

A NEW RANKING APPROACH AND A REVISITED RATIO TEST FOR IMPROVING CONTENT-BASED IMAGE RETRIEVAL

Massimo Mauro, Riccardo Leonardi

Department of Information Engineering, University of Brescia, Brescia, Italy

ABSTRACT

Geometric Verification (GV) is the last step for most visual search systems. It consists of two parts: first, ratio test is used to find matches between feature descriptors; second, a geometric consistency check is applied. Both steps are computationally expensive, but all the attempts made to speed up the process deal with the geometric check part only. In this work, we focus indeed on ratio test. Using simple PCA and other tricks, a speed-up of an order of magnitude is achieved preserving good retrieval accuracy. Moreover, we propose a modified ranking approach which exploits distance information between descriptors and further improves retrieval performance.

Index Terms— image retrieval, visual search, ratio test, ranking

1. INTRODUCTION

Thanks to the great evolution of mobile phones, the field of visual search is increasing its popularity. The first mobile visual search applications [1, 2, 3] enable the user to recognize objects (buildings, CD, paintings...) and to receive descriptive informations about them. For such applications, a query photo is acquired by a mobile camera and compared with a set of photos stored in a database.

Most current systems for image-based retrieval use a bag of features approach [4, 5, 6], in which images are represented by a set of local features [7, 8, 9] and the recognition is done by matching features between the query image and candidate database images. For fast large-scale systems, features in the database are organized in quantized cells of a vocabulary tree (VT) [10] and associated with an inverted file, i.e. a list of images in which the quantized feature appears. Based on the number of common features between the query and database images, a little number of candidate images are selected from the database. Finally, a Geometric Verification (GV) step is applied to these potentially similar images. In the GV step, features of query and database objects are matched using ratio test [7]. Then, a geometric consistency check on the locations of matching features is normally performed using RANSAC [11]. The GV process is quite slow, motivating our effort to speed it up.

1.1. Prior work

Some work has been conducted to decrease the complexity of GV. In [12, 13, 14, 4] semilocal geometric constraints are used to find or to filter out feature matches. In [15, 16] the authors have tried to speed up RANSAC estimation. In [17, 18] a way for incorporating geometry into the VT is proposed and in [5] only possible sets of hypotheses on the geometric transformation model are verified. In [19, 20] weak geometric consistency checks are used to rerank a larger set of candidate images, before a full GV process is applied on a shorter list of candidates.

1.2. Contributions

Differently from all previous works we focus our attention on ratio test, using SIFT as feature descriptors. The contributions of our work are mainly two:

- we show that a simple exploitation of PCA properties decreases complexity while giving better retrieval performance;
- we propose a modified framework for ranking which takes into account low-distance pairs of descriptors originally discarded from ratio test, proving that they are useful for reaching further improved retrieval results.

This paper is organized as follows. In Section 2 we present our matching algorithm, in Section 3 we introduce our method for ranking candidate images, and in Section 4 experimental results are shown.

2. SIFT MATCHING

Ratio test is the most common scheme used to evaluate matches between two sets of descriptors. Briefly, given a *query image* and a *candidate image*, a distance measure is calculated between their descriptors. If, for a given query descriptor, the ratio between the nearest and the second-nearest candidate descriptors is lower than a pre-fixed value r , then the considered pair is classified as a match.

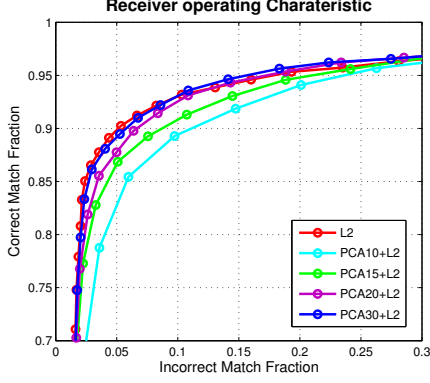


Fig. 1. Comparison of ROC curves with different number of PCA components used.

In this work we use the L_2 -norm as distance measure. In Section 2.1, 2.2 and 2.3 we explain how to modify the ratio test to improve the matching process.

2.1. Dimensionality Reduction

The dimensionality of the SIFT descriptor is 128. We use simple PCA for dimensionality reduction. We evaluated the ROC curve on a 20000-large set of matching/non-matching image patch pairs taken from the *Notre Dame* dataset [21] with a variable number of components used. We found that using 30 PCA components provides the same ROC performance than using the standard SIFT (see Fig.1).

2.2. Pruning

PCA components are ordered according to increasing eigenvalues. We exploit this fact in the calculation of the distance. We define the *cumulative distance vector* between two PCA-SIFTs D_1 and D_2 as

$$cd(D_1, D_2, n) = \sum_{k=1}^n d(D_1, D_2, k) \quad n = 1 \dots N \quad (1)$$

where k represents the index position of the PCA decomposition, $d(D_1, D_2, k)$ is the k^{th} component of the distance vector, and $N = 30$ is the maximum number of PCA components.

Fig.2 (obtained from the same dataset in 2.1) highlights a fact: it is often possible to rely on the very first components of the cumulative distance vector to determine if PCA-SIFT is a positive match or not. For example, the cumulative distance at the 5th component of 70% of non-matching pairs is above the 95%-threshold. In our algorithm, we choose to stop the calculation of the distance between two PCA-SIFTs when the cumulative distance vector exceeds the 95%-threshold.

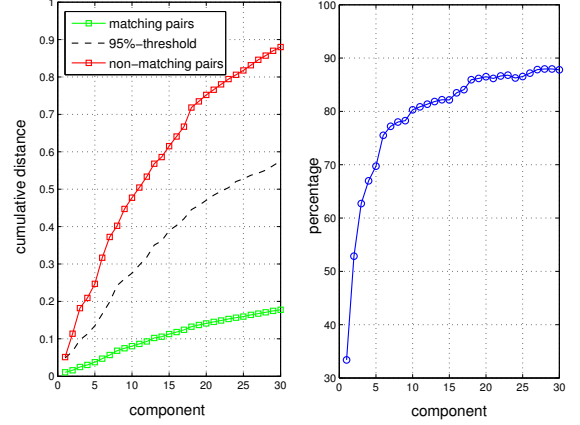


Fig. 2. Distance vector statistics with matching and non-matching pairs. The green and the red lines on the left figure show the mean cumulative distance vector between the PCA-SIFTs in case of - respectively - a matching and a non-matching pair. The black line is a threshold below which there is the 95% of the cumulative distance vectors values in case of a matching pair. The blue line on the right shows, for every component, the percentage of non-matching pairs whose cumulative distance vector values are higher than the 95%-threshold.

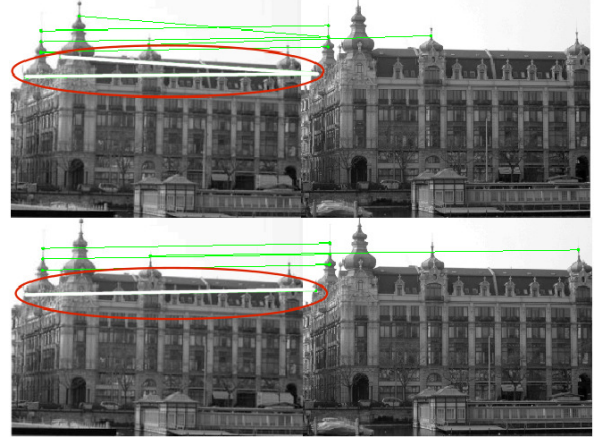


Fig. 3. Example of a multi-to-one correspondence forced to a one-to-one correspondence

2.3. A single descriptor match per region

The original matching algorithm allows multi-to-one correspondences (Fig. 3). However, it is clearly impossible that two regions match with a single region; moreover, such situation would generate matches that are geometrically unreliable, affecting the geometric consistency check performance. To avoid this, we force the algorithm to one-to-one correspondences, choosing - if needed - always the pair of matching descriptors which leads to the lowest distance.

3. IMAGE RANKING

The idea behind the ratio test is [7]:

“...correct matches need to have the closest

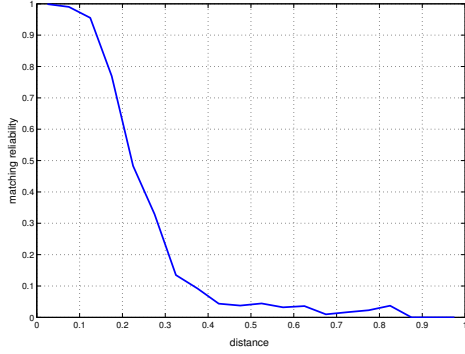


Fig. 4. Estimation of matching reliability on pairs of matching/non-matching PCA30-SIFT descriptors.

neighbor significantly closer than the closest incorrect match to achieve reliable matching. For false matches, there will likely be a number of other false matches within similar distances due to the high dimensionality of the feature space.”

But, what if there are a lot of similar regions in the images? If we consider an object with a great repetitivity (e.g. a building), there could be a lot of very similar pairs of descriptors which are discarded because ratio test classifies them as geometrically unreliable matches.

In Sections 3.1 and 3.2 we present our approach to take into account these low-distance pairs and to use them indeed for the final ranking process.

3.1. Matching reliability

We first define the *matching reliability* of a generic pair of descriptors D_1 and D_2 as

$$rel(\bar{d}) = P\left((\hat{D}_1, \hat{D}_2) \text{ is correct} \mid d(\hat{D}_1, \hat{D}_2) = \bar{d}\right) \quad (2)$$

i.e. the probability that D_1 and D_2 is a correct match, given the distance \bar{d} between them. This probability has been estimated as

$$rel(\bar{d}) \approx \frac{\#matchingpairs < \bar{d}}{\#pairs < \bar{d}} \quad (3)$$

using the PCA30-SIFTs calculated on the same *Notre Dame* dataset in Section 2. In Figure 4 the estimated probability is shown. We can see as the pairs of descriptors with a distance lower than 0.1 have a reliability $rel(\bar{d}) \approx 1$. We decide to consider for our ranking method such low-distance correspondences with $\hat{d} < 0.1$.

3.2. Modified ranking method

In the standard ranking process the matching points which come out from ratio test are passed into RANSAC to define

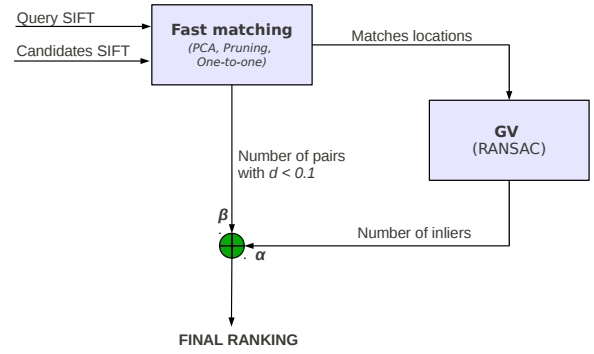


Fig. 5. Overall scheme of the proposed method.

correspondences (*inliers*) using a geometric transformation model estimated on a subset of randomly selected matching pairs. Then, the final image ranking is done according to the number of inliers found between the query image and the possible candidate images.

We propose a modified ranking method, which performs a weighed average between inliers and low-distance descriptors. Given a candidate image I_c its rank is given by:

$$rank(I_c) = \alpha \cdot \#inliers + \beta \cdot (\#pairs : \hat{d} < 0.1) \quad (4)$$

where α and β are learned through cross-validation using a SVM in ranking mode as described in [22]. The overall proposed scheme, including matching and ranking process, is in Figure 5.

4. EXPERIMENTAL RESULTS

We tested our method on 4 datasets:

- *ZuBuD*: Zurich buildings, 115 query images, 1000 reference images;
- *CTurin180*: Turin buildings, 1620 query images, 360 reference images;
- *smvs_CD*: CD covers images from Stanford Mobile Visual Search dataset, 400 query images, 100 reference images;
- *smvs_paintings*: paintings images from Stanford Mobile Visual Search dataset, 400 query images, 100 reference images.

The evaluation process - on a single dataset - is as follows. We pre-calculated PCA30-SIFTs on all the reference images. For each query image, we calculated PCA30-SIFTs and we matched them with all reference images. The value of r for the ratio test is set to 0.7. We used both the standard and our modified algorithms for matching, comparing the execution times. The number of descriptors per image is 500 on

Dataset	Measure	SIFTmatch	fastSIFTmatch	fastSIFTmatch & NewRank
ZuBuD	Top Match (%)	96.52	97.39 (+0.87)	98.26 (+1.74)
	mAP	0.813	0.833 (+0.020)	0.863 (+0.050)
CTurin180	Top Match (%)	87.00	90.00(+3.00)	95.00 (+8.00)
	mAP	0.804	0.884 (+0.080)	0.926 (+0.122)
smvs_CD	Top Match (%)	94.00	94.75 (+0.75)	95.50 (+1.50)
	mAP	0.952	0.958 (+0.006)	0.962 (+0.010)
smvs_paintings	Top Match (%)	80.77	81.87 (+1.10)	82.14 (+1.37)
	mAP	0.821	0.829 (+0.008)	0.829 (+0.009)

Table 1. Experimental results.

average. We took the 25 images with the highest number of matches, and we applied RANSAC on them. Then, the final ranking has been obtained according both to the number of inliers and to the proposed weighted average described earlier. The retrieval performance has been evaluated measuring mean Average Precision and success rate for top match as described in [23].

In Table 1 experimental results are shown. We show both results obtained using the modified matching algorithm only (*fastSIFTmatch*) and the results using also the new ranking method (*fastSIFTmatch&NewRank*), comparing them to the standard *SIFTmatch* algorithm. We can observe that *fastSIFTmatch* obtains better performance than *SIFTmatch*, while being about one order of magnitude faster. Results are even better using the proposed ranking framework, which achieves the best performances with the building datasets. This is not surprising, since in this latter case there is a high probability of repeated patterns which would have been traditionally discarded. We show instead that they turn out useful for improving retrieval accuracy.

5. CONCLUSIONS

In this work we introduce a modified ratio test for matching SIFT descriptors and we propose a new framework for ranking database images. The proposed matching algorithm achieves better performance than the standard ratio test while being ~ 10 time faster. Results are further improved by the use of our ranking approach. We introduced the idea that - especially in presence of repeated patterns - distance information between descriptors can be exploited in addition to the traditional geometric consistency check to refine the accuracy of a content-based image retrieval system.

6. REFERENCES

- [1] Google. (2009). Google Goggles [Online]. Available: <http://www.google.com/mobile/goggles>
- [2] Nokia. (2006). Nokia Point and Find [Online]. Available: <http://www.pointandfind.nokia.com>
- [3] Amazon. (2007). SnapTell [Online]. Available: <http://www.snaptell.com>
- [4] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos", in *International Conference on Computer Vision*, Washington, DC, 2003.
- [5] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases" in *Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 2008.
- [6] S. S. Tsai, D. Chen, J. Singh, and B. Girod, "Rate-efficient, real-time CD cover recognition on a camera-phone", in *ACM International Conference on Multimedia*, Vancouver, Canada, October 2008.
- [7] D. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91110, November 2004.
- [8] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust feature", in *European Conference on Computer Vision*, Graz, Austria, May 2006.
- [9] V. Chandrasekhar, G. Takacs, D. Chen, S. S. Tsai, R. Grzeszczuk, and B. Girod, "CHoG: Compressed Histogram of Gradient", in *Conference on Computer Vision and Pattern Recognition*, 2009.
- [10] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree", in *Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, June 2006.
- [11] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cryptography", in *Communications of ACM*, vol. 24, no. 1, pp. 381395, 1981.
- [12] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval", in *Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 530535, 1997.
- [13] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or how do i organize my holiday snaps?", in *European Conference on Computer Vision*, London, U.K. Springer-Verlag, 2002, pp. 414431.
- [14] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation by image exploration", in *European Conference on Computer Vision*, Prague, Czech Republic, May 2004.
- [15] O. Chum, J. Matas, and J. V. Kittler, "Locally optimized RANSAC", in *DAGM Symposium*, Magdeburg, Germany, Sept. 2003.
- [16] O. Chum, T. Werner, and J. Matas, "Epipolar geometry estimation via ransac benefits from the oriented epipolar constraint", in *International Conference on Pattern Recognition*, Cambridge, U.K., Aug. 2004.
- [17] O. Chum, M. Perdoch, and J. Matas, "Geometric min-hashing: Finding a (thick) needle in a haystack" in *Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 2009.
- [18] Z. Wu, Q. Ke, M. Isard, and J. Sun, "Bundling features for large scale partial-duplicate web image search", in *Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009.
- [19] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search", in *European Conference on Computer Vision*, Berlin, Heidelberg, 2008.
- [20] S. S. Tsai, D. M. Chen, G. Takacs, V. Chandrasekhar, R. Vedantham, R. Grzeszczuk, and B. Girod, "Fast geometric re-ranking for image based retrieval", in *International Conference on Image Processing*, Hong Kong, Sept. 2010.
- [21] S.A. Winder and M. Brown, "Learning Local Image Descriptors", in *Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007.
- [22] T. Joachims, "Training Linear SVMs in Linear Time", in *ACM Conference on Knowledge Discovery and Data Mining*, Philadelphia, USA, August 2006.
- [23] ISO/IEC JTC1/SC29/WG11/N12039, "Evaluation Framework for Compact Descriptors for Visual Search", in *96th MPEG meeting*, Geneva, Switzerland, March 2011.