International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Review - Translation, Rotation and Scale-Invariant Image Retrieval

Gajendra Paradhi[1], S. B. Nimbekar[2]

[1]Department of Computer Engineering, Pune University Maharashtra, India

[2]Asst. Professor, Department of Computer Engineering, Pune University Maharashtra, India

Abstract: ***While bag-of-features (BOF) models have been widely applied for addressing image retrieval problems, the resulting performance is typically limited due to its disregard of spatial information of local image descriptors (and the associated visual words). In this paper, we present a novel spatial pooling scheme, called extended bag-of-features (EBOF), for solving the above task. Besides improving image representation capability, the incorporation of the EBOF model with a proposed circular-correlation based similarity measure allows us to perform translation, rotation, and scale-invariant image retrieval. We conduct experiments on two benchmark image datasets, and the performance confirms the effectiveness and robustness of our proposed approach.***

Keyword: Image retrieval, bag-of-features

1. Introduction

The amount of online image data is exploding in the past decade due to the rapid growth of Internet users. Since most of such data are not properly tagged when uploading, how to search or retrieve the images of interest is still a very challenging task. This is the reason why content-based image retrieval (CBIR) attracts the attention of researchers in related fields. The use of image descriptors like SIFT [1] is popular in terms of describing the visual appearances of images.

Based on the extracted SIFT descriptors, the use of the bag of- features (BOF) model [2] provides a robust image representation, which is a histogram indicating the numbers of occurrences of each learned visual word. Although the use of BOF models has been shown to be very effective [2, 3, 4], it discards the spatial information of the visual words (or the associated image descriptors) when describing each image. To address this problem, Lazebnik *et al.* [5] proposed a spatial pyramid matching (SPM) and characterized each image by concatenating multiple BOF models at different positions and scales. Recently, Cao *et al.* [6] chose to pool the local image descriptors from each image in a particular spatial order. Instead of explicitly dividing an image into different regions for pooling, the co-occurrence of visual words were also utilized to improve the image retrieval or categorization tasks [7, 8]. In this paper, we present a novel pooling scheme for BOF, named *extended bag-of-features* (EBOF).

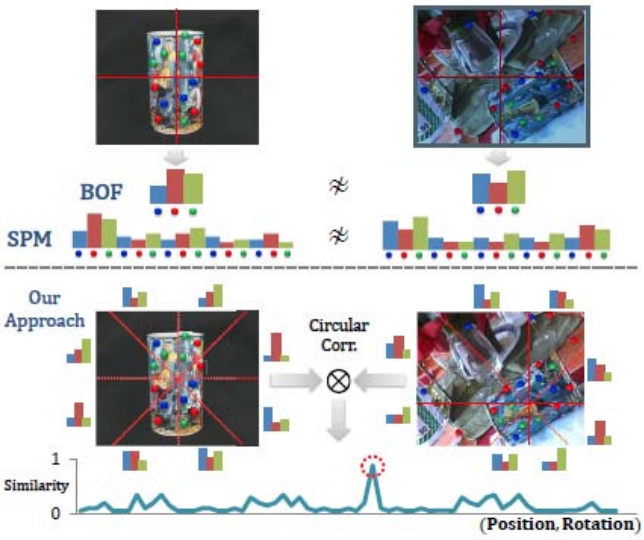

Figure 1: Advantages of our proposed spatial pooling scheme for translation, rotation, and scale-invariant image retrieval.

While the goal of EBOF is to better represent an image by preserving the spatial information of visual words, the integration of EBOF with our proposed circular-correlation based algorithm further allows us to perform translation, rotation, and scale-invariant image retrieval. It is worth noting that, when performing image retrieval, our method does not need to assume self-similarity or to calculate the co-occurrences of visual words explicitly. Later in our experiments, we will verify the effectiveness and robustness of our proposed method.

2. Literature Review

A. Image Classification

The code image classification from the paper[8] we have conclude that Previous BOF models ignore the relation among visual codes. In these algorithms, the codebook can be considered as a special graph which contains only notes

but no edges. This graph can be generated by our method when the domination region angle θ is set to 360 . In this case, our approach is similar to common BOF models. Thus, we can consider that previous BOF models are special cases of o our proposed framework. There are two factors that may affect the performance of our approach. The quality of the codebook graph partly depends on the distribution of visual codes. If the codebook unsuitably covers the local feature space, the generated graph cannot effectively describe images. In addition, to represent each local feature, we only use a simple strategy, i.e., search the nearest domination region to encode the feature. Other strategies are not studied in this paper. We believe that these factors are meaningful researches in future work.

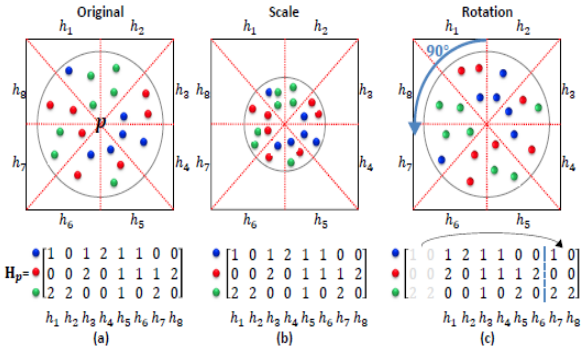

Figure 2: An example of our extended bag-of-features (EBOF) model **H***p*. (a) Original image with EBOF centered at *p*, (b) a scaled version of (a), and (c) a rotated version of (a). Note that each colored point denotes a local image descriptor with a corresponding visual word

Automatic image classification is an important and challenging problem in computer vision. There are many solutions to this problem. Currently, one of the best image classification systems contains two key parts: the bag-of-features (BOF) [7] model and the spatial pyramid matching (SPM) [11] technique. In the BOF model, an image is decomposed into a bag of local visual features which are described by a group of visual codes (codebook). After describing all features, all responses on each code are pooled over to one value by the maximum or the average operation. The image is finally represented by the responses of the codebook. The SPM technique partitions the image into spatial sub-regions, applies BOF on each sub-region and concatenates the histograms from all sub-regions. The BOF model plus the SPM technique achieves impressive performance on many databases, and plenty of extensions have been proposed.

B. Fast spatial Matching

In this paper, we tend to a large-scale object retrieval system. The user supplies a query object by selecting a region of a query image, and the system returns a ranked list of images that contain the same object, retrieved from a large corpus. We demonstrate the scalability and performance of our system on a dataset of over 1 million images crawled from the photo-sharing site, Flickr [13], using Oxford landmarks as queries. Building an image-feature vocabulary is a major time and performance bottleneck, due to the size of our dataset To address this problem we compare different scalable methods for building a vocabulary and introduce a novel quantization method based on randomized trees which we show out performs the current state-of-the-art on an extensive ground-truth. Our experiments show that the quantization has a major effect on retrieval quality. To further improve query performance, we add an efficient spatial verification stage to re-rank the results returned from our bag of words model and show that this consistently improves search quality, though by less of a margin when the visual vocabulary is large.

C. SIFT Descriptor

The purpose of image registration is to spatially align some single modality images taken at different times, or several images acquired by multiple imaging modalities. There are a lot of literatures on medical image registration[11,12], but little concerns preregistration, which is also important to registration. There are lots of computation waste on the coarsely alignment before refined alignment when registering two images difference on scale, orientation and contrast, we put forward the SIFT preregistration method to solve this problem. Medical image registration can be divided into voxel intensity based methods and feature based methods. The voxel intensity based methods use the gray level information to align images, and the registration is achieved with the transformation that maximizes similarity measurements. In feature based methods, the correspondence of presegmented features is first established, and then a certain transformation is defined. The main advantage of feature based methods is of high accuracy and high computational efficiency when accurate correspondence is available. However, it is difficult to establish the correspondence since the segmentation process is hard in most cases and measurement is not perfectly accurate. Although artificial markers provide easy correspondence, it is unwelcome for its invasiveness[3,4]. Our method possesses the advantage of feature based methods and avoids their disadvantages. It is fast to detect and describe the corresponding keypoints, and the invasive.

D. Spatial-Bag-Of-Features

In this paper, we study the problem of large scale image retrieval by developing a new class of bag-of-features to encode geometric information of objects within an image. Beyond existing order less bag-of-features, local features of an image are first projected to different directions or points to generate a series of ordered bag-of-features, based on which different families of spatial bag-of-features are designed to capture the invariance of object translation, rotation, and scaling. Then the most representative features are selected based on a boosting-like method to generate a new bag-of-features-like vector representation of an image.

The proposed retrieval framework works well in image retrieval task owing to the following three properties:
1) The encoding of geometric information of objects for capturing objects' spatial transformation,

2) The supervised feature selection and combination strategy for enhancing the discriminative power, and 3) the representation of bag-of-features for effective image matching and indexing for large scale image retrieval. Extensive experiments on 5000 Oxford building images and 1 million Panoramic images show the effectiveness and efficiency of the proposed features as well as the retrieval framework.

3. A Proposed Solution

A. A Brief Review of BOF, SPM, and SBOF

To represent an image, the bag-of-features (BOF) model [2] quantizes image descriptors such as SIFT [1] into distinct visual words. As a histogram-based representation, each at- tribute of BOF indicate the number of occurrences of each word in an image. While BOF has been applied to image retrieval or classification, it discards the spatial information of visual words and thus limits the representation capability. To address the above problem, spatial pyramid matching (SPM) [5] extends BOF by partitioning an image into several grids at different scales. It pools the BOF models from each grid and concatenates them as a final feature vector. Al- though the spatial order of the visual words is preserved by SPM, it cannot be easily extended to retrieval or classification problems in which the object of interest exhibits translation, rotation, or scale variations in an image.

B. Pooling-Bag-of-Features

Unlike SPM which pools and concatenates BOF models from different grids of an image as an one-dimensional feature vex- tor, we choose to uniformly divide an image into L fan-shaped sub-images (centered at p), as shown in Figure 2(a). For a codebook with K code words, we calculate our pooling bag-of-features (PBOF) model at center p of an image as

$$\mathbf{H}p = [\mathbf{h}\{p,1\}, \mathbf{h}\{p,2\}, ..., \mathbf{h}\{poll\}] \quad \text{.....(1)}$$

where $\mathbf{h} \in \mathbf{R}^{K\times 1}$ is the BOF of the $i$tch sub-image, and $\mathbf{H}p$ is of size $\{K,i\}\times L$. Once this PBOF is constructed, we apply a 2D Gaussian weighting function (centered at p) to suppress the contributions of visual words farther away from p. In our work, we set the standard deviations of both dime- sins of this Gaussian function as half of the longer length of the image. Finally, we normalize this calculated PBOF by $\mathbf{H}p/''\mathbf{H}p''1$ for later correlation and retrieval purposes. Comparing Figures 2(a) and (b), we see that a scale change will not affect the PBOF model, and thus scale in- variance can be achieved. As for rotation variations as shown in Figure 2(c), the resulting PBOF will be a shifted version (in column) of that of the original image. In addition to scale and rotation changes, we also need to deal with translation variations.

In our work, we consider that the object of interest is located at the center of the query image Q when calculating its PBOF $\mathbf{H}^Q$ as the image feature. Thus, the subscript p is ignored in $\mathbf{H}^Q$ for simplicity. For the target images to be retrieved, we uniformly divide each image I into $5 \times 5 = 25$ grids.

C. Image Retrieval with PBOF

1. Circular-correlation based image retrieval

We now discuss how we utilize the proposed PBOF model in addressing the retrieval task. Given a query image Q and a target image I in the database, we need to determine the similarity score between their PBOF models $\mathbf{H}^Q$ and $\mathbf{H}^I$. Re- call that we only construct one PBOF for the query (centered at the query Q), and we have 25 PBOFs for I at different centers. We now determine $\mathbf{S}^{\{Q,I\}} = (\mathbf{H}^Q \otimes \mathbf{H}_p)$ as a K-by-L, correlation matrix, and each row $\mathbf{r}_k$ of $\mathbf{S}^{\{Q,I\}}$ is calculated by

$$\mathbf{r}_k[l] = \mathbf{H}^Q[k,m]\mathbf{H}^I[k, \mathrm{mod}(l+m-1, L)],$$

where $l = 1, 2, ..., L$ denotes the number of rotation angles. From (2), one can see that we perform circular correlation between the $k$ith rows of the PBOF models $\mathbf{H}^Q$ and $\mathbf{H}^I$, and thus the resulting vector $\mathbf{r}_k$ indicates the similarity of the $k$ith visual word between these two images across different rotation angles. Once all rows of $\mathbf{S}^{\{Q,I\}}$ are obtained, we have each column of $\mathbf{S}^{\{}$ as the correlation response (i.e., similarity) between the BOF models between images Q and I. To assess which rotation angle is most likely to be the match between Q and the image I, we apply the cosine Simi- laity as the metric for determining the normalized similarity score between each column of $\mathbf{S}^{\{Q,I\}}$ and the autocorrelation output vector of the query Q. Note that the autocorrelation output vector of Q is calculated as $\mathbf{a} = dig(\mathbf{H}^Q \cdot (\mathbf{H}^Q)^T)$, in which each entry indicates the energy of the BOF model for the corresponding sub-image., this normalized similarity $Sims(\mathbf{H}^Q, \mathbf{H}^I)$ between images Q and I across L different rotation angles is calculated as:

$$Sim(\mathbf{H}^Q, \mathbf{H}^I) = [\mathrm{cost}(\mathbf{a}, \mathbf{s}1), \mathrm{cost}(\mathbf{a}, \mathbf{s}2), ..., \mathrm{cost}(\mathbf{a}, \mathbf{s}L)]. \quad (3)$$

By identifying the largest value in $Sims(\mathbf{H}^Q, \mathbf{H}^I)$, the rotation angle at which Q and I are most similar to each other can be determined. We then repeat the above correlation process for $\mathbf{H}^I$ at different centers p for translations invariance. The maximum output across different $Sims(\mathbf{H}^Q, \mathbf{H}^I)$ is the final similarity score for retrieval,

2. Translation, rotation, and scale invariance

To deal with translation variations when performing image retrieval, we consider that the object of interest is presented at the center of the query image Q without loss of general- it. Thus, only one EBOF model $\mathbf{H}^Q$ is constructed (i.e., the one centered at Q). As for the image I in the database to be retrieved, we uniformly divide I into $5 \times 5 = 25$ grids and consider p as the centers of each grid when extracting the corresponding EBOF models. The EBOF models at 25 dive- ferment locations in I are calculated for representing this imp- age. We perform the above circular-correlation

based price- dare and consider the maximum normalized similarity output across 25 different *Sims*($\mathbf{H}^Q$, $\mathbf{H}^I$) as the final retrieval score. If *p* is located at/near the center of the object of interest in *I*, the corresponding EBOF model at a particular rotation an- gel would produce the highest similarity score. This is how translation-invariant image retrieval is achieved.

To verify the above setting is sufficient for translation-invariant retrieval performance, Figure 4 plots the mean a-enrage precision (MAP) scores of the ETHZ Toys Dataset [9] using different numbers of grids (from 1×1 up to 9×9). From this figure, it can be seen that the use of 5 × 5 = 25 grids is sufficient for producing improved retrieval results (compared to 1 × 1 without shift invariance), and uses of larger numb- beers of grids are not necessary. This because that our retrieval algorithm is based on the *maximum* correlation score. Thus, our choice is preferable for producing satisfactory translation- invariant results.

As discussed earlier in Section 2.2, our proposed PBOF model is robust to *scale* variations when describing an imp- age. Since *rotation* variations would produce shifted PBOF models $\mathbf{H}p$ in columns, we calculate the similarity between the resulting PBOF models for rotation invariance. By idem- tidying the rotation angle of *I* which results in the asocial- acted rotated/shifted version to be most similar to *Q*, rotation- invariant image retrieval can be achieved.. We also observed that *L* from 6 to 10 achieved comparable imp- proved results as those with smaller *L* values. Therefore, our choice of *L* = 8 is sufficient for producing rotation-invariant results

4. Conclusion and Future Work

We proposed an pooling bag-of-features (PBOF) model for image retrieval. Our PBOF is able to exploit the spatial information of visual words presented in images. Together with a circular-correlation based similarity measure, the use of PBOF has been shown to achieve translation, rotation, and scale-invariant and visual image retrieval. Unlike prior retrieval works, the effectiveness and robustness of our proposed method.

5. Acknowledgement

With profound respect and gratitude, I take this opportunity to extend my appreciation to my guide **Prof.S.B.Nimbekar.** Computer Engineering Department, Sinhgad Institute of Technology, Lonavala Engineering, Pune for sharing knowledge and giving me guidance in successfully completing my paper work. I also thank **Prof. T.J.Parvat** (H.O.D of Computer Engg. Dept.) and **Prof. M. S. Chaudhari** (M.E. Coordinator) who provided me the opportunity to present this research paper. Simultaneously; I would like to thanks to IJSRP journals to publish my paper. Also, I would like to thank my parents for their continual encouragement and the positive support. I would also like to thank my wonderful colleagues and friends for concern my ideas, and suggestions for improving my ideas.

References

[1] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” in *Int. J. Computer Vision*, 2004
[2] G. Csurka, C. R. Dance, L. Fan, J. Williamowski, and C. Bray, “Visual categoriztion with bags of keypoints,” in *ECCV Work- shop on Statistical Learning in Computer Vision*, 2004.
[3] J. Yang, T.-G. Jiang, A. G. Hauptmann, and G.-W. Ngo, “Eval- uating bag-of-visual-words representations in scene classifica- tion,” in *ACM SIGMM Int. Workshop on Multimedia Informa- tion Retrieval*, 2007.
[4] D. Li, L. Yang, X.-S. Hua, and H.-J. Zhang, “Large-scale ro- bust visual codebook construction,” in *ACM Multimedia*, 2010.
[5] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of fea- tures: spatial pyramid matching for recognizing natural scene categories,” in *IEEE CVPR*, 2006.
[6] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang., “Spatial- bag-of-features,” in *IEEE CVPR*, 2010.
[7] Y. Zhang, Z. Jia, and T. Chen, “Image retrieval with geometry- preserving visual phrases,” in *IEEE CVPR*, 2011.
[8] C.-F. Chen and Y.-C. F. Wang., “Exploring self-similarity of bag-of-features for image classification,” in *ACM Multimedia*, 2011.
[9] V. Ferrari, T. Tuytelaars, and L. V. Gool, “Simultaneous object recognition and segmentation from single or multiple model views,” *Int. J. Computer Vision*, 2006.
[10] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Ob- ject retrieval with large vocabularies and fast spatial matching,” in *IEEE CVPR*, 2007.
[11] H. M uller, X. W. Gao, and S. Luo, “From medical imaging to medical informatics,” Computer Methods and Programs in Biomedicine, vol. 92, no. 3, pp. 225–226, 2008.
[12] P. Welter, B. Fischer, R. W. Gnther, and T. M. Deserno, “Generic integration of content-based image retrieval in computer-aided diagnosis.” Computer Methods and Programs in Biomedicine, vol. 108, no. 2, pp. 589–599, 2012.
[13] J. C. Caicedo, A. Cruz-Roa, and F. A. Gonzalez, “Histopathology image classification using bag of features and kernel functions,” in Conference on Artificial Intelligence in Medicine, ser. Lecture Notes in Computer Science, vol. 5651, 2009, pp. 126–135.
[14] H. Mller, T. Deselaers, T. M. Deserno, J. Kalpathy-Cramer, E. Kim, and W. Hersh, “Overview of the imageclefmed 2007 medical retrieval and medical annotation tasks.” in CLEF, ser. Lecture Notes in Computer Science, C. Peters, V. Jijkoun, T. Mandl, H. Mller, D. W. Oard, A. Peas, V. Petras, and D. Santos, Eds., vol. 5152. Springer, 2007, pp. 472–491.
[15] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2, ser. CVPR ’06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2169–2178.