

## Research Article

# Mobile Visual Recognition on Smartphones

Zhenwen Gui,<sup>1</sup> Yongtian Wang,<sup>1</sup> Yue Liu,<sup>2</sup> and Jing Chen<sup>2</sup>

<sup>1</sup> School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China

<sup>2</sup> School of Optoelectronics, Beijing Institute of Technology, Beijing, China

Correspondence should be addressed to Zhenwen Gui; [quizhenwen1983@bit.edu.cn](mailto:quizhenwen1983@bit.edu.cn)

Received 31 May 2013; Revised 23 September 2013; Accepted 28 September 2013

Academic Editor: Eugenio Martinelli

Copyright © 2013 Zhenwen Gui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper addresses the recognition of large-scale outdoor scenes on smartphones by fusing outputs of inertial sensors and computer vision techniques. The main contributions can be summarized as follows. Firstly, we propose an ORD (overlap region divide) method to plot image position area, which is fast enough to find the nearest visiting area and can also reduce the search range compared with the traditional approaches. Secondly, the vocabulary tree-based approach is improved by introducing GAGCC (gravity-aligned geometric consistency constraint). Our method involves no operation in the high-dimensional feature space and does not assume a global transform between a pair of images. Thus, it substantially reduces the computational complexity and memory usage, which makes the city scale image recognition feasible on the smartphone. Experiments on a collected database including 0.16 million images show that the proposed method demonstrates excellent recognition performance, while maintaining the average recognition time about 1 s.

## 1. Introduction

In recent years, smartphone has developed rapidly, almost all of inexpensive smartphones are equipped with cameras, GPS, wireless network, and gravity sensing. The improvements in imaging capabilities and computational power have given rise to many exciting mobile applications. Among these is mobile visual location recognition where users can take pictures of the place of interest by using their smartphone, to find the corresponding information related to the captured landmark anywhere [1–3].

Most current applications adopt client-server (C/S) mode to transfer image information [4, 5] (such as compressed image, image descriptors, and image location) to a remote server through, wireless network or 3G, on which a searching process will be carried out to, then the related information will be returned to phones for observation. In such systems, sets of local features [6–9] are used to represent images information, and image matching algorithms are based on vocabulary tree (VT) [10–12]. Features of the query image are quantized into visual words through the VT algorithm and then scalable textual indexing and retrieval schemes are

applied to find similar candidate image from the database [10]. However, there are some inherent limits in the existing systems. For example, the growing city scale candidate images need more time for retrieval, which will affect the efficiency of mobile visual recognition applications. Moreover, the words quantization losing discriminative power and spatial relations of the features will reduce the recognition accuracy.

We propose to realize city scale mobile visual recognition directly on smartphone to solve above problems. First, ORD algorithm is developed to plot the candidate images to a specific region according to the captured GPS information. It can obviously reduce searching time since it narrows the matching range largely. Second, a GAGCC algorithm is designed to refine geometric voting score of matching descriptors to rerank retrieval result.

The rest of this paper is organized as the follows. Section 2 gives the related work. Section 3 gives the overview of the recognition system. Section 4 presents database construction. Section 5 gives the outdoor scenes recognition. Section 6 shows the experimental results. Finally we conclude in Section 7.

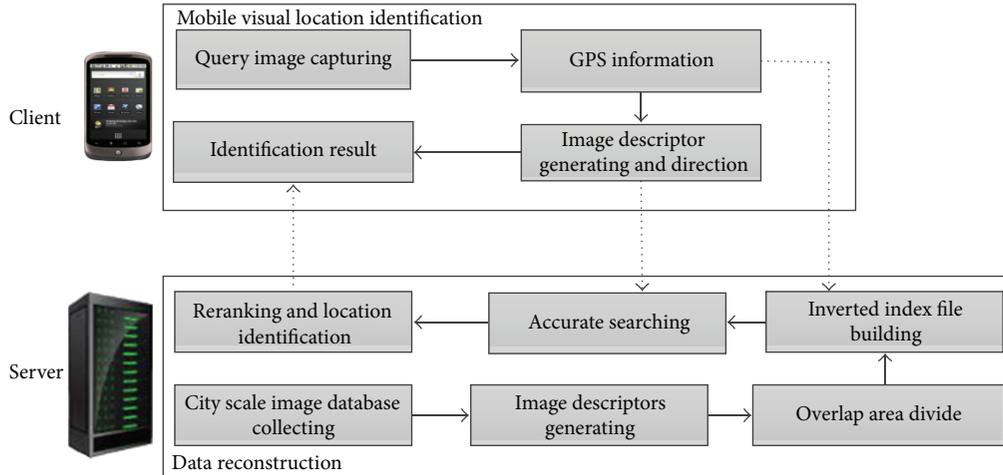


FIGURE 1: Smartphone recognition system process flow.

## 2. Related Work

Outdoor scenes recognition at the city scale is closely related to the image retrieval or large-scale object recognition problems. A common used scheme [10] builds on top of the bag-of-features model, extracting local image features, quantizing their descriptors to visual words and applying methods from text search for image retrieval [10, 13]. However, most of these works do not exploit enough spatial relationship among features and rely mainly on visual words for recognition.

Several researchers have improved the retrieval accuracy of this approach from different perspectives; for example, introducing a postspatial verification by PROSAC [14]; applying the efficient reranking in the initial retrieval results as queries [15]; imposing weak geometry constraints [3, 16] and contextual weighting assign to local features in both descriptor and spatial domains [17]. Most of these methods involve considerable calculations, for example, the spatial verification or candidate rerank is at the cost of greater complexity. Instead, our method exploits fast geometric consistency constraints, gets partial geometrical information without explicitly estimating a global transform between the query and database images. The method is integrated into the inverted file and can efficiently be applied to all images.

Meanwhile, in order to improve the accuracy, some researchers use additional sensors such as GPS to assist the mobile landmark recognition. Takacs et al. [1] use GPS information to retrieve only images falling in nearby location cells. Kumar et al. [18] use GPS to search for local vocabulary trees to speed up the visual searching process. Whereas they ignore the boundary region and GPS error, it is a problem to determine which area these images belong to when the users are located in the boundary of the two regions. When receiving the noisy GPS by satellites or network, using GPS to search for the nearest neighbor descriptor is inaccurate. We also design our system so that the mobile subscribers can query information of interest anywhere in the city.

In this research, we first propose an ORD method to overcome the problems of traditional fixed region bringing

boundary and GPS error problems and provide more accurate search results, which facilitates the city scale outdoor scenes recognition applications.

## 3. Overview of the Recognition System

This section gives an overview of the proposed outdoor scenes recognition system on mobile devices. As can be seen from Figure 1, the system can be divided into two parts: data preparation and visual location recognition. While the data preparation process is carried out on a server, the visual, GPS, and gravity direction information is obtained directly from phones.

Data preparation mainly deals with the problems of database images collection, image descriptors generating and area divide, and inverted index file constructing. With the database images reconstruction, we firstly use the method proposed in Section 4 to generate the image descriptors. Secondly, we plot overlap area and build a memory-efficient inverted index structure by integrating the information from both the visual and additional sensors in Section 4. Visual location recognition relies on the image information itself such as feature descriptor, GPS, and orientation to support.

To perform location recognition on mobile devices, GPS information will firstly be used to carry out a coarse search process to find the region candidate image set by using the method proposed in Section 4. Then, the image descriptor of the query image will be generated to find several similar images from the candidate set. Finally, the integration method proposed in Section 5 will be taken to implement effective reranking for the search results to fulfill the location recognition task.

## 4. Database Construction

We collect the outdoor scene images database by many people using high-resolution smartphone camera. The GPS and direction information of each image are recorded by using the built-in GPS and gravity sensor, respectively. We write



FIGURE 2: Procedures for collecting sensor data on smartphone.

an Android application to capture the phone-side image, GPS information, gravity acceleration data and as shown in Figure 2.

We are working with a dataset for Beijing that contains 0.16 million architectural images of many famous scenic spots, shopping malls, libraries, food city households, campus and so on. Each outdoor scene contains 10 images and a total of 16 k scenes are visited to generate the database which will be used in our mobile visual location recognition system. Figure 3 illustrates some dataset images.



FIGURE 3: Dataset examples.

**4.1. Overlap Region Divide.** Previously, in order to avoid matching the entire city sample images, city landmarks retrieval system usually divides the whole city to some regions according to the geographic information as the upper part of Figure 4. Each image is assigned to a fixed area based on the shooting scene GPS. When querying an image, you only need to match with region images other than all scene samples.

However, there are two problems on the fixed area dividing method. First, due to the existence of GPS error, the shooting scenes cannot be correctly positioned for the real region; second, when the user is on the boundary of the two regions, the user’s current query image cannot be determined to belong to which regions. So, we propose an ORD method to solve the above problems.

The ORD method divides a city into a number of fixed-size regions, then makes four adjacent regions as an overlap area. It can be seen from the upper part of Figure 4 that a city with the same size of 6 regions, four adjacent regions form an overlap area and a total of three overlap areas. Each overlap area represented by the center point coordinates. As shown in Figure 4, the center of the overlap area is marked with blue characters O1, O2, and O3 and we can get the center point location through (1). We suppose the four corner points GPS value of the overlapping area as  $(x_1, y_1)$ ,  $(x_2, y_1)$ ,  $(x_1, y_2)$ ,  $(x_2, y_2)$ , the center point as  $(c_1, c_2)$ .

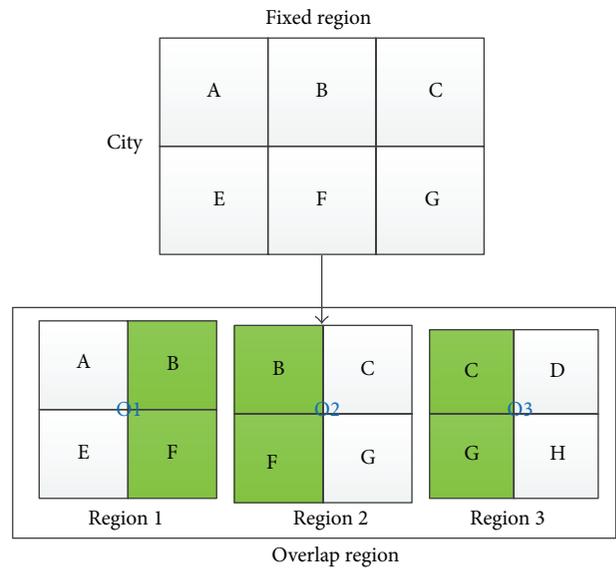


FIGURE 4: Overlap region divide.

Consider

$$\begin{aligned} c_1 &= \frac{x_1 + x_2}{2}, \\ c_2 &= \frac{y_1 + y_2}{2}. \end{aligned} \tag{1}$$

When determining scene which image belongs to which overlap region, simply selects the area with the shortest distance between center point of each overlap area and scene GPS point of the current scene. When the distance from the user to regional center is closer, the scene of the region can be better covered for the acquisition object of interest by user.

Each overlapping region contains two same fixed regions scene images, so the original boundary problem does not exist in the overlap region. When users go to the boundary of the adjacent regions, there are corresponding sample images. The GPS error is not greater than 100 m on the worst case, while the size of the fixed area is often measured in units of

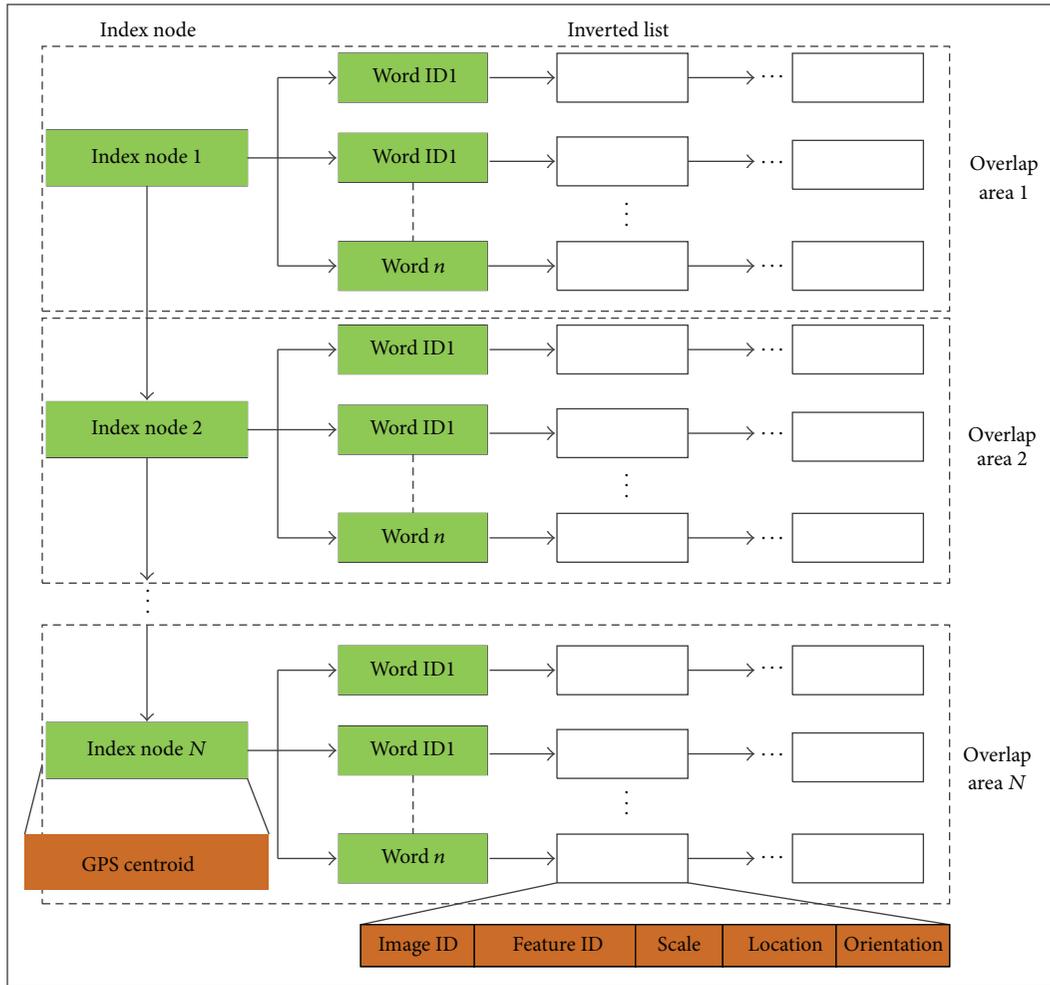


FIGURE 5: Overlap area inverted file index structure.

km that is much larger than the GPS error, so it has no impact to locate the overlap area.

**4.2. Building Inverted Index.** We store all center points GPS information of the overlap area on the server to build the inverted index structure, using the center point as an index for each overlap area to create an inverted linked list as illustrated in Figure 5.

To index a database image, we need to firstly establish a hierarchical vocabulary tree. All sample images are resized to  $320 \times 240$  resolution. We use OPENCV 2.3.1 SURF algorithm to extract SURF features from the database images and Hierarchical K-means algorithm [10] to train the vocabulary tree with depth 6 and branch factor 10. Secondly, we quantize image descriptor to word through VOC tree, and calculate the descriptor corresponding scale, location, and orientation by using the method proposed in Section 5. Finally, we use the obtained GPS and the quantized word of the current image to find the corresponding inverted list. A new entry which contains the image ID, feature ID, scale, location and orientation will be added to the found inverted list for online searching use.

## 5. Outdoor Scenes Recognition

This section gives the method to perform location recognition by integrating vision and inertial sensors. The detailed algorithm is described as follows.

*Step 1.* Find the nearest overlap area by using the GPS information of the query image to find candidate inverted lists in the found overlap area.

*Step 2.* Generate the SURF descriptors of the query image by using the method discussed in Section 4.

*Step 3.* Compute the weight vector of the query image by the vocabulary tree that has already been trained on in Section 4.

*Step 4.* Apply next section algorithm to calculate the word weights and rearrangement of similar sample image, choosing the highest score as the recognition result.

**5.1. Fast Geometric Consistency Constraint.** Currently, many reranking algorithms create some problems in image retrieval, including computational complexity and large

memory footprint as literature [17]. The literature [17] which introduced an improved algorithm requires all matching feature pairs to calculate the descriptor contextual weighting and spatial contextual weighting and also needs much memory to save the number of neighborhood characteristics, orientation and scale variables. So, we propose an FGCC (fast geometric consistency constraint) approach to optimize the contextual weighting with smaller computation and lower memory footprint.

We select the salient matching descriptors to verify the consistency of the angle and scale of a given image. When a descriptor is classified using a VT, the descriptor is compared with the children of a node and the most similar child is selected. The process starts from the root and is repeated until reaching the leaf node, to generate a path from the root to the leaf within the tree. Typically, the similar descriptors tend to be quantized along the same path. Thus, we generate matching feature pairs of query features and candidate features by quantizing. We choose the matching feature pairs set  $C$ , that is, only one query descriptor and only one candidate descriptor that has been classified to leaf node within the VT. In Figure 6, we show the matching pairs  $S$  of the query image and candidate image which falling into the red circle. When this condition is satisfied,  $S$  belongs to  $C$ .

Given two descriptors quantized to the same tree node for the set of  $C$ , we measure the fast geometric consistency of their local neighborhood and add a spatial context term  $S(i, j)$  to the matching score. This is obtained by modifying the score formalism of VOC as follows:

$$f(x, y) = \begin{cases} tf - idf(q(x)) + S(x, y), & (x, y) \in C, \\ tf - idf(q(x)), & \text{if } q(x) = q(y), \\ \text{otherwise.} & \end{cases} \quad (2)$$

Variables  $x, y$  represent descriptor  $q(x), q(y)$  quantizing descriptor into the tree node,  $tf - idf()$  is the weight function of the node,  $S(x, y)$  indicates the angle and scale similarity score, which is calculated by

$$S(i, j) = \max(W_{i,j}^s, W_{i,j}^\theta). \quad (3)$$

$W_{i,j}^s, W_{i,j}^\theta$  can be calculated using the method mentioned in the literature [17]. A SURF feature  $f_0 = \{X_0, L_0, S_0, \theta_0\}$  includes the descriptor  $x_0 \in R^D$ , location  $L_0$ , feature scale  $S_0$  (in the log domain), and orientation  $\theta_0$ . Let  $C(f_0)$  denote the neighborhood of this feature given by the disc  $(u_0, R)$ . We set the radius  $R = 12 \times S_0$  by subsequent experiments. We calculate 3 statistics of  $C(f_0)$ , that is, the descriptor density  $\rho$ , the mean relative log scale  $\Delta s$ , and the mean orientation difference  $\Delta\theta$ , with respect to  $f_0$  defined as

$$\begin{aligned} \rho &= |C(f_0)|, \\ \Delta s &= \frac{1}{|C(f_0)|} \sum_{f \in C(f_0)} |S - S_0|, \\ \Delta\theta &= \frac{1}{|C(f_0)|} \sum_{f \in C(f_0)} |\theta - \theta_0|, \end{aligned} \quad (4)$$

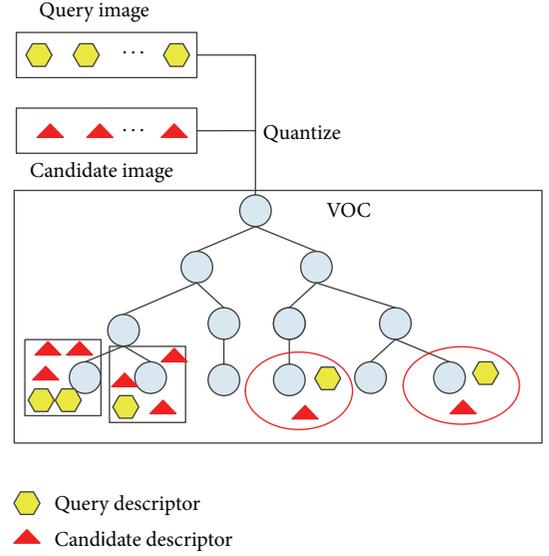


FIGURE 6: Matching pairs of the vocabulary tree.

where  $|C(f_0)|$  is the number of descriptors within  $(u_0, R)$ . These statistics are translation, scale, and rotation invariant. Given two descriptors quantized to the same tree node, we measure the consistency of their local neighborhoods and add a spatial context term  $W_{i,j}^C$  in the matching score. The matching for each statistic in the range of  $[0, 1]$  is defined as follows:

$$W_{i,j}^s = \frac{\min(\Delta S_i, \Delta S_j)}{\max(\Delta S_i, \Delta S_j)}, \quad (5)$$

$$W_{i,j}^\theta = \frac{\min(\Delta\theta_i, \Delta\theta_j)}{\max(\Delta\theta_i, \Delta\theta_j)}. \quad (6)$$

These simple angle and scale score statistics in set  $S$  effectively enhance the descriptive ability of individual features with a small computational overhead. We count the similarity scores of the entire image for reranking in matching images. It only needs less memory to save  $\rho$ ,  $\Delta s$ , and  $\Delta\theta$  for the candidate descriptor in the set  $S$ .

**5.2. Gravity-Aligned Geometric Consistency Constraint.** Many kinds of image descriptors such as SIFT, SURF, and BRISK have been designed recently for image retrieval purpose. We have found the GAFD (gravity-aligned feature descriptors) [19] to be eminently suitable for our research because they are both memory efficient and reasonably easy to implement. While flexible for large-scale image retrieval task, we do not use the SURF descriptors directly in our research since it does not take advantages of the additional sensors of mobile devices. In view of this, we propose GAGCC method to use gravity information to improve the performance of FGCC. The detailed modifications are as follows.

Firstly, in the original work, orientation aligned local features are used to make the generated SURF descriptors invariant against rotations. However, as discussed in [19]

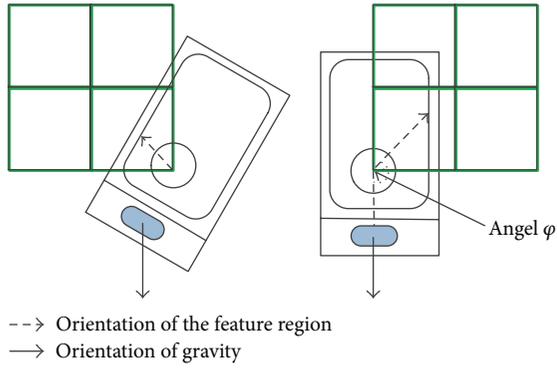


FIGURE 7: Angle between the gravity direction of the feature point and the main direction of the feature region.

the orientation computed from the pixel intensities results in problems when dealing with congruent or near-congruent features. As shown in part 1 of Figure 7, the regions around the four corners that correspond to four different natural features will be identical after normalization by using the orientations computed from the pixel intensities. While reducing the matching accuracy between local descriptors, the above phenomenon will also reduce the discriminating power of SURF descriptors obviously. The use of gravity can solve the above problem to some degrees. As shown in the left side of Figure 7, by aligning local features according to the gravity, the differences between the four normalized regions are more noticeable. In fact, the gravity has been used in [19] to solve the natural features matching problem on mobile phones. In this research, we demonstrate that the use of gravity-aligned local features can also improve the discriminating power of vocabulary descriptor obviously.

Secondly, in the original method, only local descriptors are used to generate the needed vocabulary descriptors. Another important factor that is, the orientation of local feature is absolutely ignored. This is mainly because no reference direction can be used to evaluate the orientation differences between different local features. We solve this problem by setting the gravity as the reference direction. As shown in the right side of Figure 7, the rotation angle  $\varphi$  between the gravity and the orientation computed from the pixel intensities can also be used to differentiate different local features.

We modify (6), replacing  $\theta$  with  $\varphi$ , and set a threshold angle  $= \pi/6$ , as in

$$W_{i,j}^{\varphi} = \begin{cases} \frac{\min(\Delta\varphi_i, \Delta\varphi_j)}{\max(\Delta\varphi_i, \Delta\varphi_j)}, & |\varphi_i - \varphi_j| \leq \text{Angle} \\ 0, & |\varphi_i - \varphi_j| > \text{Angle}. \end{cases} \quad (7)$$

When the absolute value of the difference of the angle between the two features is greater than the threshold value, we set the reference orientation weight  $W_{i,j}^{\varphi}$  to 0 because the two feature points are less similar if the difference of angle is larger.

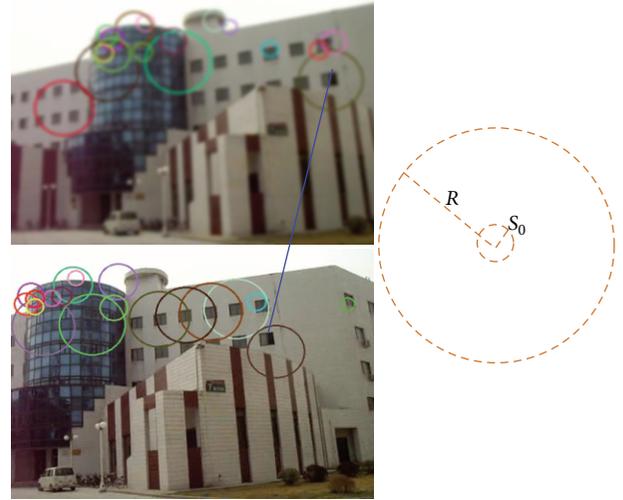


FIGURE 8: Illustration of the relationship between  $R$  and scale  $S_0$ .

## 6. Experimental Results

To test our algorithm's performance for mobile visual location recognition, we built a prototype system. On server side, the retrieval system is implemented in C++ on WIN7 64-bit operating system with 2.8 GHz processor. On client side, the system is implemented in java on HTC smartphone with 1 GHz processor.

**6.1. Measurement.** We select 5 images of each scene as the query images, others as train images in our experiments. We define correct matching for the query image and sample image in the results corresponding to the same scene.

To evaluate the performance, the parameters of average precision (AP) and average time are used. We compute an average precision score for each of the 5 queries for a scene. The average recognition time derived from the 1000 scene average processing time, including the feature extraction time, round-trip transmission time, and search time.

**6.2. Evaluation.** We test the performance of the proposed recognition at different  $R$  value.  $R$  value is related to extracted feature scale. The scale represents blur degree on image. When the image has a large-scale, it will turn greater blur and have less feature points. Figure 8 illustrates the scale of the feature point corresponding to the radius  $R$ . Therefore, we choose the radius  $R$  of the feature neighborhood to be exponentially larger than the scale of feature point. Figures 9 and 10 give the recognition accuracy and time when using different neighborhood radius  $R$  to calculate reranking score through FGCC algorithm. We can see that searching performance of our method is better when  $R$  equals  $12 \times S_0$ . The recognition accuracy is not only high, but also the average recognition time is short. So, we select the neighborhood radius  $R$  to  $12 \times S_0$ .

We test the influence of the threshold parameters angle used in our method. Figure 11 gives the search performance when varying the angle used in GAGCC method. We can see

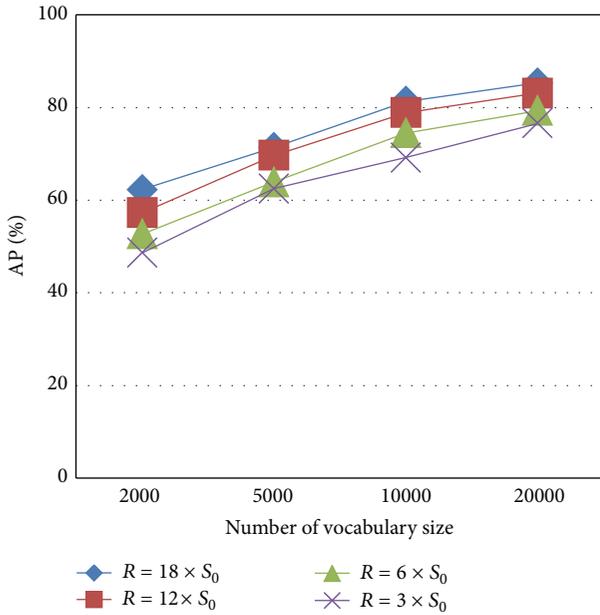


FIGURE 9: Average recognition accuracy of FGCC at different R-values.

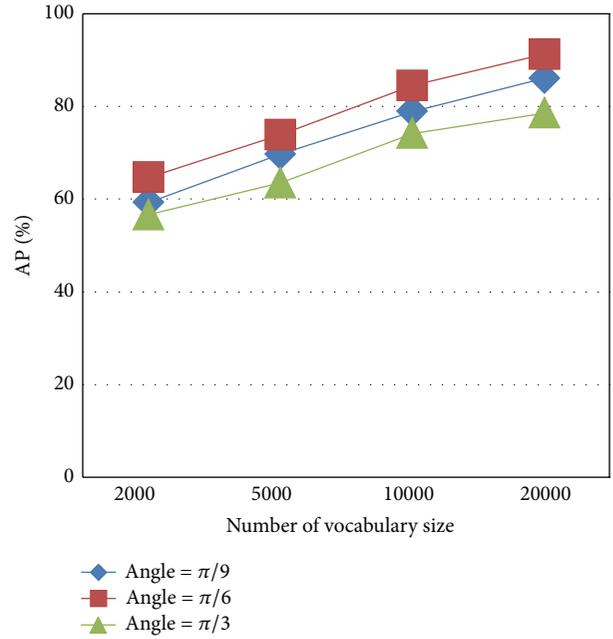


FIGURE 11: Average accuracy of GAGCC at different angles.

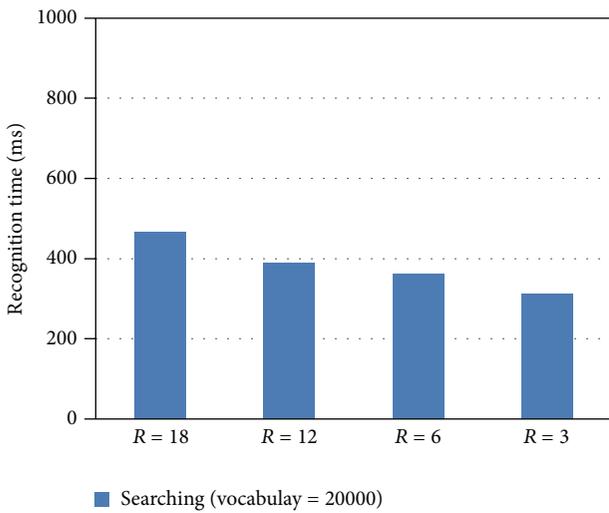


FIGURE 10: Average recognition time of FGCC at different R-values.

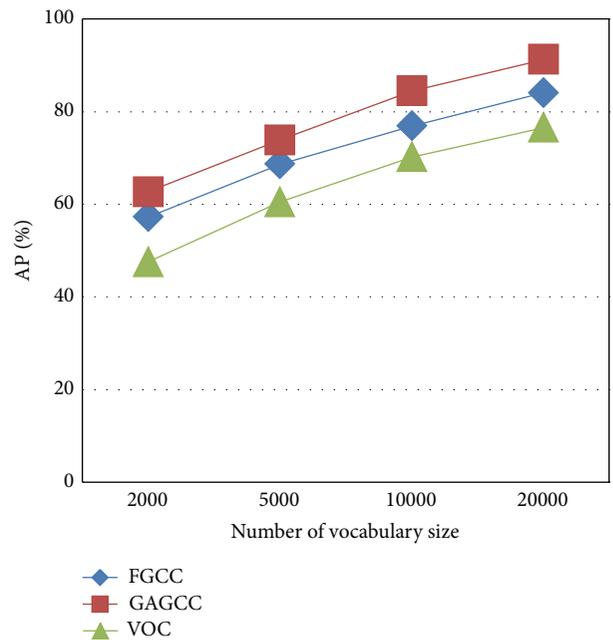


FIGURE 12: Performance comparison of the original VOC and proposed method.

that the search performance is not always improved with the decreasing of angle values. This is mainly because the errors of gravity sensor and noise of pixel intensity orientation will impair the discriminating power of the orientation parts when the angle value is too small. In our system, we set angle to  $\pi/6$ , which is found to be a reasonable tradeoff between the discriminating power and noise restraint.

We compare the performance of our approach with the classical VOC retrieval method [10]. Recognition accuracy between the proposed methods and VOC is shown in Figure 12. By incorporating ORD and FGCC methods, or ORD and GAGCC methods, the recognition result of proposed method is significantly higher than that of VOC

method Because smaller query range and more optimized scoring of the proposed method have a clear advantage over the VOC without any assistance.

Figure 13 shows the average time for whole outdoor scene recognition process, including the generating descriptor time on smartphone, transmission time, and searching database time. The total time of single scene image recognition is about 1s under the WIFI network. Table 1 shows total recognition

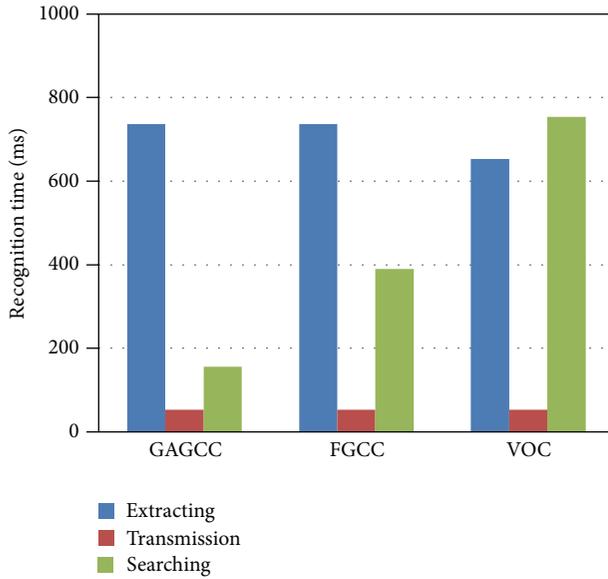


FIGURE 13: Average time for recognition.

TABLE 1: Average time for recognition in a database containing 0.16 million image.

Method	ORD + GAGCC	ORD + FGCC	VOC
Time (s)	1.18	0.95	1.46

time for one query image on 0.16 M dataset. Our method consumes less time than VOC [10], because VOC needs to search for all region candidate images.

## 7. Conclusions

Outdoor scenes recognition on smartphone is a challenging task. Several methods are developed in this research to make recognition performance become stable; the first method is ORD which narrows the search range to target object, so it reduces the recognition time and the probability of matching error on similar scenes in the different areas; the second method is utilizing fast geometric consistency constraints to rerank retrieval result. These methods enable mobile visual recognition more effectively on large-scale image sets.

## Acknowledgment

The authors thank Xue Kang and Zhang Jianguan for reviewing the paper. This research is supported by the National Natural Science Foundation of China (NSFC) under Grant no. 61072096 and the National Science and Technology Major Project of China under Grant no. 2012ZX03002004.

## References

- [1] G. Takacs, V. Chandrasekhar, N. Gelfand et al., “Outdoors augmented reality on mobile phone using loxel-based visual feature organization,” in *Proceedings of the 1st International ACM Conference on Multimedia Information Retrieval (MIR '08)*, pp. 427–434, New York, NY, USA, August 2008.
- [2] G. Schroth, R. Huitl, D. Chen, M. Abu-Alqumsan, A. Al-Nuaimi, and E. Steinbach, “Mobile visual location recognition,” *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 77–89, 2011.
- [3] S. S. Tsai, D. Chen, G. Takacs et al., “Fast geometric re-ranking for image-based retrieval,” in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 1029–1032, Hong Kong, September 2010.
- [4] D. M. Chen, G. Baatz, K. Köser et al., “City-scale landmark identification on mobile devices,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 737–744, Providence, RI, USA, June 2011.
- [5] G. Baatz, K. Koeser, D. Chen, R. Grzeszczuk, and M. Pollefeys, “Handling Urban Location Recognition as a 2D Homothetic Problem,” in *Proceedings of the IEEE European Conference on Computer Vision*, pp. 266–279, 2010.
- [6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [7] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: speeded up robust features,” in *Proceedings of the 9th IEEE European Conference on Computer Vision*, pp. 404–417, Graz, Austria, May 2006.
- [8] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, “CHoG: compressed histogram of gradients a low bit-rate feature descriptor,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 2504–2511, Miami, Fla, USA, June 2009.
- [9] S. Leutenegger, M. Chli, and R. Y. Siegwart, “BRISK: binary robust invariant scalable keypoints,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2548–2555, Barcelona, Spain, November 2011.
- [10] D. Nistér and H. Stewénius, “Scalable recognition with a vocabulary tree,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 2, pp. 2161–2168, June 2006.
- [11] z. Wu, Q. Ke, M. Isard, and J. Sun, “Bundling features for large scale partial-duplicate web image search,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 25–32, Miami, Fla, USA, June 2009.
- [12] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, “A multi-sample, multi-tree approach to bag-of-words image representation for image retrieval,” in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 1992–1999, Kyoto, Japan, October 2009.
- [13] J. Sivic and A. Zisserman, “Video google: a text retrieval approach to object matching in videos,” in *Proceedings of the 9th IEEE International Conference on Computer Vision*, vol. 2, pp. 1470–1477, Nice, France, October 2003.
- [14] O. Chum and J. Matas, “Matching with PROSAC-Progressive sample consensus,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 220–226, June 2005.
- [15] X. Wang, M. Yang, and K. Yu, “Efficient re-ranking in vocabulary tree based image retrieval,” in *Proceedings of the 45th Asilomar Conference on Signals, Systems and Computers*, pp. 855–859, 2011.
- [16] H. Jégou, M. Douze, and C. Schmid, “Hamming embedding and weak geometric consistency for large scale image search,”

- in *Proceedings of the European Conference on Computer Vision*, pp. 304–317, October 2008.
- [17] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, and T. X. Han, “Contextual weighting for vocabulary tree based image retrieval,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 209–216, Barcelona, Spain, November 2011.
- [18] A. Kumar, J.-P. Tardif, R. Anati, and K. Daniilidis, “Experiments on visual loop closing using vocabulary trees,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [19] D. Kurz and S. Ben Himane, “Inertial sensor-aligned visual feature descriptors,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 161–166, Providence, RI, USA, June 2011.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

