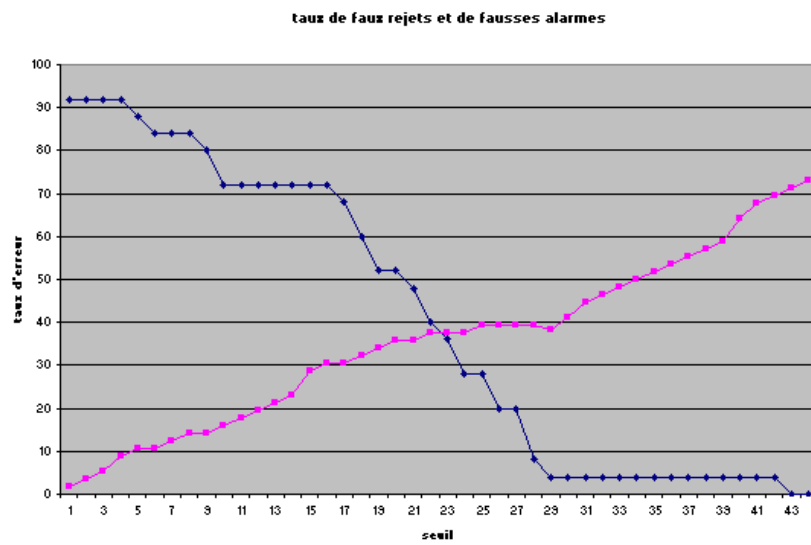


(b)

FIG. 11 – taux de faux rejets et de fausses détections dans la base de données de difficulté moyenne. Une classification avec les cinq plus proches voisins est utilisée. (a) : nouveau descripteur. (b) : moments de Zernike. L'axe des abscisse est le seuil de décision. Le trait le plus foncé correspond au taux de faux rejets et le plus foncé correspond au taux de fausses alarmes.



(a)



(b)

FIG. 12 – taux de faux rejets et de fausses détections dans la base de données de difficulté forte. Une classification avec les 7 plus proches voisins est utilisée. (a) : nouveau descripteur. (b) : moments de Zernike.

5 Conclusion, améliorations possibles et perspectives

Nous avons présenté dans ce rapport un nouveau descripteur permettant de faire de la détection de structures spécifiques après une étape d'apprentissage. Ce descripteur utilise une extraction de caractéristiques radiales et une extraction de caractéristiques angulaires. Ces caractéristiques apportent essentiellement des informations géométriques et sont invariantes par rotation. Cependant, étant donné qu'une hypothèse est faite à la base sur la forme de l'objet, des variabilités importantes sur la taille de l'objet ou la présence d'autres objets à côté de la structure à détecter peuvent dégrader la qualité de la détection. Les performances de ce descripteur sont comparées à un autre descripteur classique également invariant par rotation. L'évaluation est faite sur une base de données de ronds-points constituée au cours du stage à partir d'images satellitaires SPOT5 à 2,5 mètres de résolution. On voit que ce nouveau descripteur donne des résultats satisfaisants compte tenu du fait que l'on souhaite un algorithme de détection simple à mettre en œuvre.

Pour améliorer ce système, on pourrait à présent fournir la possibilité à l'utilisateur de donner un feedback qui permettrait d'améliorer les réponses ultérieures du système. On pourrait également faire une première passe rapide de calcul pour accélérer le processus de détection en évitant à l'algorithme de calculer les deux descripteurs en chaque point, cette première passe éviterait de faire le calcul dans des zones où des calculs plus rapides permettent de voir que l'objet ne s'y trouve pas et apporterait donc un gain de temps.

Ce travail peut également être vu comme le premier pas d'un travail plus complet de traitement de grandes bases d'images utilisant la LDA (Latent Dirichlet Affectation), qui est un modèle probabiliste génératif d'un ensemble de données discrètes. Il est souvent utilisé pour modéliser des ensembles de textes. Chaque élément est vu comme résultant d'un mélange de "sujets" sous-jacents, chaque "sujet" étant modélisé par un ensemble de probabilités. Celles-ci fournissent une représentation explicite du document. Cette méthode peut également s'appliquer à des d'images, les objets détectés par le descripteur constituant les éléments discrets qui seront gérés par le modèle LDA. Ce modèle permettrait ainsi de traiter de grandes bases d'images tout en conservant les informations statistiques importantes.

Références

- [1] Luc Brun and Mario Vento, editors. *Graph-Based Representations in Pattern Recognition, 5th IAPR International Workshop, GbRPR 2005, Poitiers, France, April 11-13, 2005, Proceedings*, volume 3434 of *Lecture Notes in Computer Science*. Springer, 2005.

- [2] A Choksuriwong, H Laurent, and B Emile. Comparison of invariant descriptors for object recognition. In *IEEE International Conference of Image Processing*, 2005.
- [3] Gyuri Dorkó and Cordelia Schmid. Object class recognition using discriminative local features. Submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.
- [4] Güray Erus and Nicolas Loménie. Automatic learning of structural models of cartographic objects. In *GbRPR*, pages 273–280, 2005.
- [5] A. Khotanzad and Y. Hua Hong. Invariant image recognition by zernike moments. *IEEE Transactions on pattern Analysis and Machine Intelligence*, 12(5) :489–497, 1990.
- [6] David G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, Corfu, Greece, September 1999.
- [7] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91–110, 2004.
- [8] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *International Conference on Computer Vision*, pages 523–531, July 2001.
- [9] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. In *Conference on Computer Vision and Pattern Recognition*, June 2003.
- [10] Ellen L. Walker and Kenji Okuma. Automatic extraction of invariant features for objects recognition. In *Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society*, pages 163–172, Boston, June 2000. GA.

A Moments de Hilbert

Soit $I(x,y)$ le niveau de gris d'un pixel de l'image I , on définit le moment d'ordre $p+q$ ($p, q > 0$) d'une image I par :

$$m_{pq} = \int_{\mathbb{R}^2} x^p y^q I(x,y) dx dy$$

Le centroïde (x_0, y_0) de la fonction I est donné par $(x_0 = \frac{m_{1,0}}{m_{0,0}}, y_0 = \frac{m_{0,1}}{m_{0,0}})$. L'image centrée $I_T(x,y) = I(x + x_0, y + y_0)$ est invariante par translation. Le moment central d'ordre $p+q$ de la fonction I s'écrit sous la forme :

$$m_{pq} = \int_{\mathbb{R}^2} x^p y^q I(x + x_0, y + y_0) dx dy$$

Les moments centraux sont invariants par translation. On introduit les moments normalisés comme suit :

$$\mu_{p,q} = \frac{v_{p,q}}{v_{0,0}^{1+(p+q)/2}}$$

Ces moments sont invariants par translation et changement d'échelle. Les moments de Hilbert se calculent à partir des moments normalisés et sont invariants par translation, rotation et changement d'échelle :

$$\begin{aligned} \phi_1 &= \mu_{2,0} + \mu_{0,2} \\ \phi_2 &= (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2 \\ \phi_3 &= (\mu_{3,0} - 3\mu_{1,2})^2 + (3\mu_{2,1} - \mu_{0,3})^2 \\ \phi_4 &= (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{2,1} - \mu_{0,3})^2 \\ \phi_5 &= (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{2,1} + \mu_{0,3})^2] + (3\mu_{2,1} - \mu_{0,3})(\mu_{2,1} + \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3})^2] \\ \phi_6 &= (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3})^2] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{2,1} + \mu_{0,3}) \\ \phi_7 &= (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{2,1} + \mu_{0,3})^2] - (\mu_{3,0} - 3\mu_{1,2})(\mu_{2,1} + \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3})^2] \end{aligned}$$

B Moments de Zernike

Les polynômes de Zernike sont des polynômes complexes qui forment un ensemble orthogonal complet sur le disque unité. Notons ces polynômes $V_{nm}(x, y)$. La forme de ces polynômes est :

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta)$$

où

n est un entier positif ou nul.

m est un entier positif ou négatif inférieur en valeur absolue à n .

$R_{nm}(\rho)$, le polynôme radial est défini de la façon suivante :

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s}$$

Ces polynômes vérifient :

$$\int \int_{x^2+y^2 \leq 1} V_{nm}(x, y)^* V_{pq}(x, y) dx dy = \frac{\pi}{n+1} \delta_{np} \delta_{mq}$$

Les moments de Zernike sont la projection de la fonction image sur ces fonctions de base orthogonales. Le moment A_{nm} est calculé par :

$$A_{nm} = \frac{n+1}{\pi} \int_x \int_y f(x, y) V_{nm}^*(\rho, \theta), x^2 + y^2 \leq 1$$

Pour calculer les moments, le centre de l'image est pris comme origine et les coordonnées des pixels sont ramenés à l'échelle du cercle unité. Les pixels sortant du cercle unité ne sont pas pris en compte dans le calcul. Notons aussi que $A_{nm}^* = A_{n,-m}$.

Reconstruction d'une image à partir des moments de Zernike Pour reconstruire une image à partir des moments A_{mn} que l'on suppose connus jusqu'à l'ordre n_{max} , on utilise, par orthogonalité de la base, la formule de reconstruction :

$$f_r(x, y) = \sum_{n=0}^{n_{max}} \sum_m A_{nm} V_{nm}(\rho, \theta)$$

Caractéristiques invariantes par rotation obtenues à partir des moments de Zernike. Plaçons nous dans un cas continu et considérons une image ayant subi une rotation d'angle α et notons la $f^r(\rho, \theta)$. La relation entre l'image originale et l'image transformée est :

$$f^r(\rho, \theta) = f(\rho, \theta - \alpha)$$

Par ailleurs, l'expression des moments de Zernike peut s'écrire en coordonnées polaire :

$$A_{nm} = \frac{n+1}{\pi} \int_{\rho=0}^1 \int_{\theta=0}^{2\pi} f(\rho, \theta) V_{nm}^* \rho d\rho d\theta$$

Le moment de Zernike de l'image ayant subi une rotation s'écrit :

$$A_{nm}^r = \frac{n+1}{\pi} \int_{\rho=0}^1 \int_{\theta=0}^{2\pi} f(\rho, \theta - \alpha) R_{nm}(\rho) \exp(-jm\theta) \rho d\rho d\theta$$

Par le changement de variable $\theta_1 = \theta - \alpha$:

$$\begin{aligned} A_{nm}^r &= \frac{n+1}{\pi} \int_{\rho=0}^1 \int_{\theta=0}^{2\pi} f(\rho, \theta_1) R_{nm}(\rho) \exp(-jm(\theta + \alpha)) \rho d\rho d\theta \\ A_{nm}^r &= \frac{n+1}{\pi} \int_{\rho=0}^1 \int_{\theta=0}^{2\pi} f(\rho, \theta_1) R_{nm}(\rho) \exp(-jm\theta_1) \rho d\rho d\theta_1 \exp(-jm\alpha) \\ A_{nm}^r &= A_{nm} \exp(-jm\alpha) \end{aligned}$$

On voit donc par cette dernière équation qu'une rotation sur l'image entraîne simplement un décalage de phase de tous les moments. Le module des moments reste par contre le même. Les $|A_{mn}|$ peuvent donc être pris comme des caractéristiques invariantes par rotation de l'image sous-jacente. De plus, $A_{n,-m} = A_{nm}^*$ entraîne que $|A_{n,-m}| = |A_{nm}|$, on peut donc simplement garder les moments A_{mn} pour lesquels $m \geq 0$.

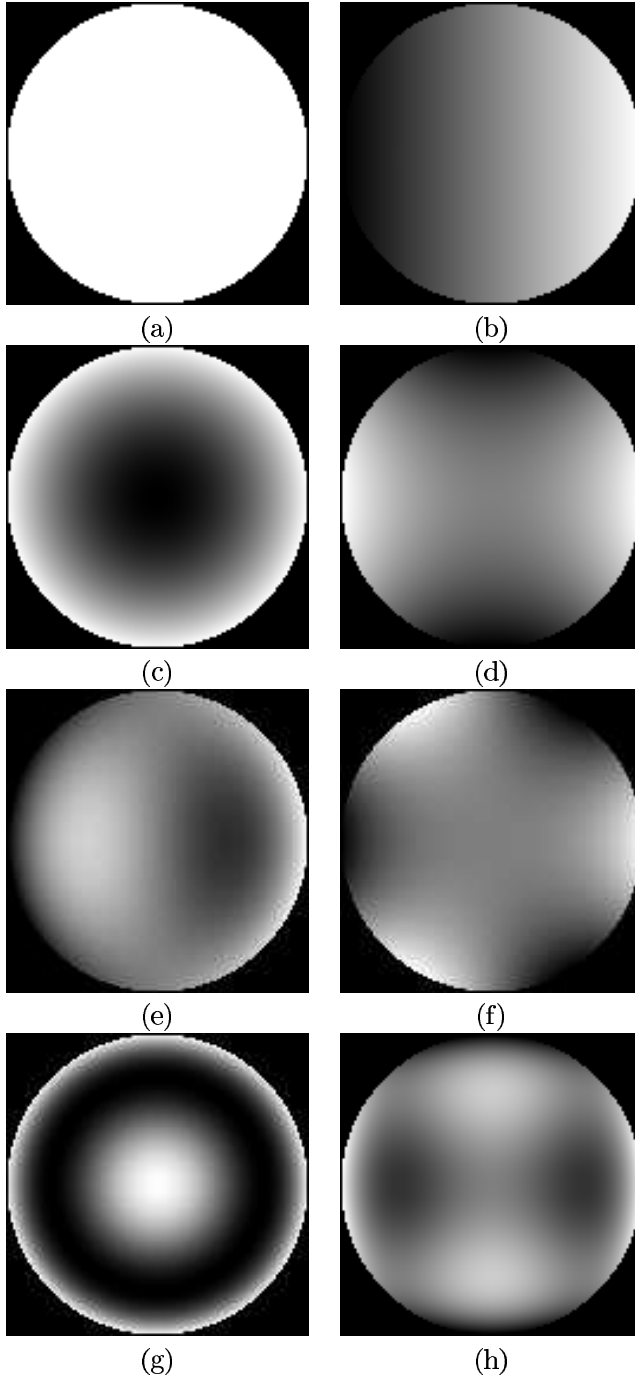


FIG. 13 – Partie réelle des polynômes de Zernike. (a) V_{00} , (b) V_{11} , (c) V_{20} , (d) V_{22} , (e) V_{31} , (f) V_{33} , (g) V_{40} , (h) V_{42}



FIG. 14 – image 512*512 de stockholm contenant deux rond-points.

C Autre résultat de détection

Les figures 14 et 15 présentent un autre exemple de détection par le nouveau descripteur sur une image de Stockholm.



FIG. 15 – fonction de coût du descripteur associée à l'image, la couleur blanche correspond à un coût élevé et la couleur noire correspond à un coût faible. Cette détection a été faite avec 10 exemples d'apprentissage en utilisant une méthode des trois plus proches voisins. On voit donc que les deux rond-points sont ici bien localisés et qu'un simple seuillage suffit à les détecter.