STRUCTURAL HIGH-RESOLUTION SATELLITE IMAGE INDEXING

Gui-Song Xia[1], Wen Yang[2], Julie Delon[1], Yann Gousseau[1], Hong Sun[2], Henri Maître[1]

[1]CNRS-LTCI, TELECOM ParisTech, 46 rue Barrault, 75013 Paris, France
e-mail: {*xia, delon, gousseau, henri.maitre*}*@enst.fr*
[2]Signal Processing Lab., School of Electronic Information, Wuhan University,
LuoJia Hill, Wuhan 430072, China
e-mail: {*yw, sh*}*@eis.whu.edu.cn*

KEY WORDS: retrieval and classification, high-resolution satellite image, structure, texture

ABSTRACT:

Satellite images with high spatial resolution raise many challenging issues in image understanding and pattern recognition. First, they allow measurement of small objects maybe up to 0.5 m, and both texture and geometrical structures emerge simultaneously. Second, objects in the same type of scenes might appear at different scales and orientations. Consequently, image indexing methods should combine the structure and texture information of images and comply with some invariant properties. This paper contributes to the indexing of high-resolution satellite images. We suggest a satellite image indexing method relying on topographic maps and a shape-based image indexing scheme. The proposed approach contains both the textural and structural information of satellite images and is also robust to changes in scale, orientation and contrast. Experimental analysis on a real satellite image database confirms the efficiency of the approach.

1 INTRODUCTION

Remote sensed satellite imaging has been widely applied to agriculture, geology, forestry, regional planning, and many other applications for analyzing and managing natural resources and human activities. In the past few years, with the development of imaging techniques, satellites with very high spatial resolution imaging systems have been launched, e.g. IKONOS, QuickBird, World-View-1, GeoEye-1, which enable satellite imagery to provide more accurate earth observation and measure small objects on the surface up to 0.5 m.

However, satellite images of high spatial resolution present many challenging problems in image understanding, information mining, and pattern recognition. First, with the enhancement of spatial resolution, more details on the earth surface emerge in satellite imagery. Unlike the case of low-resolution satellite images, where texture and intensity cues have been proved to be efficient for recognition (Li and Castelli, 1997, Richards and Jia, 2005, Ruiz et al., 2004), structures become more important for analyzing high-resolution satellite images. It is of great interest to investigate new image indexes, which can describe both the structure and texture information for high-resolution satellite image recognition. Second, in satellite images of high spatial resolution, objects contained in the same type of scenes might appear at different scales and orientations. For instance, the buildings in urban areas or the bridges on the river always show at various sizes and orientations. Moreover, if satellite images were taken under different weather conditions, there might be lighting changes between images of the same type. For these reasons, image indexing methods should comply with some invariant properties, such as scale invariance, orientation invariance and contrast invariance.

In order to extract structural features from optical satellite images of high-resolution, (Ünsalan and Boyer, 2004) proposed to use statistics of straight lines and their spatial arrangement over relatively small neighborhoods. (Bhattacharya et al., 2007, Bhattacharya et al., 2008) suggested to use geometrical information, e.g. *edge* and *Junction density*, from the extracted road network and segmented urban regions for structural satellite image indexing. As inspired by the works in computer vision, (Newsam and Yang, 2007) investigated interest point descriptors, such as *Scale-invariant feature transform* SIFT, for characterizing remote sensed images. Other structural features are computed from the pre-segmentation of images. One main disadvantage of this kind of approaches is that they rely on some pre-analysis of images, such as edge detection and segmentation, which are in themselves challenging problems. In addition, when these indexing schemes focus on structure information, they ignore the use of texture cues.

This paper contributes to structural indexing of high-resolution optical satellite images. The proposed indexing scheme is based on a complete morphological image representation, called *Topographical Map* (Caselles et al., 1999), which is made of all the connected components of the level lines of images. More precisely, the indexing of satellite images follows the shape-based indexing scheme, proposed by (Xia et al., 2009). First, satellite images are decomposed into tree of shapes, by using a morphological transformation, named *Fast Level Set Transformation* (FLST) (Monasse and Guichard, 2000). Then, image features are developed from those shape ensembles and their relationships. The derived shape-based features describe the structure distributions of images. They also encode the texture information of images, if taking shapes as *textons* (Zhu et al., 2005). Furthermore, the developed satellite images features are invariant to geometric transformations involving scaling and rotating and are robust to illumination changes.

The remainder of the paper is structured as follows. In Section 2, we give a brief review on our shape-based image indexing scheme. We then detail the proposed approach for high-resolution satellite images in Section 3. Section 4 provides the experimental analysis and Section 5 concludes the paper.

2 SHAPE-BASED IMAGE INDEXING FRAMEWORK

This section sketches the basics of our work, *i.e.* the topographic map and the shape-based image indexing framework.

2.1 Topographic map

For a gray-scale image u, the upper and lower level sets are defined respectively as $\chi_\lambda(u) = \{x \in \Omega;\ u(x) \geq \lambda\}$ and $\chi^\lambda(u) = \{x \in \Omega;\ u(x) \leq \lambda\}$, for $\lambda \in R$. The topographic map (Caselles et al., 1999) of the image u is made of the connected components of the topological boundaries of the upper level sets of the image (they could be equivalently defined from the lower level sets). Observe that the connected components of upper level sets (respectively of the lower level sets) are naturally embedded in a tree structure. Monasse and Guichard combined these two redundant tree structures, by drawing on the notion of *shape*, and developed an efficient way to compute a hierarchical representation of images (Monasse and Guichard, 2000), named FLST[1] as mentioned before. A ***shape*** is defined as the interior of a level line (the boundary of a level set). Figure 1 shows an example of the topographic map representation of a synthetic image.

It is shown that topographic map, the tree of shapes of an image has many impressive properties. First, it inherits a hierarchical structure from the nesting properties of level sets and it's a scale space without any geometrical degradation. Secondly, it's a complete image representation and can encode both the geometric and radiometric information simultaneously. And it's also invariant to any contrast changes.

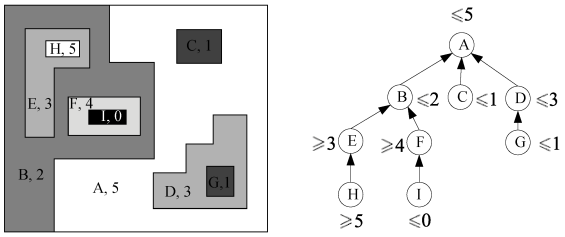


Figure 1: Representation of an image by its topographic map. Left: an original digital image; Right: representation of the image by its tree of shapes, where $(A, B, \ldots, I)$ denote the corresponding shapes.

2.2 Shape-based image indexing framework

By relying on the topographic map representation, Xia *et al.* proposed a shape-based invariant image indexing scheme in (Xia et al., 2009). A flowchart of the scheme is provided in Figure 2. The idea is to decompose images into shapes (by using FLST) and then develop image features from the shape ensembles and their relationships. It has been shown that the framework is very efficient for achieving geometric invariant texture features and obtain state-of-the-art performance on invariant texture recognition task.

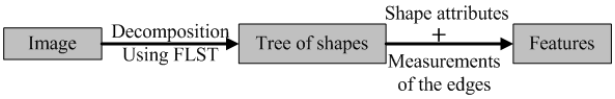


Figure 2: Shape-based image indexing framework.

As the topographic map provides a complete representation of images, the modeling of texture u is converted to the modeling of the tree of shapes (S, T), as $p(u) = p(T, S)$. The invariant texture features first rely on classical shape moments, then make use of the hierarchical structure of the topographic map.

[1]The codes of FLST are included in the free software MegaWave, and can be downloaded at http://megawave.cmla.ens-cachan.fr/.

The $(p+q)$th order central moments of shape s is defined as

$$\mu_{pq} = \int\int_s (x - \overline{x})^p (y - \overline{y})^q \, dxdy, \tag{1}$$

where $(\overline{x}, \overline{y})$ are the center of mass of s. According to the framework of (Xia et al., 2009), three different invariant features are developed from the invariant moments of shape ensembles S:

- EH: histogram of elongation $\epsilon = \frac{\lambda_2}{\lambda_1}$;
- CH: histogram of compactness $\kappa = \frac{1}{4\pi\sqrt{\lambda_1 \lambda_2}}$;
- CtH: histogram of contrast: $\gamma(x) = \frac{u(x) - \mathrm{mean}_{s(x)}(u)}{\sqrt{\mathrm{var}_{s(x)}(u)}}$.

where $s(x)$ is the smallest shape containing pixel x, λ_1 and λ_2 are the two eigenvalues of the normalized inertia matrix of shape s, with $\lambda_1 \geq \lambda_2$.

To develop features from the tree structure T, the parent-children relationships are used by defining an ancestor family of s, $\mathcal{N}_s^M$, as shapes containing the m-th ($m \leq M$) cascaded parents of s. The feature is:

- SH: histogram of scale ratio $\alpha(s) = \frac{\mu_{00}(s)}{\langle \mu_{00}(s') \rangle_{s' \in \mathcal{N}^M}}$;

where $\langle \cdot \rangle_{s' \in \mathcal{N}^M}$ is the mean operator on $\mathcal{N}^M$.

Observe that EH is invariant to similarity (translation, scaling and rotation) changes and CH, SH and CtH are invariant to affine transformations. Furthermore, all the four features are invariant to increasing contrast changes.

3 STRUCTURAL SATELLITE IMAGE INDEXING

This part devotes to the structural indexing of high-resolution satellite images, under the shape-based image indexing framework.

3.1 Structures of satellite imagery

Panchromatic images and multispectral images are the main types of optical satellite images acquired by optical remote sensing sensors. For satellite imagery in panchromatic format, all the structure information is, of course, contained in the gray scale image. However, for a multispectral satellite image $U = \{u_1, u_2, \ldots, u_L\}$ of L bands, an L-dimensional vector is stored for each pixel. In this case, we suppose that the main structure information of U is included by its p-energy channel $\mathcal{L}$, defined as

$$\mathcal{L} = \left(\sum_{u_i \in U} (u_i)^p \right)^{\frac{1}{p}}, \ p \geq 1. \tag{2}$$

Actually, Caselles *et al.* have proved that the main geometric information of natural color images are contained in their luminance channel (Caselles et al., 2002) (where $p = 1$). In the context of this work, as we shall see, we will only deal with natural color satellite images, so the analysis of structure information is based on the luminance channel. The same scheme could be applied to multispectral images using the p-energy channel $\mathcal{L}$.

3.2 High-resolution satellite image indexing

High-resolution satellite images allow to accurately represent small objects on the earth surface, such as cars, airplanes and buildings. An important and discriminative measurement for the objects is made of the shapes of their contour. For instance, bridges usually have elongated shapes, and the outlines of cars are usually compact and not too elongated.

First, following the shape-based image indexing scheme presented in Section 2, we also use the *elongation histogram* (EH), *compactness histogram* (CH), *scale ratio histogram* (SH), and *contrast histogram* (CtH) mentioned in Section 2 as indexes of satellite images.

Secondly, for the purpose of this paper, we propose to add several features, all invariant to similarity, a necessary invariance for satellite images. For instance, the orientation distribution of an image is also an available structure feature.
Orientation distribution: For a shape s, its orientation θ is calculated as

$$\theta = \frac{1}{2}\arctan\frac{2\mu_{11}}{\mu_{20}-\mu_{02}} \tag{3}$$

The resulted structure feature is the histogram of θ, named *Orientation Histogram* (OH), on all the shapes contained by the image.

We can also develop more structural features from the tree structure T, by using the M-order ancestor family $\mathcal{N}^M$:
Nested contrast: For a shape s, we define its nested contrast $\delta(s)$ as,

$$\delta(s) = \langle |u(s) - u(s')| \rangle_{s' \in \mathcal{N}^M}. \tag{4}$$

Maximum axis ratio: It's defined as ratio between the maximum eigenvalue of shape s and the average maximum eigenvalue among its ancestor family, as

$$\rho(s) = \frac{\lambda_1(s)}{\langle \lambda_1(s') \rangle_{s' \in \mathcal{N}^M}}. \tag{5}$$

where $\langle\cdot\rangle_{\mathcal{N}^M}$ is the mean operator on $\mathcal{N}^M$. We can similarly define the axis ratio by replacing λ_1 with λ_2. However, observe that $\mu_{00}(s) \propto \lambda_1\lambda_2$, so the similar definition involving to λ_2 will be redundant with the scale ratio $\alpha(s)$.
The corresponding features are the histograms of δ and ρ of all the shapes on the tree, called *nested contrast histogram* (NCH) and *maximum axis ratio histogram* (MAH), respectively. They are invariant to similar transformations.

input : Satellite image u
output: A set of indexes of u

Compute the luminance channel $\mathcal{L}(u)$ of the image;
Decompose $\mathcal{L}(u)$ into a tree of shapes $\{S, T\}$;
for *each* s_i *in* S **do**
    Compute the shape attributes $\epsilon_i, \kappa_i, \gamma_i, \theta_i$;
end
for *each* s_i *and* $\mathcal{N}^M_{s_i}$ *on* T **do**
    Compute α_i, δ_i, *and* ρ_i;
end
for *each attribute* ξ^i_k *in* $\{\epsilon_i, \kappa_i, \gamma_i, \theta_i, \alpha_i\, \delta_i, \rho_i\}$ **do**
    compute histogram $H(\xi_k) = \frac{\#\{\xi^i_k = \xi_k, 1 \le i \le N\}}{N}$.
end

Algorithm 1: Structural satellite image indexing.

A given high-resolution satellite image u is then characterize by 7 1-D histograms as detailed above (EH, CH, SH, CtH, OH, NCH, MAH). The pipeline of the indexing steps is given by Algorithm 1.

Remark that in order to compute the dissimilarity between two satellite images, we use the Kullback-Leibler divergence to compute distances between single descriptors and then add them together. Especially, for comparing two OH's, one is circularly shifted to compute the Kullback-Leibler divergence with the other and the minimum divergence among them is taken as the dissimilarity.

4 EXPERIMENTAL ANALYSIS AND DISCUSSION

In this part, we illustrate the proposed analysis scheme on high-resolution satellite image recognition tasks. In order to evaluate the efficiency of the proposed approach, we also compare it with other features for satellite image indexing.

As the structural image indexing approach proposed by (Bhattacharya et al., 2007) relies on edge or junction density in small patches of images, it is not invariant to scale changes and of course not comparable to our approach. The interest point based features, such as SIFT descriptor, might have some potential for the structural analysis of satellite images. For example, it has been used by (Newsam and Yang, 2007) for satellite image retrieval. However, one main disadvantage of interest points based approaches is that the detection of interest points depends on the contrast of the images. For some images where no contrasted structure presents, for instance, the meadow and forest displayed in Figure 3, there will be few interest point detected, which makes the indexing of those images difficult.

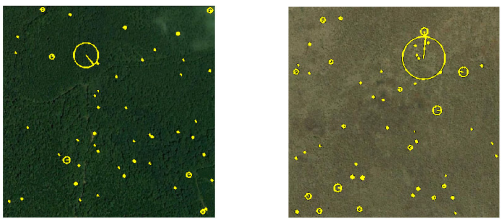


Figure 3: Two examples where few SIFT descriptors are detected. ***Left***: a forest image; ***Right***: a meadow sample.

In this work, we simply compare the proposed structural indexes with texture features based on Gabor filters with 6 scales and 8 orientations. In order to achieve rotation invariant, we average on all the 8 orientations.

4.1 High-resolution Database

To test the proposed satellite image indexing method, we collect a set of satellite images exported from Google Earth[2], which provides high-resolution satellite images up to 0.5 m. Some samples of the database are displayed in Figure 4.1. (The database can be downloaded at (Xia, 2009).) It contains 12 classes of meaningful scenes in high-resolution satellite imagery, including *Airport, Bridge, River, Forest, Meadow, Pond, Parking, Port, Viaduct, Residential area, Industrial area, and Commercial area*; For each class, there are 50 samples. It's worth noticing that the image samples of the same class are collected from different regions in satellite images of different resolutions and then might have different scales, orientations and illuminations.

[2] http://earth.google.com/

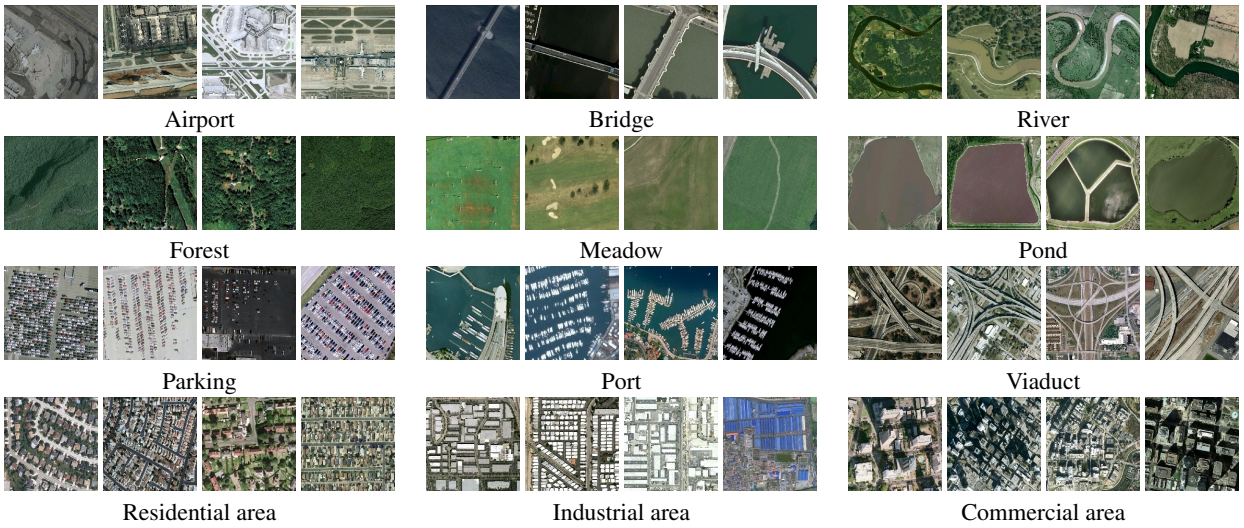


Figure 4: Some samples of the testing high-resolution satellite image database. For each class, there are 50 samples, and 4 of which are shown here.

4.2 Retrieval and classification

We apply the proposed analysis scheme to two common satellite image recognition tasks: retrieval and classification. For retrieval, one sample is used as a query image (thus removed from the database) and the N_r most similar samples are retrieved from the database. One after the other, all samples in the database are used as query images, and the average recall is computed in function of the number N_r of retrievals for evaluating the performance.

In the classification experiment, N_c samples from each class are randomly chosen as a training set and the remaining samples are classified thanks to a nearest-neighbor classifier. The rate of correct classifications is then computed as a function of N_c. In order to consolidate the results, classification rates are averaged on a sequence of 200 random training sets.

Figure 5(a) shows the average retrieval performance on the whole database. It indicates that by using only one sample, averagely 52.84% samples of the same class can be correctly retrieved among the first 49 matches. However, in the same context, if we use the mean and standard deviation of Gabor filter responses, only 21.19% samples can be retrieved averagely. According to the performance curve, when the number of matches is extended to 200, 90.50% samples can be retrieved by using the structural indexes. But the same percentage to Gabor features is 54.13%.

Some illustrations of the retrieval results are displayed in Figure 6, 8(a), 8(b), where a query image is followed by its first 49 closest samples. The retrieval results for all samples can be found at (Xia, 2009).

Figure 6 shows a retrieval result of bridge category, which is very structured. Observe that even though there are large illumination changes between samples and the query image, the method works well, thanks to the contrast invariance of the indexing scheme. It's also interesting to inspect the false alarms and observe that there often contains some structures similar to bridges, see the parts framed in red inside Figure 7. Figure 8(a) and 8(b) illustrate two retrieval examples respectively on river and viaduct class. Even structures in this two classes are complicated, the proposed approach works well.

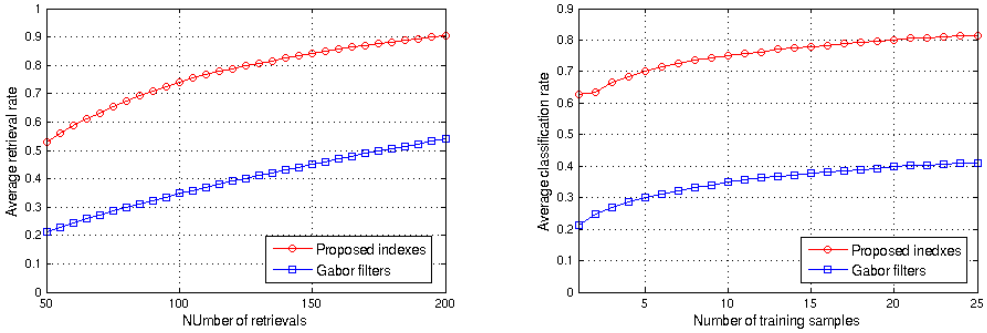
Figure 5(b) shows the average classification performance by using nearest-neighbor classifier, when the number of training images ranges from 1 to 25. It indicates that the structural indexes outperforms the Gabor features dramatically. Furthermore, Table 1 shows the average classification rate for each class of the database. We can see that Gabor features are efficient only on some texture classes, e.g. forest and meadow. The proposed structural indexes work well on classes with complicated structures such as viaduct and airport, and also on classes containing more textures.

However, we found that the structural indexes can not distinguish industrial and residential classes well. This is because those two categories share many similar structures, and some semantic information of the scene might be helpful (Bordes and Maître, 2007).

Figure 7: Some false alarms of bridge retrievals. The parts framed in red really contain some bridge-like structures.

Category	1 training sample		25 training sample	
	StructInd	GaborF	StructInd	GaborF
Airport	53.39	12.72	82.31	39.93
Bridge	40.45	7.73	83.42	27.57
Commercial	48.74	23.17	82.02	43.06
Forest	80.94	42.54	94.62	74.45
Industrial	42.93	20.11	78.37	37.75
Meadow	75.06	35.08	95.45	55.01
Parking	68.74	10.58	81.92	24.89
Pond	63.96	29.91	82.95	48.09
Port	50.93	12.79	73.80	26.66
Residence	32.17	19.04	48.87	40.07
River	60.23	27.28	88.66	49.79
Viaduct	64.19	13.24	86.76	23.78

Table 1: Average classification rate (%) of Structural indexes (using **StructInd** for short) and Gabor Features (using GaborF for short) on each category of the database, with the number of training samples as 1 and 25, respectively.



(a) retrieval performance (b) classification performance

Figure 5: Average retrieval (a) and classification (b) performance.

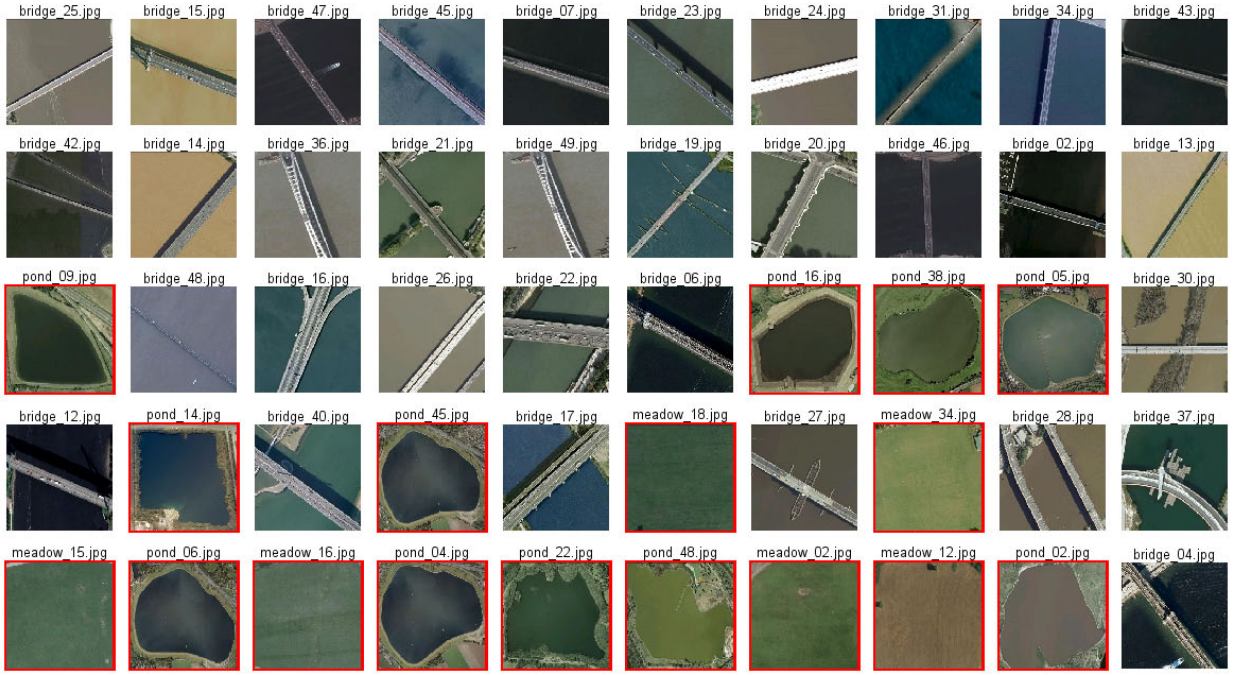


Figure 6: A retrieval result of the bridge category obtained by using the proposed indexing scheme. The query image is in the first position and the 49 most similar samples follow, ordered by their matching scores. The false samples are framed in red.

5 CONCLUSION

In this paper, we have developed some structural features for indexing high-resolution satellite image, based on the topographic map and under the shape-based image indexing framework. The experimental analysis shows that the indexes can balance the structures and textures information in high-resolution satellite images and provide impressive image recognition performances.

However, we should observe that we simply adopted the nearest-neighbor classifier for classification. The recognition performance, of course, benefits from some more powerful classification scheme, e.g. SVM.

ACKNOWLEDGEMENTS

This work was partly supported by a grant from the National Natural Science Foundation of China (No.40801183, 60890074).

REFERENCES

Bhattacharya, A., Roux, M., Maître, H., Jermyn, I. H., Descombes, X. and Zerubia, J., 2007. Computing statistics from man-made structures on the earth's surface for indexing satellite images. International Journal of Simulation Modelling 6(2), pp. 73–83.

Bhattacharya, A., Roux, M., Maître, H., Jermyn, I. H., Descombes, X. and Zerubia, J., 2008. Indexing of mid-resolution satellite images with structural attributes. In: Proc. 21th ISPRS Congress, Commission IV, Beijing, China, pp. 187–192.

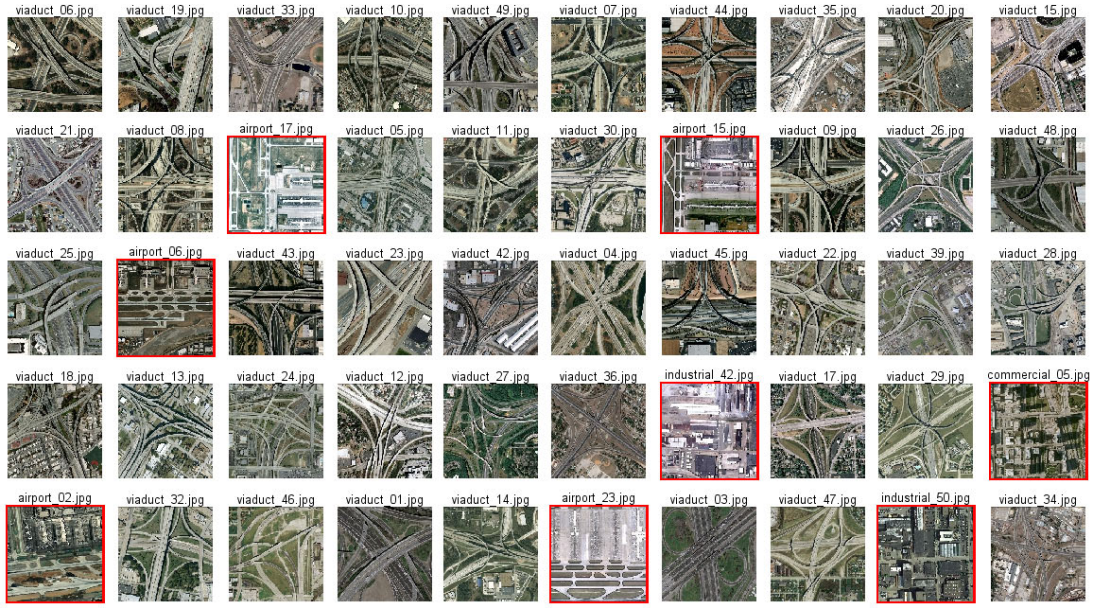
Bordes, J.-B. and Maître, H., 2007. Semantic annotation of satellite images. In: MLDM Posters, pp. 120–133.

Caselles, V., Coll, B. and Morel, J.-M., 1999. Topographic maps and local contrast changes in natural images. International Journal of Computer Vision 33(1), pp. 5–27.

Caselles, V., Coll, B. and Morel, J.-M., 2002. Geometry and color in natural images. J. Math. Imaging Vis. 16(2), pp. 89–105.

Li, C.-S. and Castelli, V., 1997. Deriving texture feature set for content-based retrieval of satellite image database. In: Proceed.

(a) A retrieval example of river category



(b) A retrieval example of viaduct category

Figure 8: Some retrieval results obtained on the database using the proposed indexing scheme. The query image is in the first position and the 49 most similar samples follow, ordered by their matching scores. The false samples are framed in red.

International Conference on Image Processing, IEEE Computer Society, Washington, DC, USA, p. 576.

Monasse, P. and Guichard, F., 2000. Fast computation of a contrast invariant image representation. IEEE Trans. Image Processing 9(5), pp. 860–872.

Newsam, S. and Yang, Y., 2007. Comparing global and interest point descriptors for similarity retrieval in remote sensed imagery. In: GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, ACM, New York, NY, USA, pp. 1–8.

Richards, J. A. and Jia, X., 2005. Remote Sensing Digital Image Analysis: An Introduction. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Ruiz, L. A., Fdez-Sarria, A. and Recio, J. A., 2004. Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study. In: Proc. 20th ISPRS Congress, Commission IV, pp. 1109–1114.

Ünsalan, C. and Boyer, K. L., 2004. Classifying land development in high resolution panchromatic satellite images using straight line statistics. IEEE Trans. Geosci. Remote Sens 42, pp. 907–919.

Xia, G.-S., 2009. High-resolution satellite image indexing project webpage. http://www.tsi.enst.fr/~xia/satellite_image_project.html.

Xia, G.-S., Delon, J. and Gousseau, Y., 2009. Shape-based invariant texture indexing. International Journal of Computer Vision, to appear.

Zhu, S.-C., Guo, C.-E., Wang, Y. and Xu, Z., 2005. What are textons? International Journal of Computer Vision 62(1-2), pp. 121–143.