Local Descriptors without Orientation Normalization to Enhance Landmark Regconition

Dai-Duong Truong, Chau-Sang Nguyen Ngoc, Vinh-Tiep Nguyen,
Minh-Triet Tran, and Anh-Duc Duong

Abstract. Derive from practical needs, especially in tourism industry; landmark recognition is an interesting and challenging problem on mobile devices. To obtain the robustness, landmarks are described by local features with many levels of invariance among which rotation invariance is commonly considered an important property. We propose to eliminate orientation normalization for local visual descriptors to enhance the accuracy in landmark recognition problem. Our experiments show that with three different widely used descriptors, including SIFT, SURF, and BRISK, our idea can improve the recognition accuracy from 2.3 to 12.6% while reduce the feature extraction time from 2.5 to 11.1%. This suggests a simple yet efficient method to boost the accuracy with different local descriptors with orientation normalization in landmark recognition applications.

1 Introduction

In context-aware environment, applications or services are expected to wisely recognize and understand current user contexts to generate adaptive behaviors. Context information can be classified into external and internal contexts. External contexts include information can be about location, time, or environmental factors such as light condition, temperature, sound, or air pressure. Internal contexts are related to information that mostly specified by users, e.g. events, plans, or even emotional states. Various methods and systems are proposed to capture and process the wide variety of context information, such as location-based services using network [1] or GPS [2] data, sensor-based [3], audio-based applications [4], or visual based

Dai-Duong Truong · Chau-Sang Nguyen Ngoc · Vinh-Tiep Nguyen · Minh-Triet Tran
Faculty of Information Technology, University of Science, VNU-HCM, Vietnam
e-mail: {nvtiep,tmtriet}@fit.hcmus.edu.vn

Anh-Duc Duong
University of Information Technology, VNU-HCM, Vietnam
e-mail: ducda@uit.edu.vn

V.-N. Huynh et al. (eds.), *Knowledge and Systems Engineering, Volume 1*,
Advances in Intelligent Systems and Computing 244,
DOI: 10.1007/978-3-319-02741-8_34,

systems, e.g. visual search [5], gesture recognition [6], natural scenes categorization [7], template matching [8], etc. With the continuous expansion of capability of personal devices and achievements in research, more and more approaches are proposed to explore contexts to provide better ways to interact with users.

The portability of mobile devices helps people access information immediately as needed. This property makes mobile devices become an essential and promising part of context-aware systems. Derive from social practical demands, especially in tourism industry; landmark recognition is one of the problems with increasing needs. This problem is a particular case of natural scene classification but limits only for places of interest. Landmark recognition applications on mobile devices usually use a general architecture in which the first step is extracting features from captured images. The two most popular approaches for this step are dense sampling [9] and local features extraction. Dense sampling can yield a high accuracy but require a high computational cost which may not be appropriate for mobile devices. Our research focuses on providing a simple method to boost the accuracy of landmark recognition using the local feature approach.

Most of local feature based landmark recognition systems [10, 11, 12] take the traditional approach using local descriptors with orientation normalization. However we show that it is not a good choice for landmark recognition problem. Our idea is inspired by the result of Zhang [13] which shows that descriptors equipped with different levels of invariance may not always outperform the original ones. We take a further step to conduct experiments to prove that in landmark recognition problem, rotation invariance not only is unnecessary but also may decrease the accuracy. This can be explained by the loss of discriminative information of a descriptor during the computation process to make it rotation invariant. It should be noticed that users tend to align their cameras so that images are usually captured in landscape or portrait orientation. This makes rotation invariance become not much efficient in this case.

In this paper, we present our proposed idea then experiments to show that the elimination of the orientation normalization step can enhance the accuracy of common local features. We test three different descriptors, including SIFT, SURF, and BRISK, that require identifying dominant orientation(s) and orientation normalization. We use Bag of Visual Words (BoVWs) [14] which is the basic framework of many state-of-the-art methods in the problem of landmark recognition. Experiments are conducted on two standard datasets: Oxford Buildings [15] and Paris[16] datasets. The remainder of our paper is organized as follows. In section 2, we briefly introduce landmark recognition problem and some state-of-the-art approaches using BoVWs model. In section 3, we describe our proposal. Experimental results are presented and discussed in section 4 while the conclusion and future work are in section 5.

2 Background

One of the major goals of intelligent systems is to provide users natural and simple ways to interact with while still get necessary inputs to generate the response. In

order to achieve that, these systems need to be equipped with the capability to recognize external environment and context to generate appropriate behaviors. Visual data is one of the most informative data source for intelligent systems. A popular problem in vision based systems is static planar object recognition with a wide range of applications, especially in augmented reality [17, 18]. The basic solution for this problem is directly matching a query image with existing templates [8]. Different measures can be used to obtain different levels of robustness and distinctiveness. One limitation of template matching is that the relative positions of pixels to be matched should be preserved. A small change in viewpoint may lead to a wrong match. To overcome this disadvantage, proposed methods are mostly focus on efficient ways to detect key points [19, 20], interest points that are stable to brightness, illumination, and transformation variations, and describe them [21]. However, in problems related to scene or object classification with a wide intra-class variation, these methods are not good enough to provide an adequate accuracy. State-of-the-art systems usually use highlevel presentations to eliminate this intra-class variation [9, 14, 22].

To deal with the wide intra-class variation, state-of-the-art approaches in landmark recognition systems obtain a high-level presentation of features in images. The method is called Bag of Features (BoFs) or Bag of Visual Words (BoVWs) [14]. This idea is borrowed from text retrieval problem where each document is described by a vector of occurrence counts of words. In specific, a codebook containing a list of possible visual words corresponding to common patches in scenes is built. Each image will then be described by the histogram of distribution of these visual words. Using a histogram, BoVWs loses the information about spatial distribution of these visual words. Spatial Pyramid Matching (SPM) [22] solves this drawback by introducing a pyramid of histograms in which each histogram captures the distribution of visual words in a specific region at a particular resolution. Locality-constrained Linear Coding (LLC) [9] takes a further step to loosen the constraint that each feature can only belong to a single visual word. In LLC, each feature can belong to several different visual words in its vicinity.

In existing methods of image classification and landmark recognition, uniform dense sampling or local detectors are used to extract key points from images. Despite the advantage of high accuracy, dense sampling requires a high computational cost which is impractical to mobile devices. Therefore, local detector is of preference. The traditional way of using local detectors is taking features with as much invariance as possible. Among the invariant properties of a local feature, rotation invariance is considers an essential property. That is the reason why the orientation normalization step is used in calculating descriptors of popular local features such as SIFT, SURF, and BRISK.

In landmark recognition applications, each landmark appears in users images with minor changes in orientation. Besides, a gyroscope, which is equipped on almost every smart phone, allows estimating the orientation of the image easily. These factors make the rotation invariance of local feature become redundant. Inspired by the result of Zhang [13] which shows that descriptors equipped with different levels of invariance may not always outperform the original ones, combine with two

observed properties of landmark recognition problem mentioned above, we propose the idea of eliminating orientation information in extracting local features. In literature, the idea of eliminating orientation information used to be applied in SURF [23]. However, the motivation of the authors is to optimize speed which is different from our motivation. Georges Baatz [24] also conducts an experiment to compare upright-SIFT, SIFT with zero-orientation, with traditional SIFT and concludes that upright-SIFT give the better performance than SIFT. Nevertheless, this experiment is conducted on a dataset containing landmarks whose orientations are normalized to exactly zero. Clearly, this dataset is far different from reality and also cause no rotation difficulty to upright- SIFT. In our experiments, we use standard datasets whose images are collected from real life to prove that in general, eliminating orientation information gives an enhancement in accuracy with variety of local features (SIFT, SURF, and BRISK).

3 Proposed Method

In order to present and illustrate our proposed idea, we need to put the idea in the context of a specific system so that the efficiency of the proposed idea can be evaluated. Through approaches mentioned in section 2, BoVWs is widely used as the core framework for many state-of-the-art systems. Therefore, we also use BoVWs in our proposal of using local features without orientation normalization to recognize landmarks.

In section 3.1, we briefly describe the BoVWs method. We make a small modification in phase 1 local descriptors extraction to integrate the idea. In section 3.2, we describe the orientation normalization processes of common local features and explain why they cause different level of discriminative information loss. In section 3.3, we describe our system in details.

3.1 Bag of Visual Words (BoVWs)

Basically, BoVWs can be divided into six small phases which are described in specific below.

Phase 1: local descriptors extraction. Local features from all images in the training set are extracted. Depending on the problem requirements, various local features can be used to obtain the local descriptors of each image.

Phase 2: codebook building. Every descriptor extracted from the training set is clustered into k clusters. Then, descriptors in one cluster will be represented by the cluster centroid. A centroid, which can be seen as a visual word, describes the most common features of descriptors inside the cluster that are frequently repeated in the images. The set of these visual words forms a codebook. In most systems, k-means is used in the clustering phase. The higher the value of parameter k is, the more discriminative properties of descriptors are preserved. Phase 1 and phase 2 of BoVWs method are illustrated in Fig. 1.

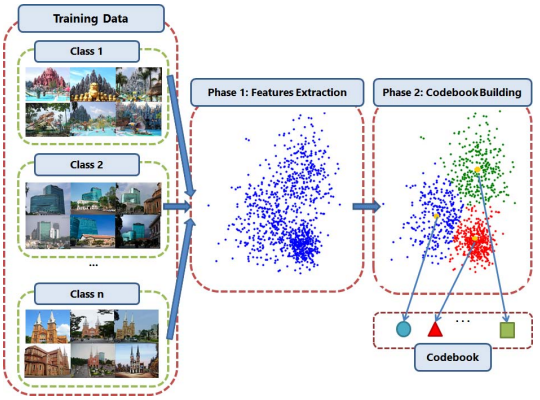

Fig. 1 Features extraction phase and codebook building phase of BoVWs

Phase 3: bag of visual words building. Each local descriptor is characterized by its most similar visual word (nearest in distance). Then, instead of describing an image by all of its local descriptors, the image is presented by a set of visual words. The new representation of an image is conventionally called bag of visual words. This phase is illustrated in Fig. 2.

Phase 4: pooling. We do not stop at representing an image by a list of visual words but keep building a higher-level presentation. A histogram, which has the equal size with the codebook, is taken by counting the number of times each visual word appears in an image. The normalized histogram vectors are then used as the input for the building model phase.

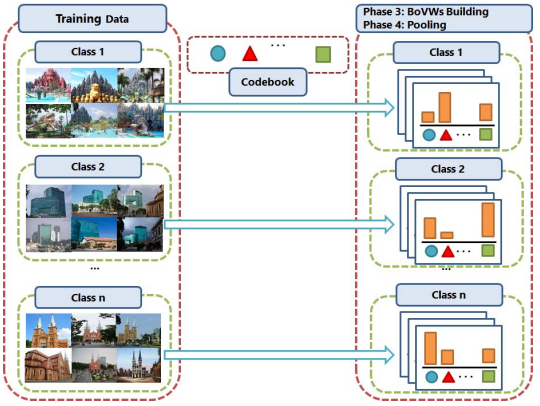

Fig. 2 Bag of Visual Words Building phase of BoVWs

Phase 5: model training. Many classification methods, such as Nave Bayes, Hierarchical Bayesian models like Probabilistic latent semantic analysis (pLSA) and

latent Dirichlet allocation (LDA), or support vector machine (SVM) with different kernel, can be used in this phase to build the model.

Phase 6: prediction. To classify a landmark in an image, that image needs to go through phase 1, 3, and 4. The output, which is a histogram vector, will be predicted the label using the model obtained from phase 5. Fig. 3 illustrates this phase in details.

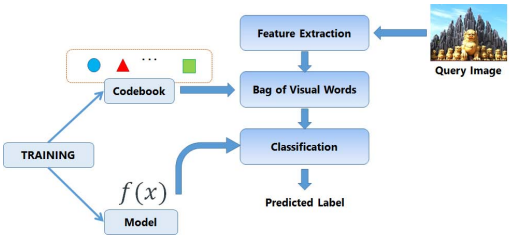

Fig. 3 Predicting a new landmark using BoVWs model

3.2 Orientation Normalization Process

Let us take a further look into the orientation identifying schemes of common local features to understand why it causes the loss of discriminative information. With SIFT, at each pixel in the neighbor region of each key point, the gradient magnitude and orientation are computed using pixel differences. The orientation space is divided into equal bins. Each neighbor pixel votes an amount determined by its gradient magnitude weighted with a Gaussian centered at the key point for the bin to which its gradient orientation belongs. The highest peaks in the histogram along with peaks that are above 80% the height of the highest is chosen to be the orientations of that key point. From Fig. 4, we can see that this orientation identifying scheme is quite sensitive to the change of camera pose and light condition. A small variance of these properties can lead to a significant change of dominant orientations and result in a wrong classification.

SURF makes some modifications in the scheme of determining dominant orientation of SIFT. At each pixel, SURF uses a Haar-wavelet to compute the gradient vector. And instead of choosing multiple orientations for each key point, SURF chooses only one. In comparison to SIFT, SURF provides a less sensitive method to compute local gradient at each pixel. A small change in intensity value of a pixel can be immediately reflected in SIFT local gradient. Whereas, it needs to be a trend of a local region to be reflected in SURF local gradient. Moreover, SURF does not yield different orientation descriptors for a key point which clearly makes it more discriminative than SIFT. Therefore, SURF loses less discriminative information than SIFT. Experiments in section 4 show that there is not much difference in performance between SURF without orientation normalization and the original.

BRISK uses a more complex scheme. In a circle around the key point, which contains n pixels, a local gradient will be obtained at each pair of pixels (from

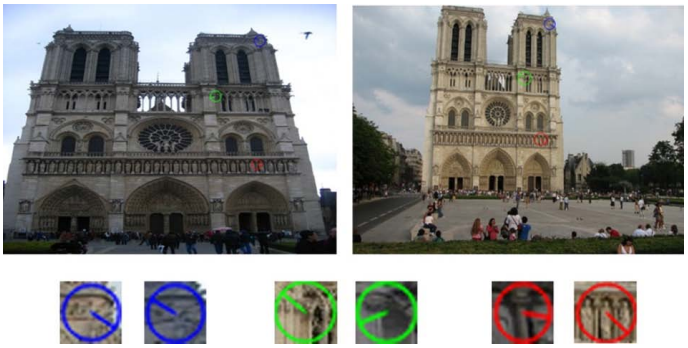

Fig. 4 Dominant orientations detected by SIFT of the same landmark spots at different viewpoints and light conditions

$n(n+1)/2$ pairs) instead of each pixel like SIFT or SURF. The authors smooth the intensity values of two points in a pair by a Gaussian with σ proportional to the distance between the points. The local gradient is computed using smooth intensity differences. Instead of using bins partition and voting scheme, BRISK directly computes the orientation of the key point by taking the average local gradients of the pair of pixels whose distance over a threshold. Therefore, BRISK prefers long-distance pairs than short-distance pairs. It makes BRISK gradient become less local. This leads to a huge loss of discriminative information which can be seen in section 4.

In conclusion, the rotation invariance is not necessary in the problem of landmark recognition. Moreover, rotation invariant descriptor might reduce the classification performance. Besides, different orientation identification schemes cause different levels of loss of discriminative information. In order to confirm our hypothesis, in the experiments section, we test through SIFT, SURF, and BRISK on the Oxford Buildings and the Paris dataset.

3.3 Our Specific System

As mention above, we use the BoVWs model to evaluate the performance of our proposal. In phase 1, we respectively detect key points using detectors of SIFT, SURF, and BRISK. With the set of key points detected by SIFT detector, we describe them by two ways. The first way is using the original SIFT descriptor. The second way is using SIFT descriptor but ignoring the step of rotating the patch to its dominant orientation. We also eliminate key points differed only by their dominant orientation. After this phase, with the set of SIFT key points, we obtain two sets of descriptors: original SIFT descriptors and SIFT without orientation normalization descriptor. In a similar way, with each set of SURF and BRISK key points, we have two sets of descriptors. In phases 2, we use k-means clustering with k-means++ algorithm for choosing initial centroids. We test through different sizes of codebook which range from 512 to 4096 with step 512. We use SVM in phase 5 to train the model.

Multiple binary one-against-one SVM models were obtained to form a multi-class SVM model.

4 Experiments Result

In this session, we report the results of our idea when applied to BoVWs. Experiments are conduct on two standard datasets for landmark recognition problem: Oxford Buildings and Paris. We test through three different local descriptors, which are SIFT, SURF, and BRISK, with varieties of codebook sizes to confirm the efficiency of eliminating orientation information.

4.1 Oxford Buildings

The Oxford Buildings dataset [15] consists of 5062 images from 11 different landmarks. These images range between indoor and outdoor scenes with a great variation in light conditions and viewpoints. Landmarks in some images are difficult to identify because of cluttered backgrounds and occlusions. The similarity of buildings in the dataset even make the classification becomes harder. Some images from the dataset are presented in Fig. 5.

Fig. 5 Some images from the filtered Oxford Buildings dataset

Because images in the Oxford Buildings dataset appear both indoor and outdoor, we filter out the dataset to keep only outside scenes which are suitable for the problem of landmark recognition. All the images in the filtered dataset are resized to be no larger than 640×480 pixels for faster computation. We randomly divide the filtered dataset into 2 subsets: 80% for training purpose and 20% for testing. The codebook and the classification model are built on the training set while samples from test set are used to compute the testing accuracy. We repeatedly run this process 5 times. The average testing accuracy is taking for the final result.

We denote SIFT*, SURF*, BRISK* respectively are SIFT, SURF, BRISK descriptor without orientation information.

Table 1 Accuracy (%) of descriptors with and without orientation information on the Oxford Buildings dataset

Descriptor	Codebook Size							
	512	1024	1536	2048	2560	3072	3584	4096
SIFT	71.80	74.36	75.21	76.92	81.20	76.92	78.63	76.92
SIFT*	**79.83**	**79.83**	**81.54**	**81.37**	**82.05**	**81.37**	**81.37**	**81.20**
SURF	80.34	80.34	81.20	82.91	82.05	82.91	85.47	83.76
SURF*	**81.20**	**82.90**	**83.76**	**85.47**	**88.03**	**87.18**	**88.03**	**87.18**
BRISK	51.45	54.87	55.56	59.83	57.61	57.44	57.44	55.21
BRISK*	**67.01**	**70.43**	**70.43**	**70.09**	**67.69**	**69.74**	**67.18**	**67.86**

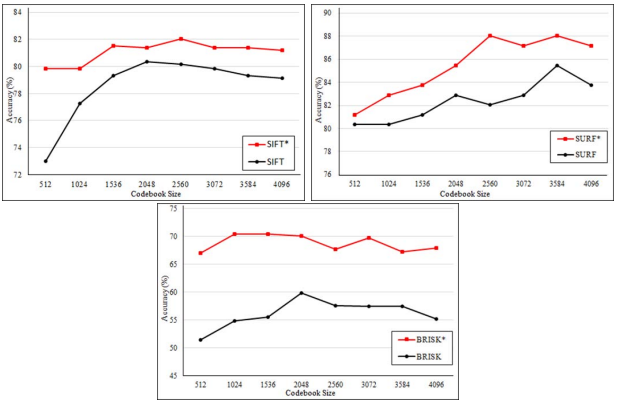

Fig. 6 The experiment result on the Oxford Buildings dataset

The experiment result shows that with or without orientation information, SURF yields the highest precision. On the other hand, BRISK gives the lowest recognition accuracy. Even though landmarks from the dataset do not always appear in the same orientation (Fig. 5), the experiment shows that descriptors without orientation normalization still yield a better performance in comparison to orientation-invariant ones. Table 1 and Fig. 6 show that this elimination gives a remarkable enhancement: on average, it helps boost the accuracy about 4.6% on SIFT, 3.1% on SURF, and 12.6% on BRISK. From the amount of classification performance enhancement of each descriptor, we can conclude that SURF orientation identifying scheme is the one causing least loss of discriminative information. In contrast, BRISK causes a huge loss which means the proposal can give a significant improvement to BRISK. In the best case, it can even raise the accuracy about 15.6%. Another conclusion can be derived from the experiment is that larger codebook size does not always increase the performance while it costs more time for building codebook and training model. In this case, the best codebook size lies between 2048 and 2056.

4.2 Paris

In a similar way to the Oxford buildings dataset, the Paris dataset [16] is collected from Flickr by searching for particular Paris landmarks. The dataset consists of 6412 images from 10 classes. In comparison to the Oxford Buildings dataset, the Paris dataset presents an easier challenge. Landmarks are less obscured. Also, the dissimilarity between classes is quite clear. Some images from the dataset are presented in Fig. 7.

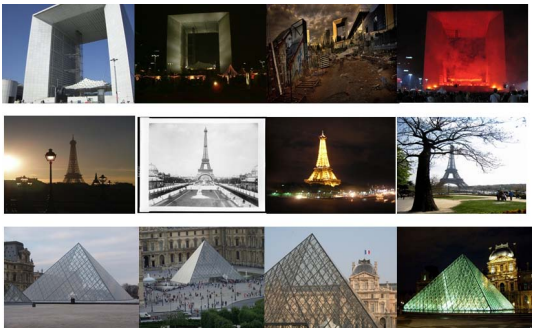

Fig. 7 Some images from the filtered Paris dataset

We conduct a similar experiment to experiment 1. The dataset is filtered and resized and randomly divided into 2 subsets: training set and test set. The classification performances of 3 pairs of descriptors are respectively examined on a variety of codebook sizes.

Table 2 Accuracy (%) of descriptors with and without orientation information on the Paris dataset

	Codebook Size							
Descriptor	512	1024	1536	2048	2560	3072	3584	4096
SIFT	82.19	85.29	85.48	86.84	85.81	85.74	86.32	86.32
SIFT*	**86.45**	**87.36**	**87.81**	**89.23**	**89.03**	**89.03**	**88.90**	**89.68**
SURF	**83.87**	87.74	88.07	87.74	89.68	89.03	**90.32**	**89.68**
SURF*	83.55	**88.07**	**89.03**	**88.71**	**90.00**	**90.00**	89.36	89.03
BRISK	58.97	61.42	61.03	62.77	61.29	61.68	61.55	60.39
BRISK*	**71.03**	**72.71**	**72.65**	**74.58**	**75.10**	**74.97**	**75.42**	**74.84**

Through Table 2 and Fig. 8 we can see that the proposal continue to bring a remarkable enhancement: about 2.9% on SIFT, 0.3% on SURF, and 12.8% on BRISK. It is not surprising that all three descriptors yield better results on the Paris dataset

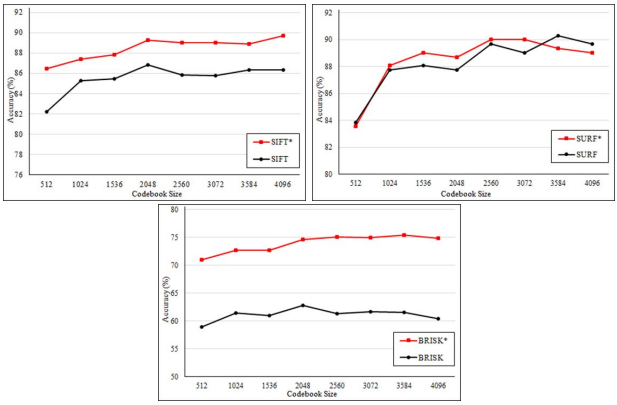

Fig. 8 The experiment result on the Paris dataset

than on the Oxford Buildings dataset. This confirms the hypothesis that the Paris dataset introduces an easier challenge than the Oxford Buildings dataset. Besides, it can be seen form Fig.8 that eliminating orientation information does not always enhance the performance of SURF. However, in most of the cases, it helps boost the accuracy. Moreover, time to extract feature can be save up to 2.5 to 11.1% depend on specific feature.

5 Conclusion

In this paper, we propose the idea of eliminating orientation normalization in calculating visual local descriptors to be applied in landmark recognition problem. The proposal can improve the overall recognition performance of different commonly used local descriptors, such as SIFT, SURF, and BRISK, with a remarkable improvement (12.6% in the best case) while cut down the duration for extracting features. This provides a simple way to boost the efficiency of landmark recognition systems in general.

The results in this paper encourage us to further study and develop landmark recognition systems with more advanced and complex methods, such as LLC or SPM. Besides, we also study to apply results in the field of neuroscience, such as autoencoder, to enhance the accuracy of landmark recognition systems.

Acknowledgment. This research is partially supported by research funding from Advanced Program in Computer Science, University of Science.

References

1. Ahmad, S., Eskicioglu, R., Graham, P.: Design and Implementation of a Sensor Network Based Location Determination Service for use in Home Networks. In: IEEE International Conference on Mobile Ad Hoc and Sensor Systems, pp. 622–626 (2006)
2. Sundaramurthy, M.C., Chayapathy, S.N., Kumar, A., Akopian, D.: Wi-Fi assistance to SUPL-based Assisted-GPS simulators for indoor positioning. In: Consumer Communications and Networking Conference (CCNC), pp. 918–922 (2011)
3. Hsu, C.-H., Yu, C.-H.: An Accelerometer Based Approach for Indoor Localization. In: Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, pp. 223–227 (2009)
4. Jarng, S.: HMM Voice Recognition Algorithm Coding. In: International Conference on Information Science and Applications, pp. 1–7 (2011)
5. Adamek, T., Marimon, D.: Large-scale visual search based on voting in reduced pose space with application to mobile search and video collections. In: IEEE International Conference on Multimedia and Expo, pp. 1–4 (2011)
6. Wilhelm, M.: A generic context aware gesture recognition framework for smart environments. In: International Conference on Pervasive Computing and Communications Workshops, pp. 536–537 (2012)
7. Devendran, V., Thiagarajan, H., Wahi, A.: SVM Based Hybrid Moment Features for Natural Scene Categorization. In: International Conference on Computational Science and Engineering, pp. 356–361 (2009)
8. Dai-Duong, T., Chau-Sang, N., Vinh-Tiep, N., Minh-Triet, T., Anh-Duc, D.: Realtime arbitrary-shaped template matching process. In: 12th International Conference on Control Automation, Robotics and Vision, pp. 1407–1412 (2012)
9. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA (2010)
10. Chen, T., Li, Z., Yap, K.-H., Wu, K., Chau, L.-P.: A multi-scale learning approach for landmark recognition using mobile devices. In: 7th International Conference on Information, Communications and Signal Processing, ICICS 2009, Macau (2009)
11. Bandera, A., Marfil, R., Vzquez-Martn, R.: Incremental Learning of Visual Landmarks for Mobile Robotics. In: 20th International Conference on Pattern Recognition (ICPR), Istanbul (2010)
12. Lee, L.-K., An, S.-Y., Oh, S.-Y.: Efficient visual salient object landmark extraction and recognition. In: International Conference on Systems, Man, and Cybernetics (SMC), Anchorage, AK (2011)
13. Zhang, J., Marszalek, M., Lazabnik, S.: Local features and kernels for classification of texture and object categories: A comprehensive study. In: Conference on Computer Vision and Pattern Recognition Workshop, CVPRW 2006 (2006)
14. Fei-Fei, L., Perona, P.: A Bayesian Hierarchical Model for Learning Natural Scene Categories. In: Computer Vision and Pattern Recognition, pp. 524–531 (2005)
15. The Oxford Buildings Dataset, `http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/`
16. The Paris Dataset, `http://www.robots.ox.ac.uk/~vgg/data/parisbuildings/`
17. Shin, C., Kim, H., Kang, C., Jang, Y., Choi, A., Woo, W.: Unified Context-Aware Augmented Reality Application Framework for User-Driven Tour Guides. In: International Symposium on Ubiquitous Virtual Reality (ISUVR), Gwangju (2010)

18. Chen, D., Tsai, S., Hsu, C.-H., Singh, J. P., Girod, B.: Mobile augmented reality for books on a shelf. International Conference on Multimedia and Expo (ICME), Barcelona (2011)
19. Nistér, D., Stewénius, H.: Linear Time Maximally Stable Extremal Regions. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 183–196. Springer, Heidelberg (2008)
20. Rosten, E., Porter, R., Drummond, T.: Faster and Better: A Machine Learning Approach to Corner Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(1), 105–119 (2010)
21. Chandrasekhar, V., Takacs, G., Chen, D., Tsai, S., Grzeszczuk, R., Girod, B.: CHoG: Compressed histogram of gradients A low bit-rate feature descriptor. In: Conference on Computer Vision and Pattern Recognition, CVPR 2009, Miami, FL (2009)
22. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Computer Vision and Pattern Recognition, CVPR 2006 (2006)
23. Bay, H., Ess, A., Tuytelaars, T., Gool, L.: SURF: Speeded Up Robust Features. In: Computer Vision and Image Understanding (CVIU), pp. 346–359 (2008)
24. Baatz, G., Köser, K., Chen, D., Grzeszczuk, R., Pollefeys, M.: Handling Urban Location Recognition as a 2D Homothetic Problem. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 266–279. Springer, Heidelberg (2010)