Novel local features with hybrid sampling technique for image retrieval

Conference Paper · January 2010

DOI: 10.1145/1871437.1871671 · Source: DBLP

Novel Local Features with Hybrid Sampling Technique for Image Retrieval

Leszek Kaliciak, Dawei Song, Nirmalie Wiratunga
Robert Gordon University
St Andrew Street, AB25 1HG
Aberdeen, Scotland
{l.kaliciak, d.song, n.wiratunga}@rgu.ac.uk

Jeff Pan
Aberdeen University
AB24 3UE
Aberdeen, Scotland
jeff.z.pan@abdn.ac.uk

ABSTRACT

In image retrieval, most existing approaches that incorporate local features produce high dimensional vectors, which lead to a high computational and data storage cost. Moreover, when it comes to the retrieval of generic real-life images, randomly generated patches are often more discriminant than the ones produced by corner/blob detectors. In order to tackle these problems, we propose a novel method incorporating local features with a hybrid sampling (a combination of detector-based and random sampling). We take three large data collections for the evaluation: MIRFlickr, ImageCLEF, and a collection from British National Geological Survey. The overall performance of the proposed approach is better than the performance of global features and comparable with the current state-of-the-art methods in content-based image retrieval. One of the advantages of our method when compared with others is its easy implementation and low computational cost. Another is that hybrid sampling can improve the performance of other methods based on the "bag of visual words" approach.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models

General Terms

Algorithms, Experimentation, Performance

Keywords

Content-based image retrieval and representation, Local features, Global features, Keypoints, Interest points detectors, Sparse sampling, Dense sampling, Random sampling, Local descriptors, Co-occurrence matrix, Colour moments, K-means algorithm, Vector quantization

CIKM'10, October 26–30, 2010, Toronto, Ontario, Canada.

1. INTRODUCTION

Recently, content-based image retrieval (CBIR) based on local features has become popular. Global approaches find it hard to capture the local patterns in content of an image. When it comes to retrieval by texture for example, real-life images often consist of small patches of uniform texture. A popular approach to characterize the local patterns is based on "bag of visual words" or "bag of features"[1].

Most of these methods, however, produce high dimensional vectors. This, in turn, requires the application of dimensionality reduction techniques which leads to higher computational cost. Moreover, most existing approaches (in the retrieval of generic real-life images) are based on interest points detectors, but quite often more discriminant image patches are generated at random (see [7]).

To tackle this problem, we propose a novel and simple to implement method based on local features. Based on the observation that quite often randomly generated sample points are more discriminant than the points detected by corner (blob) detector, we decide to develop a hybrid sampling technique. Points detected by corner (blob) detector tend to concentrate on objects, while randomly generated points help to capture properties of the background. A comparison between random, dense, pure corner-based, and hybrid sampling shows the superiority of our method. The evaluation is performed on three large data collections: ImageCLEF 2007, MIRFlickr 25000 and an image collection from British National Geological Survey (BGS). We also test the influence of the spatial information on the retrieval performance and compare the global features with the local methods.

For the description of the image patches around the sample points, we propose to use: co-occurrence matrix computed at eight different orientations, and three colour moments to capture the local colour properties. Co-occurrence matrix has proven to be an effective way of texture representation. By considering multiple orientations, we make it invariant to rotation. Colour moments are fairly insensitive to changes in viewpoint, and their computation is trivial. In order to make them retain light-independent properties we switch from RGB to HSV colour space. Moreover, patches characterized by colour moments are also able to capture the local textural information.

[1]The terms "bag of visual words" and "bag of features" will be used interchangeably in this paper.

2. RELATED WORK

The "bag of features" (B.O.F.) approach based on SIFT detector and descriptor and its application to nude images detection is described in [1, 2, 3, 4].

Authors of [5] incorporated and tested some methods derived from textual information retrieval domain into CBIR: term weighting, stop word removal, feature selection. They also conducted experiments to test the influence of the vocabulary size (number of clusters) and spatial information on the retrieval performance.

The method proposed in [6] learns the similarity measure and uses a random sampling technique to retrieve previously unseen objects. The algorithm is based on randomized binary trees (a combination of scale invariant feature transform and geometry) and randomly selected rectangular regions of random size.

Nowak and others ([7]) compare different sampling techniques (corner detectors, random sampling, dense sampling). The experimental results showed that the number of sampled patches significantly influenced the retrieval perfor–mance. Surprisingly, for a large number of patches the keypoints detected by corner/blob detector tend to lose their discriminative power. However, there was no comparison with a combination of random and detector-based sampling.

The novel descriptor proposed in [8] uses Gaussian filter banks to characterize local patches. The method seems to outperform SIFT but only on a specific small data collection.

Finally, Athanasakos et al. ([9]) undermine the surprisingly good retrieval performance reported so far by various researchers, claiming that the methods were finely-tuned for collection specific properties. That is why it is so important to test the models on different data collections.

3. THE RETRIEVAL ALGORITHM

Our local features algorithm comprises following stages:

- **Image sampling:** Apply Shi and Tomasi method (see [15]) combined with random points generator to sample an image. This will generate a sparse representation of an image in the form of a set of sub-images (local patches).

- **Description of local patches:** Characterize local patches in the form of co-occurrence matrix or colour moments. In case of co-occurrence matrix, extract the meaningful statistics - energy, entropy, contrast, and homogeneity. Compute the features for individual colour channels.

- **Feature vector construction:** Represent local patterns as 9 dimensional (moments) or 12 dimensional (co-occurrence) vectors.

- **Visual dictionary generation:** Apply clustering (in our work K-means) to the training set in order to obtain the codebook of visual words.

- **Histogram computation:** Create a histogram of visual word counts by calculating manhattan distance between image patches and cluster centroids and generate a vector representation for each image.

- **Similarity measurement:** Measure the distance between multidimensional vectors by using Minkowski's fractional distance measure.

3.1 The Sampling Technique

A sampling technique can have a significant influence on the retrieval performance. The most commonly used sampling methods are based on corner/blob detectors. However, as we shall see, this kind of sampling may not be the best choice for the generic real-life image retrieval. Other methods include dense and random sampling.

As mentioned before, the hybrid sampling method combines Shi and Tomasi corner detection with a random number generator. The Shi and Tomasi method is based on the Harris corner detector. For the detailed description of the algorithm the reader is referred to [15].

We apply the afore-mentioned detector but generate each second sample point at random. This method will take into account the properties of the foreground as well as the properties of the background which are especially important in the retrieval of images depicting natural scenes.

Other sampling techniques implemented for comparison purposes are: corner detector only, purely random and dense sampling. In case of the latter one, each image is divided into the same number of identical rectangular subimages.

3.2 Region Descriptors

We characterize each local patch in an image as an eight orientation co-occurrence matrix and three colour moments in HSV colour space.

The co-occurrence matrix describes the way certain gray–scale pixel intensities occur in relation to other grayscale pixel intensities. It counts the number of such patterns. The most discriminating statistics extracted from co-occurrence matrix are: contrast, inverse difference moment, entropy, energy, homogeneity, and variance.

The method based on three colour moments assumes that the distribution of colour can be treated as probability distribution. Three statistics extracted from individual colour channels are mean, standard deviation and skewness. The first moment can be interpreted as an average colour value, the second as a square root of the variance of the distribution, and the third as the measure of asymmetry in the distribution. Colour moments can also capture the textural properties of an image and are fairly insensitive to viewpoint changes. By computing them in HSV colour space we can make the statistics insensitive to illumination changes.

3.3 Similarity Measurement

The similarity between histograms is calculated based on Minkowski's fractional distance measure. It was experimentally proven (see [10]) that the fractional measures from Minkowski's family of distances yield good results in CBIR.

3.4 Global Features for Comparison Purposes

The global features implemented for comparison purposes are: colour histogram in HSV (Hue, Saturation and Value) colour space, texture features based on co-occurrence matrix, colour moments and methods based on edge detectors which incorporate Canny edge detector and our novel detector based on bilateral filtering (image smoothing that preserves the edges), four directional derivatives and thresholding. The distribution of edges is captured by co-occurrence matrices.

Table 1: ImageCLEF2007 results, 250 sample points

	Hybrid	Random	Dense	Corner
MAP	0.042	0.041	0.037	0.031

Table 2: ImageCLEF2007 results, 900 sample points

	Hybrid	Random	Dense	Corner
MAP	0.050	0.051	0.048	0.040

4. EXPERIMENTS AND DISCUSSION

The data collections used for experimental purposes were ImageCLEFphoto2007, MIRFlickr and BGS (British Geological Survey).

ImageCLEFphoto2007 consists of 20000 everyday real–world photographs. There are 60 query topics consisting of example images that do not belong to the collection. The ground truth data is provided for each topic.

The MIRFlickr collection comprises 25000 images from the Flickr website which are freely redistributable for research purposes. For a detailed description of the collection the reader is referred to [13].

The last data collection consists of 7432 images from BGS. Some of the categories are obviously very difficult for the visual features to capture. This is due to the highly abstract categories with semantic meaning like "economic geology" or "geological hazards".

For each of the two latter collections, 100 query images are randomly selected from the collection, and the ground truth is approximated based on the sharing of same categories between images. We retrieve 1000 images for each query and compute the Mean Average Precision (MAP) based on the ground truth data and general topics.

4.1 Experimental Setup

We experiment with the following sampling techniques: pure corner-based, random, dense and hybrid. Most existing approaches set the number of keypoints to be between 300 and 1400. The larger number of sample points is expected to give more accurate results, but the trade-off is the higher computational cost. In our experiments, we tested two different settings: 250 and 900 respectively. For each sampled point in an image, a 10×10 square patch around it is characterized as a multidimensional vector by applying a local descriptor. The impact of the spatial information is taken into account by dividing an image into 9 sub-images of equal size. Each image patch has 9 and 12 dimensions accordingly, and the codebook size is 40.

4.2 Experimental Results and Discussion

Tables 1 and 2 illustrate the performance of our local features on the ImageCLEF2007 data collection, for 250 and 900 sample points accordingly.

In term of the MAP, for the relatively small number of sample points, the best performing method is the one with hybrid sampling. When the number of sample points increases, however, the discriminative power of randomly sampled patches grows and the set of keypoints detected by corner detector gets saturated. In other words, the corner detector is unable to find more discriminative patches. That is why the best performing method, for 900 sample points, is the one incorporating random sampling. We can also observe that it is desirable to have a large number of sample points which improves the overall performance of local features. Nevertheless, the larger the number of local patches is, the higher the computational and the storage costs will be.

Table 3: MIRFlickr results, 250 sample points

	Hybrid	Random	Dense	Corner
MAP	0.610	0.607	0.609	0.592

We have also tested the influence of the spatial information on the retrieval performance. For a small number of sample points, the inclusion of spatial information improves the retrieval performance. Thus, the MAP for the local features with random sampling and 250 patches increases to MAP$\approx$0.0427. However, when the number of sample points is high, e.g. 900, the spatial information can even hamper the retrieval. For instance, the retrieval performance of local features with random sampling, spatial information included and 900 patches, drops to MAP$\approx$0.045.

The influence of the number of sample points and the inclusion of spatial information on retrieval performance validates the findings reported in [5]. It turns out that when it comes to content based retrieval of generic real-life images, the sophisticated sampling methods are effective only if the number of sample points is relatively low. Otherwise, simple random sampling is the best one due to its ability of finding a higher number of discriminative image patches.

The local features with a different descriptor based on the co-occurrence matrix (with hybrid sampling and 250 sample points) performed worse, obtaining MAP$\approx$0.011.

According to [14], the best performing global method obtained MAP$\approx$0.026. They conducted their experiments on ImageCLEF 2008 collection which consists of the same 20000 images and same 60 topics as ImageCLEF 2007. Thus, in this case, the MAP of the local features is almost twice as high as MAP of the global methods.

The performance of local features with different sampling techniques on MIRFlickr dataset is quite similar (Table 3). It is rather surprising, considering the differences in performance on individual queries. This suggests that the combination of the features with different sampling should improve the results. This will be investigated further in the future.

Figure 1 presents the retrieval results for the global features. The MAP of local features with hybrid sampling is also depicted for comparison purposes. We can observe that the local features performed better on this image collection with our hybrid sampling approach (local hybrid) being the best in term of MAP.

The local features with a different descriptor based on the co-occurrence matrix (with hybrid sampling) performed worse, obtaining MAP$\approx$0.599.

The best results for BGS data collection are obtained by the local features with hybrid sampling as shown in Table 4. The difference between this best performing method and others (approximately 17% of improvement over the most commonly used corner/blob detector-based sampling) is more significant than the difference on the ImageCLEF and MIRFlickr collections.

Figure 2 depicts the MAP for global features and the best performing local method (hybrid sampling). Again, the local features outperformed the global ones.

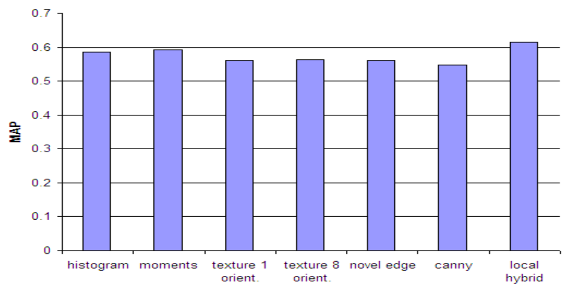

Figure 1: Global versus local features, MIRFlickr.

Table 4: BGS - results, 250 sample points

	Hybrid	**Random**	**Dense**	**Corner**
MAP	0.183	0.171	0.178	0.156

The local features with a different descriptor based on the co-occurrence matrix (with hybrid sampling) performed worse, obtaining MAP≈0.159.

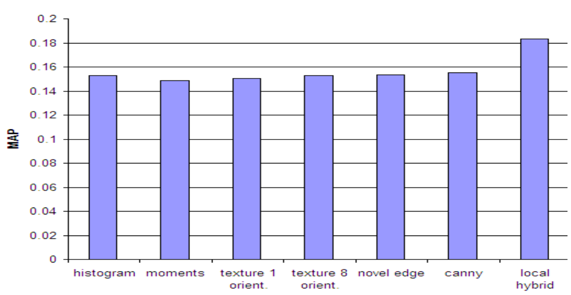

Figure 2: Global versus local features, BGS.

5. CONCLUSIONS AND FUTURE WORK

In this paper we propose a novel method based on the local features, incorporating an easy to implement descriptor and a hybrid sampling technique. We systematically compare different sampling methods: hybrid (a combination of random and detector-based sampling), purely random, purely detector-based and dense on three large data collections.

Our approach is easy to implement, not sophisticated, with low computational and data storage cost (mostly because our vectors are low dimensional), and the hybrid sampling technique can be used in other methods based on the "bag of visual words" to improve the retrieval performance. Moreover, the evaluation of the proposed method has been conducted on three different large data collections without changing the initial setup. In this way we avoid "fine-tuning" of parameters to the specific data collections. This makes the results more general and reliable.

We plan to extend our evaluation of different sampling techniques to various other detectors and descriptors, such as Scale Invariant Feature Transform (SIFT). We will extend the experiments to different numbers of sample points and measure the influence of the codebook size on retrieval performance.

6. ACKNOWLEDGMENTS

This work is funded in part by the Scottish Northern Research Partnership and the UK's Engineering and Physical Sciences Research Council (grant number: EP/F014708/2).

7. REFERENCES

[1] D.G. Lowe. Object recognition from local scale-invariant features. *In Proceedings of the International Conference on Computer Vision*, 2:1150–1157, 1999.

[2] K. Mikolajczyk, C. Schmidt. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1615–1630, 2005.

[3] T. Lindeberg. Scale-space. *Encyclopedia of Computer Science and Engineering*, 4:2495–2504, 2009.

[4] A.P.B. Lopes, S.E.F. De Avila, A.N.A. Peixoto, R.S. Oliveira, A.A. Araujo. A bag-of-features approach based on hue-SIFT descriptor for nude detection. *In Proceedings of the 17th European Signal Processing Conference, Glasgow, Scotland*, 2009.

[5] J. Yang, Y.G. Jiang, A.G. Hauptmann, C.W. Ngo. Evaluating bag-of-visual-words representations in scene classification. *In Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval*, 206, 2007.

[6] E. Nowak, F. Jurie. Learning visual similarity measures for comparing never seen objects. *IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, 2007.

[7] E. Nowak, F. Jurie, B. Triggs. Sampling strategies for bag-of-features image classification. *Lecture Notes in Computer Science*, 3954(490), 2006.

[8] T. Li, I. S. Kweon. A semantic region descriptor for local feature based image categorization. *ICASSP*, 8:1333–1336, 2008.

[9] K. Athanasakos, V. Stathopoulos, J.M. Jose. A framework for evaluating automatic image annotation algorithms. *ECIR*, 5993, 217–228, 2010.

[10] H. Liu, D. Song, S. Ruger, R. Hu, V. Uren. Comparing dissimilarity measures for content-based image retrieval. *AIRS*, 44–51, 2008.

[11] R. Collins. Lecture 06: Harris Corner Detector. http://www.cse.psu.edu/∼rcollins/CSE486/lecture06_6pp.pdf

[12] M. Grubinger, P. Clough, A. Hanbury, H. Muller. Overview of the ImageCLEFphoto 2007 photographic retrieval task. *Advances in Multilingual and Multimodal Information Retrieval*, 433–444, 2008.

[13] M.J. Huiskes, M.S. Lew. The MIR flickr retrieval evaluation. *Proceeding of the 1st ACM International Conference on Multimedia Information Retrieval*, 39–43, 2008.

[14] S. Overell, A. Llorente, H. Liu, R. Hu, A. Rae, J. Zhu, D. Song, S. Rueger. MMIS at ImageCLEF 2008: experiments combining different evidence sources. *Working Notes for the CLEF 2008 Workshop*, 2008.

[15] J. Shi, C. Tomasi. Good features to track. *IEEE Conference on Computer Vision and Pattern Recognition*, 593–600, 1994.