FAST RETRIEVAL ON REMOTE SENSING IMAGERY BASED ON ORB FEATURE

[1]LEI MA, [2]BITAO JIANG, [3]ZHENGLAN XU

Institute of Automation CAS , Beijing Institute of Remote Sensing Information, Institute of Automation CAS
E-mail: [1]lei.ma@ia.ac.cn, [2]Jiangbt2000@163.com, [3]zhenglan.xu@ia.ac.cn

Abstract - The Content Based Image Retrieval (CBIR) systems have been an active area of research in remote sensing image precessing. Local invariant features, such as SIFT, have been applied to remote sensing image retrieval. In this paper, we employ ORB feature as an alternative to SIFT, to improve the exibility and e_ciency of BoWs model in the codebook training, high-dimensional feature exaction, and quantization. Our method can achieves a reduction in the number of computations of about one order of magnitude, while obtaining the more accuracy. The e_ciency of ORB descriptor are evaluated with UC Merced Land Use-Land Cover data set.

Index Terms—Image Retrieval, Local invariant features, BRIEF feature, ORB feature, Bag of visual words.

I. INTRODUCTION

Remote sensing is currently undergoing a technical revolution with the appearance and blooming development of remote sensors. An increasing number of Earth observation commercial satellites with high-resolution sensors have been launched, so it is difficult to someone to quickly find out what he/she concern.

Automated processes must be developed and refined, which can eliminate the requirement of a human-in the- loop for creating large-scale searchable image repositories. Content-based image retrieval (CBIR) is an increasingly popular retrieval method for large-scale image databases. CBIR queries are not performed in a traditional relational database management system (RDBMS) of image metadata.

CBIR is the most important and effective image retrieval method and widely studied in both academia and industry area. Content Based Image Retrieval (CBIR) seems to have originated with the work of Kato[1] for the automatic retrieval of the images from a database. It is different from retrieval by keywords, tags, or descriptions associated with the image, CBIR search and analyzes the contents of the image depends on colors, shapes, textures, or any other information that can be derived from the images themselves.

Aptoula [2]apply a couple of texture descriptors, the circular convariance histogram and the rotation-invariant point triplets, to represent the content of the images. His work are evaluated with the LULC data set. Because visual knowledge is often related to shape characteristics of objects, shape descriptors always are employed to represent image. Scott et al. [3] construct geospatial information retrieval and indexing system (GeoIRIS), and present a novel indexing structure and shape descriptor.

Oppositely, local feature can overcome the limitation of global feature that occlusions, geometric transformations and illumination changes. Local features detect a set of salient keypoints, then describe the image patch around the key point with a vector as descriptors. Yang and Newsam [4] contribute the first work to investigate local invariant features for geographic image retrieval. Currently two classes of image local features have been proposed: the non-binary descriptor and binary descriptor. The most popular non-binary descriptor include SIFT, SURF, HOG, PCA-SIFT and so on. Lowe proposed the scale invariant feature transform (SIFT) [5] which can be seen as a milestone in the research process of local invariant features.

For the other class of local features, the binary descriptor such as BRIEF [6], BRISK [7], ORB [8] and so on compare the intensities of pixels at different locations to form a series of binary code. Calonder et al. [6] proposed BRIEF descriptor for super-fast description and matching and consists of a binary string computed by simple intensity difference test at random pre-determined pixel locations in the image patches. Despite the obvious advantage in speed and efficiency of this approach, it suffers in terms of reliability and robustness as it has limited tolerance to image rotation, scale changes. BRISK is designed by Leutenegger et al. [7]. It first efficiently detects keypoints in the continuous scale-space based on a detector inspired by FAST and GFAST. Then the BRISK descriptor is composed as a binary string by concatenating the results of simple brightness comparison tests. It is different apart from the BRIEF, BRISK can obtains location, scale and orientation clues for each keypoint, so it achieves orientation-invariance and scale-invariance. ORB descriptor of Rublee et al. [8] is original from BRIEF, but it extract keypoints with a novel o-FAST, and has invariance to the rotation. Despite the limitations in the aspects of distinctiveness and robustness comparing with SIFT, it is a obvious faster than SIFT.

Our approach is inspired by earlier work [4] that showed that investigate local invariant features for geographic image retrieval.

II. METHOD

In this paper, there are several main procedures for image retrieval. First, many key points in each image are detected, then the ORB descriptors with great discriminative power describe the patches around key points within in images. Second, bag of words (BoW) model is applied to quantize the local descriptors into ”visual words”, then represent each image as a vector of words and generate a vocabulary. Final, query image search the result images in vector space rely on the similarity measure.

A. Feature Descriptor

Local feature descriptors play a key role in the image retrieval task. We employ ORB feature descriptor in our methods. ORB improve the binary feature descriptor BRIEF, so in this section, we have a brief introduction in BRIEF firstly.

1) BRIEF: BRIEF (Binary Robust Independent Elementary Features)[9] is a descriptor that relies on a relatively small number of intensity difference tests to represent an image patch as a binary string.
In order to reduce sensitivity for noise, a pre-smoothing the patch is necessary. In the Gaussian smoothing step, a kernel window size is set 9×9 commonly. Then we define test p on patch p of size S ×S as

$$\tau(p;x,y) := \begin{cases} 1 & p(x) < p(y) \\ 0 & otherwise \end{cases} \qquad (1)$$

where p(x) is the intensity of pixel at x in patch p. A set of binary tests are defined by choosing a set of n_d(x,y) location pairs. So BRIEF descriptor can be the n_d-dimensional bitsting

$$f_{n_d}(p) := \sum_{1 \le i \le n_d} 2^{i-1} \tau(p;x_i,y_i) \qquad (2)$$

B. ORB: ORB feature (Oriented FAST and Rotated BRIEF)

ORB is a brand new feature descriptor building on the well-known FAST keypoint detector and the recently-developed BRIEF descriptor. This feature is about an order of magnitude faster than SURF, and over two orders faster than SIFT. ORB feature starts by detecting FAST feature points in the image. The original FAST proposal implements a set of binary tests over a patch, by varying the intensity threshold between the center pixel and those in a circular ring around the center.

The Harris corner measure has been used to provide an evaluation of the corner intensity. However, FAST features do not have an orientation component, which is important for description extraction. In ORB, the intensity centroid assumes that a corner’s intensity is offset from its center. The missing orientation information of FAST are instead complemented with Rosin’s corner intensity. In particular, the moment m_{pq} of a patch can be computed as:

$$m_{pq} = \sum_{pq} x^p y^q I(x,y) \qquad (3)$$

and with these moments the centroid can be found by

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \qquad (4)$$

The vector $\vec{OC}$ can be defined from the corner’s center O to the centroid C. The orientation of the patch then simply is:

$$\theta = a\tan 2(m_{01}, m_{10}) \qquad (5)$$

where atan2 is the quadrant-aware version of arctan. The binary test and the feature fn(p) is defined in the same way. Due to BRIEF can not be invariant to in-plane rotation, a steer BRIEF according to the orientation of keypoints is applied in ORB. For any feature set of n binary tests at location (x_i,y_i), define the 2×n matrix

$$S = \begin{pmatrix} x_1, & \dots, & x_n \\ y_1, & \dots, & y_n \end{pmatrix} \qquad (6)$$

A ”steered” version S_θ of S can be constructed by the patch orientation θ and the corresponding rotation matrix R_θ:

$$S_\theta = R_\theta S \qquad (7)$$

The steered BRIEF operator is defined by:

$$g_n(p,\theta) := f_n(p) \mid (x_i, y_i) \in S_\theta \qquad (8)$$

Finally, ORB descriptor search among all possible binarytest to find a set of uncorrelated test with high variance as result. As shown in Fig.2, ORB has better matching result than SIFT and SURF.

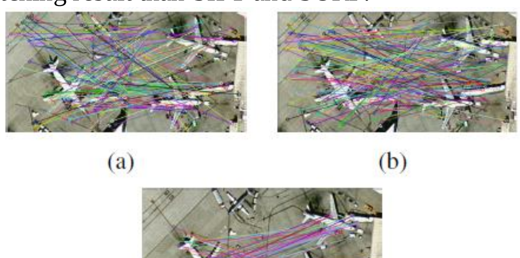

Fig. 1. Matching result using local feature descriptors on remote sensing
images. (a) SIFT. (b) SURF. (c) ORB.

C. Descriptors Quantization and Indexing

Due to the variational number of interest points for each image, a large number of local features must be quantized. We quantize the local descriptors into visual words. Then a standard k-means clustering is applied to generate a dictionary of visual words or codebook with a large number of ORB descriptors. During the iterative k-means clustering and quantization, the distances between SIFT or SURF descriptors and cluster-centroid are computed rely on the Mahalanobis or Euclidean measure. However, ORB descriptors as binary feature descriptors extracted form image are assigned the label of the closest cluster centroid by Hamming distance. So the speed is greatly improved. Based on the vocabulary, query images are represented as a sparse vector of visual word, we can obtain the similar images by calculating the similarity between the query vector and each candidate vector.

III. EXPERIMENTS AND RESULTS

In our experiments, we test our method on UC Merced LULC data set, which consists of 21 categories. There are 100 RGB images (256 pixels by 256 pixels) in each category. The images were manually extracted from large images from the USGS National Map Urban Area Imagery collection for various urban areas around the country. The pixel resolution of this public domain imagery is 1 foot.

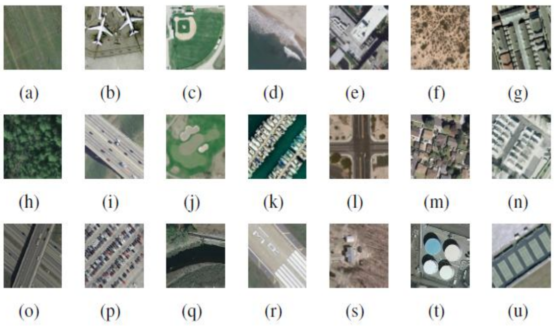

Fig.2. Example images from each class of the UC Merced LULC data set.

(a) Agricultural. (b) airplane. (c) baseball diamond. (d) beach. (e)buildings. (f)chaparral.(g) dense residential. (h)forest. (i)freeway. (j) golf course. (k) harbor.(l) intersection. (m) medium residential. (n) mobile home park. (o) overpass.(p) parking lot. (q) river. (r) runway. (s) sparse residential. (t) storage tanks.(u) tennis courts. The retrieval quality can be measured by ANMRR (average normalized modified retrieval rank), which is accepted by the CBIR community as MPEG-7 retrieval effectiveness evaluation. ANMRR considers not only the number of the ground truth items that appear in the top retrievals, but the order of these. Given a query q ,the size of ground truth is NG(q), Rank(k) represents the position of the kth ground truth itemin the retrieval list. We conduct a series of comparison experiments to demonstrate the effectiveness of our proposed algorithm. Our method based ORB feature is compared with classical feature SIFT and SURF. In BoW step, codebooks is constructed through k-means clustering using three local features. For measuring the similarity between two features, the Euclidean distance is applied to SIFT and SURF, Hamming distance is applied to ORB. For each class of 21 target categories in database, we query 6 times, and 100 retrieval images are obtained each query. Then we average the ANMRRs of 6 times. Fig.3 shows the retrieval performance of the various local features in each 21 classes and average value. The averages of ANMRR of SIFT, SURF and ORB are 0.602787, 0.614163, 0.615101. From ANMRR value as performance description, the tree features is close. We can employ the Confusion matrix to compare the retrieval precision. The rows indicate the query class and the columns the target classes. As shown in Fig.4, ORB has most correct prediction in tree features, due to it has higher the diagonal values of the confusion matrix Oviously. In our experiment, feature descriptors have been tested with a 2.80GHz Intel Core2 Processor and 4G RAM.

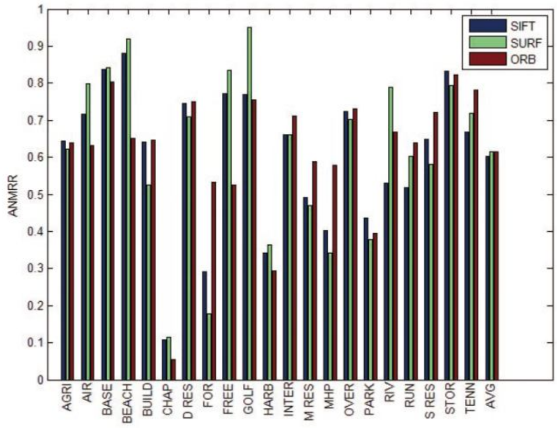

Fig. 3. Per Class retrieval performance of the various tested feature descriptors in terms of ANMRR.

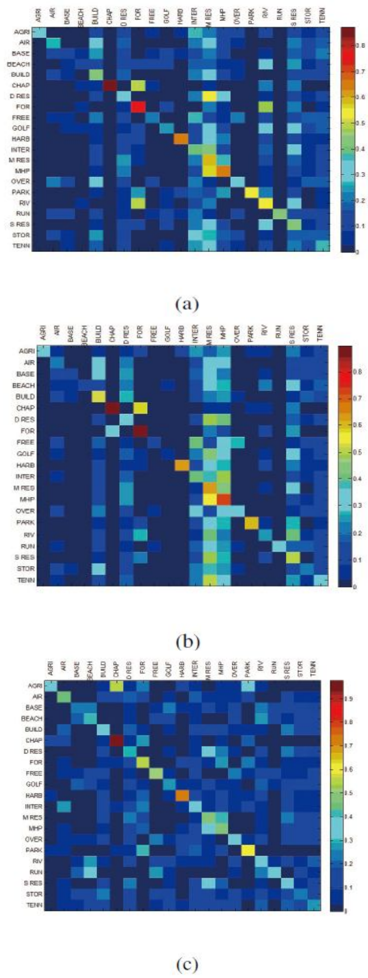

Fig. 4. Confusion matrices corresponding to (a) SIFT. (b) SURF. (c) ORB.

CONCLUSION

In this letter, we employed a new local feature ORB, as an alternative to SIFT, to describe the remote sensing images in retrieval task. Our method can achieves a reduction in the number of computations of about one order of magnitude, while obtaining the more accuracy. Experiment carries out on UC Merced LULC data set, and shows the efficient of our method.

REFERENCES

[1] T. Kato, “Database architecture for content-based image retrieval,” in SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology. International Society for Optics and Photonics, 1992, pp. 112–123.

[2] E. Aptoula, “Remote sensing image retrieval with global morphological texture descriptors,” 2014.

[3] G. J. Scott, M. N. Klaric, C. H. Davis, and C.-R. Shyu, “Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 49, no. 5, pp. 1603–1616, 2011.

[4] Y. Yang and S. Newsam, “Geographic image retrieval using local invariant features,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 51, no. 2, pp. 818–832, 2013.

[5] D. G. Lowe, “Distinctive image features from scaleinvariant keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004. [Online]. Available: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94

[6] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “Brief: Binary robust independent elementary features,” in Computer Vision–ECCV 2010. Springer, 2010, pp. 778–792.

[7] S. Leutenegger, M. Chli, and R. Y. Siegwart, “Brisk: Binary robust invariant scalable keypoints,” in Computer Vision (ICCV), 2011 IEEE

[8] International Conference on. IEEE, 2011, pp. 2548–2555. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: an efficient alternative to sift or surf,” in Computer Vision (ICCV), 2011 IEEE

[9] International Conference on. IEEE, 2011, pp. 2564–2571. M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “BRIEF: binary robust independent elementary features,” in Computer Vision - ECCV 2010, 11th

★★★