ISSN:2229-6093

Re-Ranking Process for Image Based Retrieval Using Geometric Techniques

Ravi Tej Kumar Maguluri, Podila Manoj, S.Roja Ramani,K.Naga Prakash,K.Ravi Kumar
Department of Electronics & Computer Engineering ,K L University,Vijayawada
maguluriravi6@gmail.com,manoj.podila5@gmail.com

Abstract

We present a fast and efficient geometric re-ranking method that can be incorporated in a feature based image-based retrieval system that utilizes a Vocabulary Tree (VT). We form feature pairs by comparing descriptor classification paths in the VT and calculate geometric similarity score of these pairs. We propose a location geometric similarity scoring method that is invariant to rotation, scale, and translation, and can be easily incorporated in mobile visual search and augmented reality systems. We compare the performance of the location geometric scoring scheme to orientation and scale geometric scoring schemes. We show in our experiments that re-ranking schemes can substantially improve recognition accuracy. We can also reduce the worst case server latency up to 1 sec and still improve the recognition performance. This entails finding the location of a query image in a large dataset containing 3 × 10^4 street side images of a city. We investigate how the traditional invariant feature matching approach falls down as the size of the database grows. In particular we show that by carefully selecting the vocabulary using the most informative features, retrieval performance is significantly improved, allowing us to increase the number of database images by a factor of 10. We also introduce a generalization of the traditional vocabulary tree search algorithm which improves performance by effectively increasing the branching factor of a fixed vocabulary tree.

.

1. Introduction

This paper describes an approach to retrieve images containing specific objects, scenes or buildings. The image content is captured by a set of local features. More precisely, we use so-called invariant regions. These are features with shapes that self-adapt to the viewpoint. The physical parts on the object surface that they carve out are the same in all views, even though the extraction proceeds from a single view only. The surface patterns within the regions are then characterized by a feature vector of moment invariants. Invariance is under affine geometric deformations and scaled color bands with an offset added. This allows regions from different views to be matched efficiently. An indexing technique based on vantage point tree organizes the feature vectors in such a way that a naive sequential search can be avoided. The existence of large-scale image databases of the world opens up the possibility of recognizing one's location by simply taking a photo of the nearest street corner or store-front and finding the most similar image in a database. When this database consists of millions of images of the world, the problem of efficiently searching for a matching image becomes difficult. The standard approach to image matching – to convert each image to a set of scale- and rotation-invariant feature points – runs into storage-space and search-time problems when dealing with tens of millions of feature points[1].High-end mobile phones have developed into capable computational devices equipped with high-quality color displays, high resolution digital cameras, and real-time hardware-accelerated 3D graphics. They can exchange information over broadband data connections, and sense location using GPS. This enables a new class of augmented reality applications which use the phone camera to initiate search queries about objects in visual proximity to the user.

Pointing with a camera provides a natural way of indicating one's interest and browsing information available at a particular location. Once the system recognizes the user's target it can augment the viewfinder with graphics and hyper-links that provide further information.[2]

Our goal, given a query image, is to locate its near- and partial-duplicate images in a large corpus of web images. There are many applications for such a system, for example detecting copyright violations or locating high-quality, or canonical, versions of a low-resolution or altered image. Web image search differs from image-based object retrieval, where image variations can be due to 3D view-point change, lighting, object deformations, or even object-class variability. In our case, target images are obtained by editing the original 2D image through changes in scale, cropping, partial occlusions, etc. This is a less challenging task than full object retrieval, and so the bar is set higher for a system's performance, scalability and accuracy. State-of-the-art large scale image retrieval systems [3, 4, 6, 7] have relied on quantizing local SIFT descriptors [5] into visual words, and then applying scalable textual indexing and retrieval schemes [8]. The discriminative power of local descriptors, however, is limited due both to quantization and to the large number of images (e.g. greater than a million images). Geometric verification [4, 5, 7, 8] becomes an important post-processing step for getting a reasonable retrieval precision, especially for low-resolution images. But full geometric verification is computationally expensive. In practice therefore it is only applied to a subset of the top-ranked candidate images. For web image retrieval the number of near or partial duplicates could be large, and applying full geometric verification to only these top-ranked images may not be sufficient for good recall.

2.GEOMETRIC RE-RANKING IN IMAGE MATCHING SYSTEM:

We consider a mobile visual search system with geometric re-ranking as illustrated in Fig. 1. The mobile client takes a picture of a query object and sends the compressed features to a server where the image recognition takes place. On the server, query features are first quantized using a greedy search through the VT [9]. Then, the histogram of the quantized visual words is used to perform a similarity measure between a query image and a database image. We apply geometric reranking to a subset of the top matching candidates from the VT search. This improves the final list which is passed on to the GV stage, which typically considers a few images only. In the next section, we describe how we generate a matching feature pair list M from the VT search and use the list to generate geometric similarity scores between a query feature set and a database feature set.

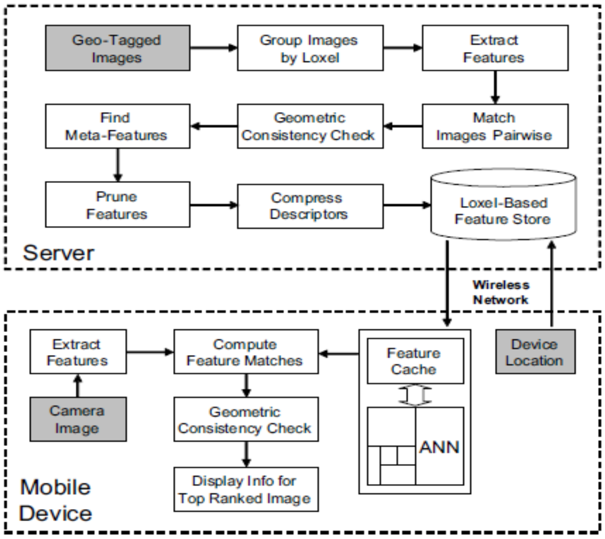


Figure 1: System block diagram. The system is divided into two major components, a mobile device and a server, which communicate over a wireless network.

3. GEOMETRIC SIMILARITY SCORING

We wish to confirm the matching pairs in M using geometry information. In the GV stage, a rigorous validation requires estimating a geometric transformation between the query image and the database image. The estimation of multiple parameters of the geometric transformation renders the process complex and time consuming; thus, we aim to estimate a single parameter instead. The simplest approach uses only orientation and scale information. If we assume a global rotation between the query image and the candidate matching image, then, matching feature pairs should have a consistent orientation difference. Using location information of features for geometric reranking can be advantageous for several reasons. First, for the client server model shown in Fig. 1, we would only need to send the location information of features, which can be compressed efficiently [11]. Second, as GV typically uses only the location information of features for finding a geometric transformation, the location information is already available for geometric similarity scoring. Furthermore, it is compatible with systems that use features that are rotation invariant, such as Rotation Invariant Fast Features[10], which do not yield orientation information. However, using location information is not intuitive when the geometric transformation involves translation, scaling, and rotation.

3.1 Location geometric similarity scoring

We generate a set of log of distance ratios from the list M:

$$S_{LDR} = \left\{ \log \left(\frac{dist(l_{q,i}, l_{q,m})}{dist(l_{d,j}, l_{d,n})} \right) \mid (i,j),(m,n) \in M \right\}$$

where dist(_ ; _) corresponds to the Euclidean distance of two points in the image (Fig. 4 (a)-(c)). For two true matching pairs, the value corresponds to the scale ratio between the query and database image. We then estimate the number of features that have similar scale ratio as follows:

$$C_{LDR}(\alpha) = \sum_{z \in S_{LDR}} I \left(\frac{\alpha}{c} \leq z < \frac{\alpha+1}{c} \right)$$

where I(_) is the indicator function, and _=c corresponds to the scale ratio difference. c is a tolerance factor that is experimentally determined. In practice, for speed and simplicity, we

implement (2) as a histogram with soft bin assignment with _ as the histogram bin index. The geometric similarity score of the two feature sets is then given by:

$$Score_{LDR} = \max_{\alpha} C_{LDR}(\alpha)$$

Using log distance ratio enables us to perform single parameter estimation, estimating the scale ratio between the query and database image. Distances are invariant to rotation, scale, and translation. Distance histograms have been used to match point sets [12]. We extend this idea and use distance ratios, while still preserving robustness against similarity transforms.

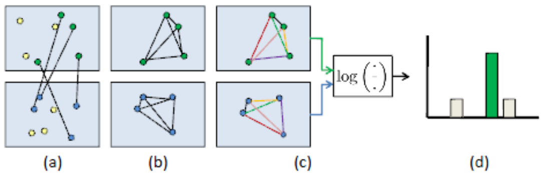


Fig. 2. The process of generating the location geometric score can be shown as the following steps: (a) features of two images are matched according to the descriptor paths, (b) distance of features within image are calculated, (c) log distance ratios of the corresponding pairs (denoted by color) are calculated , and (d) histogram of log distance ratios is formed. The maximum value of the histogram is the geometric similarity score.

3.2 Orientation geometric similarity scoring

Similar to what was described in the previous section, the orientation geometric scoring is formed as follow:

$$\begin{aligned} S_{OD} &= \{(o_{q,i} - o_{d,j}) | (i,j) \in M\}, \\ C_{OD}(\alpha) &= \sum_{z \in S_{OD}} I \left(\frac{2 \cdot \pi \cdot \alpha}{c} \leq z < \frac{2 \cdot \pi \cdot (\alpha+1)}{c} \right) \\ Score_{OD} &= \max_{\alpha} C_{OD}(\alpha) \end{aligned}$$

Intuitively, this orientation difference corresponds to the global rotation angle between the query image and the database image.

3.3 Scale geometric similarity scoring

Scale can also be compared by simply using the feature pairs in M. The scale geometric scoring is formed as follow:

$$\begin{aligned} S_{LSR} &= \left\{ \log \left(\frac{s_{q,i}}{s_{d,j}} \right) | (i,j) \in M \right\} \\ C_{LSR}(\alpha) &= \sum_{z \in S_{LSR}} I \left(\frac{\alpha}{c} \leq z < \frac{\alpha+1}{c} \right) \end{aligned}$$

$$Score_{LSR} = \max_{\alpha} C_{LSR}(\alpha)$$

In this case, the log scale difference indicates the scale difference between the query image and the database image.

4. Weak geometrical consistency

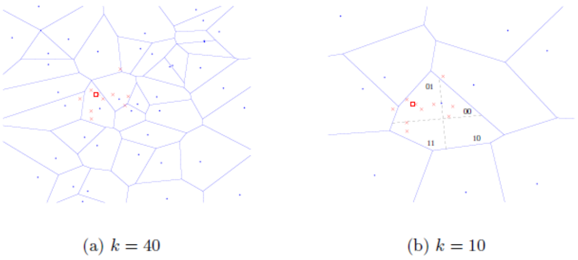
The key idea of our method is to verify the consistency of the angle and scale parameters for the set of matching descriptors of a given image. We build upon and extend the BOF formalism by using several scores sj per image. For a given image j, the entity sj then represents the histogram of the angle and scale differences, obtained from angle and scale parameters of the interest regions of corresponding descriptors. Although these two parameters are not sufficient to map the points from one image to another, they can be used to improve the image ranking produced by the inverted file.

$$s_j(\delta_a, \delta_s) := s_j(\delta_a, \delta_s) + f(x_{i,j}, y_{i'})$$

where _a and _s are the quantized angle and log-scale differences between the interest regions. The image score becomes

$$s_j^* = g\left(\max_{(\delta_a, \delta_s)} s_j(\delta_a, \delta_s)\right)$$

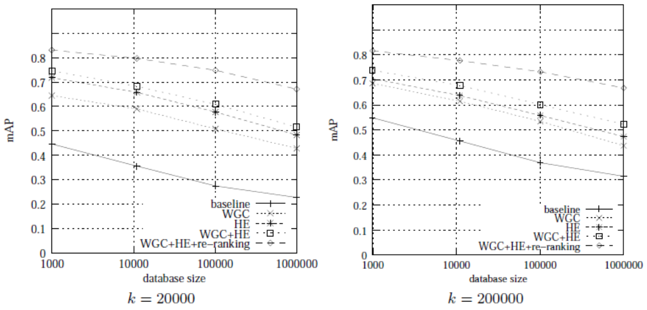
The motivation behind the scores is to use angle and scale information to reduce the scores of the images for which the points are not transformed by consistent angles and scales. Conversely, a set of points consistently transformed will accumulate its votes in the same histogram bin, resulting in a high score.



(a) $k = 40$ (b) $k = 10$

Fig. 3. Illustration of k-means clustering and our binary signature. (a) Fine clustering. (b) Low k and binary signature: the similarity search within a Voronoi cell is based on the Hamming distance. Legend: ·=centroids, _=descriptor, ×=noisy versions of the descriptor.

Re-ranking: The re-ranking is based on the estimation of an affine transformation with our implementation of [20]. Fig. 8 also shows the results obtained with a shortlist of 100 images. We can observe further improvement, which confirms the complementary of this step with WGC.



$k = 20000$ $k = 200000$

Fig. 4. Performance of the image search as a function of the dataset size for BOF, WGC, HE (ht = 22), WGC+HE, and WGC+HE+re-ranking with a full geometrical verification (shortlist of 100 images).

6. CONCLUSION

We have proposed a novel method of compressing location information for mobile image retrieval systems based on feature-based matching. a new method of incorporating geometric similarity re-ranking for mobile image matching systems. Based on the classification feature paths in the VT, a list of matching database and query feature pairs is computed. We use geometric similarity scoring to re-rank candidate matching images given by the tree search. We develop a location geometric scoring that is invariant to similarity transform, compatible with rotational invariant features, and can be conveniently integrated in a mobile visual search system. With this we can expect the fastest means of image searching techniques.

7. References

[1] G. Schindler, M. Brown, and R. Szeliski, "City-scale location recognition," in Conference on Computer Vision and Pattern Recognition, New York, NY, USA,June 2007, pp. 1–7.

[2] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, and B. Girod, "Outdoors augmented reality on mobile phone using loxel-based visual feature organization," in ACM International Conferenceon Multimedia InformationRetrieval, Vancouver, Canada, October 2008.

[3] O. Chum, J. Philbin, M. Isard, and A. Zisserman. Scalable near identical image and shot detection. In Proc. of the Int. Conf. on Image and Video Retrieval, 2007.

[4] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In ECCV, 2008.

[5] D. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 20:91–110, 2003.

[6] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In CVPR'2006.

[7] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In CVPR, 2007.

[8] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In ICCV, Oct. 2003.

[9] G. Schindler, M. Brown, and R. Szeliski, "City-scale location recognition," in Conference on Computer Vision and Pattern Recognition, New York, NY, USA, June 2007, pp. 1–7.

[10] G. Takacs, V. Chandrasekhar, S. S. Tsai, D. M. Chen, R. Vedantham, R. Grzeszczuk, and B. Girod, "Unified real-time tracking and recognition with rotation-invariant fast features," in Conference on Computer Vision and Pattern Recognition, 2010, p. submitted.

[11] S. S. Tsai, D. M. Chen, G. Takacs, V. Chandrasekhar, R. Vedantham, R. Grzeszczuk, and B. Girod, "Location coding for mobile image retrieval," in Proc. 5th International Mobile Multimedia Communications Conference, 2009.

[12] M. Boutin and M. Comer, "Faithful shape representation for 2D gaussian mixtures," in International Conference on Image Processing, 2007, pp. VI: 369–372.

[13] D. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 20:91–110, 2003.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In BMVC, 2002.

[15] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. PAMI, 27(10):1615–1630, 2005.

[16] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J.Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. IJCV, 65:43–72, 2005.

[17] H. Bay, T. Tuytelaars, and L. J. V. Gool. SURF: Speeded up robust features. In European Conference on Computer Vision, pages I: 404– 417, 2006.

[18] S. Arya and D. M. Mount. Algorithms for fast vector quantization. In IEEE Data Compression Conference, pages 381–390, 1993.

[19] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In IEEE Computer Vision and Pattern Recognition, pages II: 2161–2168, 2006.

[20]. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60 (2004) 91–110

[21]. Philbin, J., Chum, O., Isard, M., A., J.S., Zisserman: Object retrieval with large vocabularies and fast spatial matching. In: CVPR. (2007)

[22] Y. Aasheim, M. Lidal, and K. Risvik. Multi-tier architecture for web search engines. In Proc. Web Congress, 2003.

[23] D. Nist´er and H. Stew´enius. Scalable recognition with a vocabulary tree. In Proc. CVPR, 2006.

[24] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In Proc. ICCV, Oct 2003.

[25] K. Mikolajczyk, B. Leibe, and B. Schiele. Multiple object class detection with a generative model. In Proc. CVPR, 2006.

[26] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. Neural Computing, 9(7):1545–1588, 1997.

[27] F. Moosman, B. Triggs, and F. Jurie. Randomized clustering forests for building fast and discriminative visual vocabularies. In NIPS, 2006.