DETECTION OF ROUNDABOUTS IN SATELLITE IMAGES

Jean-Baptiste Bordes, Michel Roux
GET/Télécom Paris, CNRS UMR 5141
CNES-DLR-ENST Competence Center

KEY WORDS: Detection, High-resolution, Features Extraction, SPOT5, database

ABSTRACT

In this article, we present a new algorithm for the detection of objects in satellite images. Our method requires a learning of the specific structures we wish to detect but no "a priori knowledge" about them. Indeed, the user just has to provide example images of the different objects he wants to find in large images. This algorithm combines an angular local descriptor and a radial one. These descriptors are rotation invariant and provide coefficients using the Fourier analysis on some well-chosen unidimensional signals in the image. These features are computed in the learning images and an euclidan distance is applied for comparison with the same features computed in the images used for detection. The efficiency of these features is compared to the first Zernicke moments on a roundabout database extracted from high-resolution SPOT images in various urban areas.

1 INTRODUCTION

1.1 Goals of the work

During the last decades, the imaging satellite sensors (optical, synthetic aperture radar, and other sensors) have acquired huge quantities of data. Now, the storage of image archives is getting even more enormous, due to the data collected by a new generation of high-resolution satellite sensors. In order to increase the actual exploitation of satellite observations, it is very important to set up systems which are able to selectively access the information content of image archives. We are interested in this paper to increase the ability of the state-of-art systems of indexation to detect specific structures in satellite images. Most of the actual methods of object detection use a-priori knowledge about the objects. It results in time-costly adapatations when the user wants to detect other types of objects. We need in fact a method which requires no rules, but image examples of the objects the user wants to detect.

1.2 State of art

Few works have focused on the object detection in satellite images at the resolution of 2,5 meters by pixel. So, a brief state of art of very generic methods of object detection is presented here. These methods have been developed for a wide variety of applications like scene analysis, chromosome classification or target identification. Most of these methods define a set of features to represent the image while reducing its dimensionality (Walker, 2000). After the features have been computed, they are used with a classification rule to set a label to the image. Depending on the application, the invariance of the features to certain kind of transformations (rotation, scale changes, etc.) is required. Here is a brief description of the main approaches which seem relevant for our problem.

1.2.1 Object recognition with rotation invariant moments The algebric invariant features are defined as ratios or powers of moments of the image function. As they are classic examples, the Zernicke moments are presented in section 3, and the Hilbert moments are quoted here. Let $I(x, y)$ be the image function, the Hilbert moment of order $p + q, (p, q \geq 0)$ is defined by:

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

The seven invariant features of Hu are calculated from these values (Choksuriwong, 2005), they are invariant by translation, rotation, and scale changes. The value of m_{pq} corresponds to the projection of $I(x, y)$ on the polynomial basis $x^p y^q$. This basis is not orthogonal, which implies that it is time expensive to reconstruct the image from them. Moreover, it implies that the information contained in the m_{pq} is redundant. That is why the Zernicke moments appear to be more interesting and will be presented in more details later in this paper.

1.2.2 Object recognition with points of interest This method consists in extracting points of interest (for example with the Harris detector), and calculating rotation invariant features at these points (Mikolajczyk, 2003; Dorko, 2004; Lowe, 2004). At the location of the points of interest, the features are computed by deriving the local grey level values onto order three. To derive the local grey values in a rotation invariant way, the direction of derivation has to be the direction of the gradient. To detect an object, the points in the learning image are matched with the points of the test image. A strong point of this method is that it is robust to modifications of the background and to occultations. However, the information which is extracted from the object is very local because the features are only computed in a very small area near the points of interest.

1.2.3 Graph of attributes In this approach, the image is first reduced to a set of primitive shapes (segments and arcs for example). Then, from this simplified representation of the image, we build a graph whose nodes are the primitive shapes and arcs are spatial relationships between them (angle of intersection, distance, etc.). From the learning of images of a structure such as a roundabout, we can build the corresponding graph which will be used for the detection in the images. This method may be used for roundabout and bridge detection (Erus, 2005; Brun, 2005). A lot of information is extracted from the structure of the

objects, which makes this method promising. But such a system requires complicated graph matching algorithms and seems thus difficult to set up.

1.3 Presentation of the work

The method presented in this paper is based on the use of a new local descriptor which extract features on a circular region of an image. Being given a point which defines the center of the circular region of analysis, the descriptor analyses the pixels on a few circles, and on some radii of the circles. The center of the example images must coincide with the center of the objects. The descriptor is calculted in the center of each example image, and extracts a vector of features. Then, the features are calculted for each point of the image in which we want to detect the objects, and are compared with the features of the example images. This descriptor will be called the RAFA descriptor (Radial and Angular Fourier Analysis). We test the RAFA descriptor on a database of roundabouts in optical SPOT5 images at 2,5 meters of resolution. In fact, the occurrence of roundabouts in european urban areas is rather high, and their variability is not too important, which makes it is easy to evaluate this descriptor on that kind of object. We compare the efficiency of this descriptor with the Zernicke moments which are often used as rotation-invariant features for object detection. We will note that the performances of the RAFA descriptor are better than the performances of the Zernicke moments for this database.

2 PRESENTATION OF THE RAFA DESCRIPTOR

Many methods of object detection based on a learning process use local descriptors which extract vectors of features at given points of the image. These vectors of features are then compared with reference vectors in order to detect the objects. The method presented here relies on a new local descriptor which is rotation invariant and extracts geometrical characteristics of the objects. The descriptor we describe here contains in fact one descriptor which extracts features about the repartition of the pixels on concentric circles, and one descriptor which extracts features about the repartition of pixels on the radii of these circles.

2.1 Angular descriptor

For a given point O, let us consider N circles of radius kR/N of center O and the N unidimensionnal signals given by the value of the grey level pixels on these circles. The Fourier transform of each of these signals is calculated and only the modules of the p first coefficient are kept so that the features may be rotation-invariant. Let us note them: $V_1, V_2, V_3, ..., V_N$ for each circle, where $V_i = (c_1^i, c_2^i, ..., c_p^i)$. The features c_i^j are then normalized by the quantity defined by $\sqrt{\sum_{i=1}^{p}\sum_{j=1}^{N}(c_i^j)^2}$. Thus, if the image function $I(x, y)$ is multiplied by a factor λ, the features remain unchanged. To compare two objects, all the features extracted from each circle are put in a single vector and the euclidian distance is used to measure thier similarity. An other interesting way to compute a distance between two set of

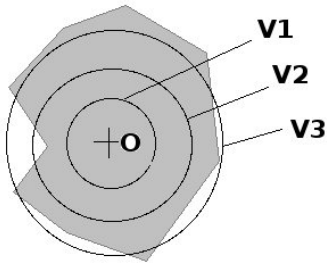


Figure 1: Explicative diagram of the angular descriptor. V1, V2, V3 are the vectors of features extracted on the unidimensional signals defined by the circles.

angular features is to compute the edition distance between the two vectors. It provides a kind of scale invariance. The length of the circumferences of the different circles being different, the length of the different signals are different, and, if we use this method of distance computation, it may be appropriate to interpolate all the different signals to the same length before computing the features. With our database, the performance of the results did not appear to be significantly better with the edition distance than with the euclidian distance. To improve the robustness of this descriptor, we can also, instead of keeping directly the grey value of a pixel of angle θ, get a ponderated mean of the value of the pixels which have an angle θ and a radius between $[r - R/2N, r + R/2N]$. The parameters that have to be fixed by the operator are: the number of circles, the radius of the biggest circle, the number of Fourier coefficients which are kept. The good choice of these parameters is crucial for the efficiency of this descriptor.

2.2 Radial descriptor

The previous descriptor extracts geometrical features only on concentric circles, and thus does not give information about the relationship between pixels which are at different distances to the center of the circles. In order to make up for this shortcoming, we introduce an other descriptor. For an image with a given point O, we draw M segments wich have a length R (the length of the biggest circle of the previous section), of angles $k2\pi/M$ with one of their extremity in O. We then calculate the Fourier transform of the unidimensional signals obtained by considering the grey value of the circles from the point O to the other extremity. Both the real and the complex part of the coefficients are kept. M vectors of features are thus computed: $V_1, V_2, ..., V_M$, one for each radius, where $V_i = (c_1^i, c_2^i, .., c_n^i)$. The coefficients c_i^j are then normalized by the quantity defined by: $\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{M}(c_i^j)^2}$. Thus, if the image function $I(x, y)$ is multiplied by a factor λ, the features remain unchanged. To compare two objects, the best alignment is computed between the vectors $V_1^1, V_2^1, V_3^1, ..., V_M^1$ and $V_1^2, V_2^2, V_3^2, ..., V_M^2$ so that the features extracted by this radial descriptor may be as rotation-invariant as possible. Thus, to calculate the

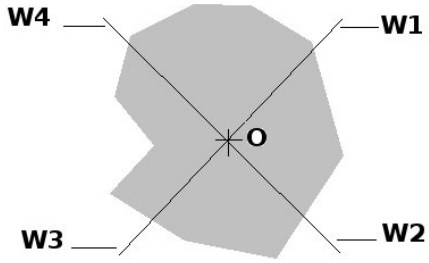


Figure 2: Explicative diagram of the radial descriptor. W1, W2, W3, W4 are the vectors of features extracted on the unidimensional signals defined by the four radius.

distance, the following expression is used:

$$\min_{\text{translations } \pi \text{ on } [1,M]} (\sum_{i=1}^{n} \sum_{j=1}^{M} (c^j_{1,\pi(i)} - c^j_{2,i})^2)$$

To improve the robustness of this descriptor, we can also, instead of keeping directly the grey value of a pixel of radius r and of angle θ, get a ponderated mean of the value of the pixels which have a radius r and an angle between $[\theta - \pi/N, \theta + \pi/N]$. The parameters that have to be fixed are: the number of radius and the number of Fourier coefficients which are kept.

2.3 Discussion about the descriptors

The descriptor is rotation-invariant, and invariant if the gray-level fonction $I(x, y)$ is multiplied by a constant factor λ. The descriptor is rather noise-robust because we just keep the first Fourier coefficients. However, the number of coefficients we keep is high. For example, if we use for the angular descriptor four circles with five coefficients for each circle, and for the radial descriptor eight radius with four coefficents for each segment, we then have already 52 coefficients. Moreover, the radius of the biggest circle is a very critical parameter, because the zone of analysis has to fit rather precisely the shape of the object. Otherwise, the circles and the radius will intersect other objects, this will result in additional noise. The center of the object must also be determined rather precisely, because both the radial and angular descriptors are not translation invariant. To detect an object in the image, we then have to calculate the descriptor in each point of the image. And for the learning stage, we have to make the center of the object coincide with the the center of the example image.

3 ZERNICKE MOMENTS

3.1 Definition of the Zernicke moments

The Zernicke moments are calculated by projecting the image fonction $I(x, y)$ on a set of complex polynomials which form a complete orthogonal set over the interior of the unit circle (Khotanzad, 1990; Choksuriwong, 2005). Let the set of these polynomials be denoted by $\{V_{mn}(x, y)\}$. The form of these polynomials is:

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

where n is a positive integer or zero, and m is a positive or negative integer subject to constraint $|m| \leq n$. $R_{nm}(\rho)$, the radial polynomial, is defined by:

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

The Zernicke moments are the projection of the image function onto these orthogonal basis functions:

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

For a digital image, the integrals are replaced by summations to get:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V^*_{nm}(\rho, \theta), x^2 + y^2 \leq R^2$$

Now, if we consider a rotation of the image $I(r, \theta)$ through angle α: $I^\alpha(r, \theta)$, with the equation:

$$I^\alpha(\alpha, \theta) = I(r, \theta - \alpha)$$

We find that:

$$A^\alpha_{nm} = A_{nm}. \exp(-jm\alpha)$$

And if we take the modules, we get:

$$|A^\alpha_{nm}| = |A_{nm}|$$

So, the modules of the moments are rotation-invariant, and we can keep them as features of the image. We can also notice that $A^*_{nm} = A_{n,-m}$ and thus $|A_{nm}| = |A_{n,-m}|$, and so we just keep the $|A_{nm}|$ coefficients, for $0 \leq m < n$.

3.2 Reconstruction of an image

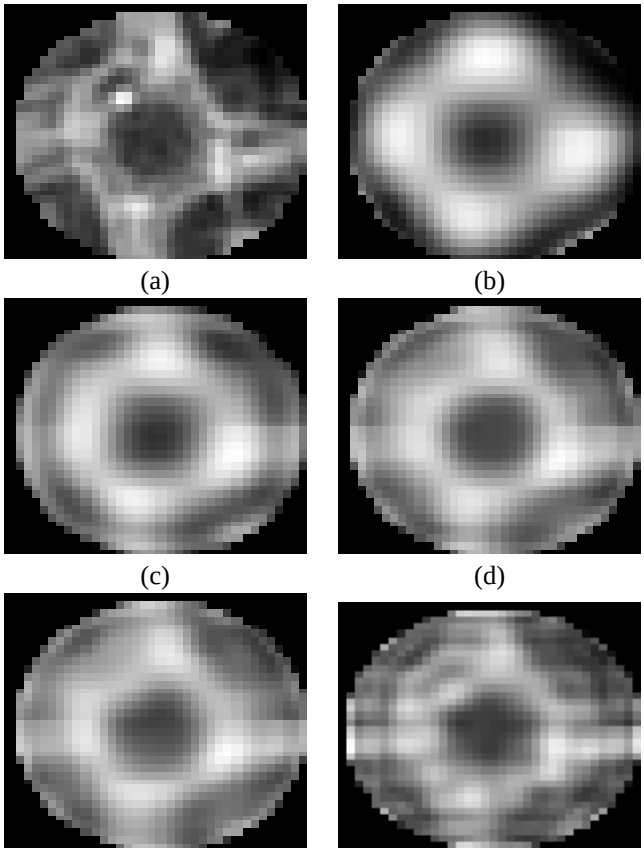
Bearing in mind the fact that the Zernicke moments are the coefficient of the decomposition of the image on a polynomial basis, we can evaluate the quality of the reconstruction of the image when we keep a certain number of coefficients. Let us note $f(x, y)$ the image function (whose value is included in [0,255]), and $f_N(x, y)$ the image function reconstructed with N moments:

$$f_N(x, y) = \sum_{n=0}^{N} \sum_{m<n} (A_{nm} V_{nm}(\rho, \theta))$$

We define a rate of error reconstruction T_N by

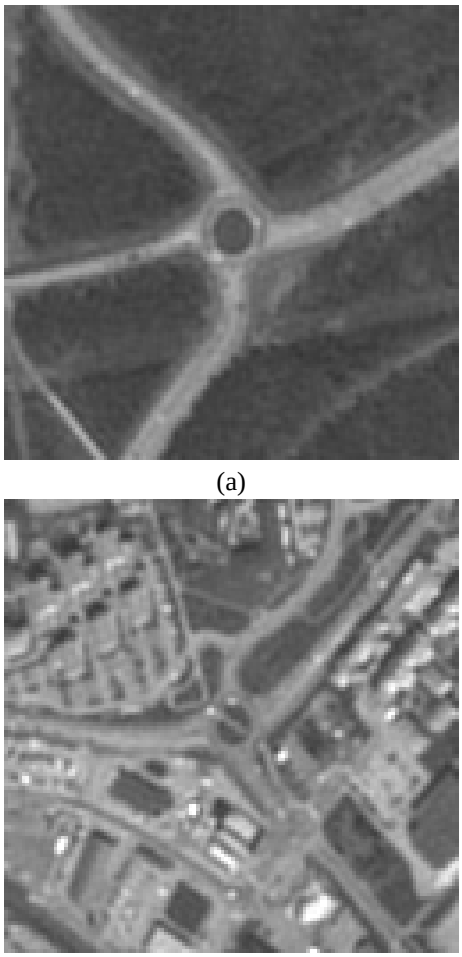
$$T_N = \frac{\sum_{x,y} (|f_N(x, y) - f(x, y)|)}{\sum_{x,y} f(x, y)}$$

On the roundabout database, we notice that, if we keep more than 65 coefficients, $T_N < 0.1$, we thus have a good reconstruction. The figure 2 shows the reconstruction of a roundabout with an increasing number of moments.



(a) (b) (c) (d) (e) (f)

Figure 3: (a): image of a roundabout ©CNES. Then, reconstruction with (b): 10, (c): 30, (d): 50, (e): 70 and (f): 90 coefficients.



(a) (b)

Figure 4: (a): clean roundabout ©CNES. (b): noisy roundabout ©CNES.

4 EXPERIMENTAL RESULTS

4.1 Presentation of the database

We have set up a database of 230 roundabouts extracted from SPOT5 images at 2,5 meter of resolution. Some roundabouts are quite isolated, but other roundabouts are not very clear and appear wih a lot of noise and other objects around them (cf Fig.4).

4.2 Qualitative results: detection of the roundabouts in SPOT5 images

4.2.1 Procedure Being given N example images of roundabouts and an image where we want to detect roundabouts, we extract one vector of features for each example image at the center of the image. Then we extract the features with our descriptor at every pixel of the image and we calculate a distance with the reference vectors by computing a ponderated mean with the k nearest reference vectors. To visualise the efficiency of the descriptor, we generate images where the value of the grey level is all the more high than the distance is low (cf Fig.5).

4.2.2 Results The results depend a lot on the noise and the environment of the roundabout in the image which make

it more or less difficult to detect. But it depends also on the good choice of the parameters, and on the learning database. If all these conditions are satisfied, we note that the results are quite good. However, as the algorithm computes the features in every pixel, the calculus can take a significant time, depending on the number of circles, the number of radius, and the size of the image. The algorithm takes approximatly ten minutes for a 512×512 image with a processor of 1,6 GHz for 5 circles and 8 radius.

4.3 Quantitative results: classification of example images

Even if the detection results of the previous sections in large images are useful to evaluate qualitatively the performances of the RAFA descriptor, it is rather difficult to obtain quantitative results in this way because it is necessary to take into account the localisation of the detection and thus to set a precise metric. An easier way to get quantitative results is to perform a classification directly on the small example images of roundabouts.

4.3.1 Procedure We divided the database of roundabouts in two sets: a learning set (85% of the roundabouts of the database) and a test set (15% of the roundabouts of the database). In the test set, we added 150 images of urban areas in which there are no roundabouts. We compute a vector of features for each image and we compare the test vectors to the learning vectors by computing a ponderated mean of the k nearest learning vectors. We make the treshold vary and we calculate the rate of false alarms and of false rejections.

4.3.2 Results Results are shown on figure 6. We can see that the performances of the RAFA descriptor are better than those of the Zernicke moments with a similar number of features (respectively 64 and 70). Indeed, the equal error rate of the RAFA descriptor is 17%, whereas the equal error rate of the Zernicke moments is 40%. However, a features selection would probably increase the performances while reducing the number of features.

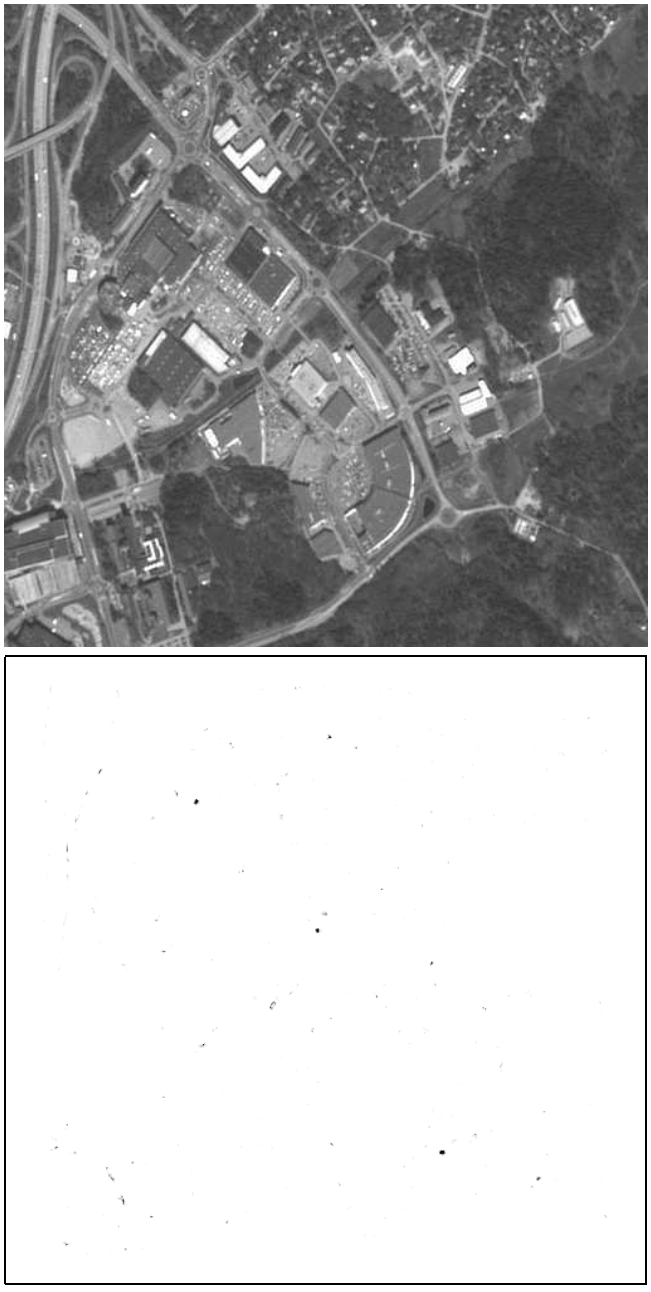
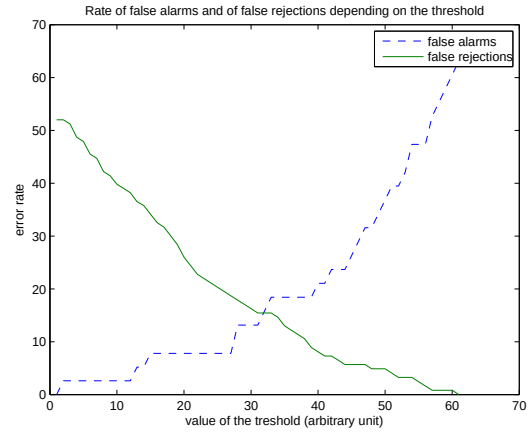


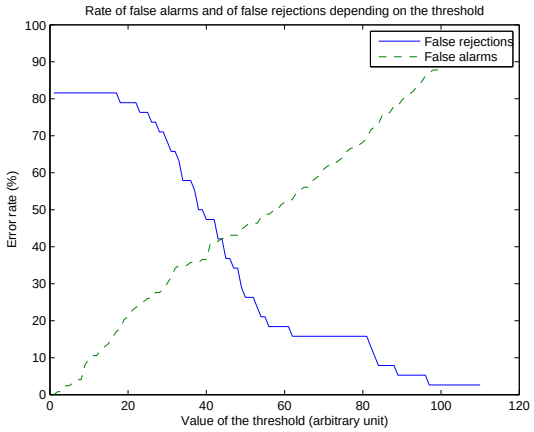
Figure 5: result of detection in a SPOT5 image ©CNES. The three darkest spots correspond to the three clearest roundabouts. So they can be detected with a well chosen threshold. However, they are other crossroads in this image which can be interpreted as roundabouts but they are not detected by the RAFA descriptor.

5 CONCLUSION

We presented in this paper a new local descriptor in order to detect objects after a specific learning from example images of the structure. This descriptor computes characteristics about radial and angular distribution of the pixels, and extracts geometric information about the object. This descriptor is invariant by rotation and by multiplication of the image function by a constant factor. However, due to the fact that this relies on a strong hypothetis about the shape of the object, the presence of other objects in the zone of analysis or important variations of the size of the object deteriorate the results. The performance of this descriptor is compared to an other local descriptor which is also rotation invariant. The evaluation is done with a database of roundabouts extracted from SPOT5 images of urban areas at 2,5 meters of resolution. On this database, we note that the results of our descriptor is better than those of Zernicke



(a)



(b)

Figure 6: (a) rate of false alarms and of false rejections with the RAFA descriptor (67 coefficients). (b) rate of false alarms and of false rejections with the 70 firsts Zernicke moments.

moments, and that the detection results are quite satisfactory, bearing in mind the simplicity of the approach. The method presented here si the begining of a work whose aim is to interpret satellite images. The objects detected in the image are seen as a collection of discrete items. Models like LDA (Latent Dirichlet Allocation) can deal with such collections of discrete items while conserving essential statistical information and give an explicit representation of an image (Blei, 2003).

We thank Mr Alain Giros from the CNES for providing us the images we used in this paper to evaluate our algorithms.

REFERENCES

Blei, D, A Ng, and M Jordan, 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022.

Brun, L and M Vento (Eds.), 2005. *Graph-Based Representations in Pattern Recognition, 5th IAPR InternationalWorkshop, GbRPR 2005, Poitiers, France, April 11-13, 2005, Proceedings*, Volume 3434 of *Lecture Notes in Computer Science*. Springer.

Choksuriwong, A, H Laurent, and B Emile, 2005. Comparison of invariant descriptors for object recognition. In *IEEE International Conference of Image Processing*.

Dorkó, G and C Schmid, 2004. Object class recognition using discriminative local features. Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence.

Erus, G and N Loménie, 2005. Automatic learning of structural models of cartographic objects. In *GbRPR*, pp. 273–280.

Khotanzad, A and Y. H Hong, 1990. Invariant image recognition by zernike moments. *IEEE Transactions on pattern Analysis and Machine Intelligence* 12(5), 489–497.

Lowe, D, 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110.

Mikolajczyk, K and C Schmid, 2003, June). A performance evaluation of local descriptors. In *Conference on Computer Vision and Pattern Recognition*.

Walker, E and K Okuma, 2000, June). Automatic extraction of invariant features for objects recognition. In *Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society*, Boston, pp. 163–172. GA.