EFFICIENT COARSE-TO-FINE NEAR-DUPLICATE IMAGE DETECTION IN RIEMANNIAN MANIFOLD

Ligang Zheng[1], *Guoping Qiu*[2] *and Jiwu Huang*[1]
[1] School of Information Science and Technology, Sun Yat-sen University, Guangzhou 510275, China
[2] School of Computer Science, The University of Nottingham, Nottingham NG8 1BB, UK

ABSTRACT

This paper presents an efficient coarse-to-fine strategy for near duplicate image detection in a Riemannian space. At the coarse level, we use the faster but less accurate log-Euclidean Riemannian metric to search the entire database to retrieve a subset of the images that are likely to contain the near duplicates of the querying image; and at the fine level, we use the more accurate but computationally more demanding affine-invariant Riemannian metric to search the coarse level results to accurately identify near-duplicates. We present experimental results to show that the new coarse to fine strategy can be over 20 times faster than existing techniques using affine-invariant Riemannian metric without sacrificing accuracy.

Index Terms— Riemannian metric, manifold, visual saliency, region covariance, near duplicate detection.

1. INTRODUCTION

Due to the advance in digital multimedia processing technology, broadband internet access and increasing popularity of online media sharing, a huge amount of multimedia (especially image, video and photo) has been flooding websites such as YouTube, Facebook, Flickr and many others. Undoubtedly, the easy available multimedia contents have largely entertained the public. However, they have also created problems such as copyright infringements and wasteful usage of storage space and network bandwidth.

A key challenge to the successful detection of a copied video and image lies in the design of effective video and image content descriptors and many schemes have been proposed in the literature [1 - 3]. These descriptors can be classified into global and local descriptors. The global descriptors are generally efficient to compute, compact to store, but less accurate in terms of their retrieval quality. On the other hand, local descriptors [2] are relatively more robust to image transformations, such as occlusion, cropping, etc., but they usually demand more memory and computational resources.

In previous recent work, we have developed a compact and robust descriptor based on visual saliency and region covariance matrices for near duplicate image and video detection [4]. The salient covariance (SCOV) descriptor has been shown to provide state of the art performances and has the advantages of being compact, discriminative and robust. Like other region covariance based descriptors, SCOV's are symmetric, semi-positive defined matrices, which form a manifold in a non-vector space. Therefore, Euclidean metrics are not applicable to the SCOV descriptors and instead the similarities of the descriptors have to be measured using the Riemannian metric. In the literature, there are two major types of Riemannian metric that can be used as a distance measures. One is the affine-invariant Riemannian metric [7, 8] which involves intensive use of matrix inverse and matrix logarithm [8] or generalized eigenvalues [7]. This metric is theoretically elegant but time-consuming. Another Riemannian metric is the log-Euclidean Riemannian metric which first uses matrix logarithm to convert the manifold into vector space and then calculate the Frobenius norm of the matrix as a distance metric [5]. When used for large scale near duplicate image detection, this kind of metric is not as accurate as the affine-invariant Riemannian metric, but it is relatively easy to calculate. An added advantage of the log-Euclidean metric is that well-developed fast search tools such as hashing techniques, e.g., locality sensitive hashing (LSH) [9] can used to enhance search efficiency.

For internet scale near-duplicate image detection, efficiency is very important. In this paper, we first present a comparative study of the two conventional Riemannian metrics for SCOV based near-duplicate image detection which showed that, on the one hand, the affine-invariant Riemannian metric is more accurate but more time consuming, and on the other, log-Euclidean metric is faster but less accurate. Based on this discovery, we have developed a simple but practical coarse-to-fine strategy to enhance the efficiency without compromising accuracy. At the coarse level, we use the faster but less accurate log-Euclidean metric to search the entire database to retrieve a subset of the images that are likely to contain the near duplicates of the querying image; and at the fine level, we use the more accurate but computationally more demanding affine-invariant Riemannian metric to search the returned subset of the coarse level search.

2. SALIENCT COVARIANCE IMAGE DESCRIPTOR

Visual saliency has recently attracted a lot of interests in the computer vision community and various methods have been developed to exploit visual saliency for various tasks such as object recognition [6]. In our previous work [4], we introduced a covariance matrix based descriptor - the salient covariance (SCOV) for near-duplicate image/video detection. The SCOV integrate visual saliency and region covariance which has been shown to have many merits, such

This work was supported by 973 Program (2011CB302204), NSFC (61003243) and the funding of Zhujiang Science & Technology (2011J2200091)

as robust to many kinds of geometric and photometric transformations, compact and discriminative.

Given an image $I \in R^{wxh}$ and let $F \in R^{wxhxd}$ be a d-dimensional feature image. There are many ways to derive the feature image and in this paper we choose F as

$$F(x,y) = \left\{ C_1, C_2, C_3, \frac{\partial I}{\partial x}, \frac{\partial I}{\partial x}, \frac{\partial^2 I}{\partial x \partial y}, \frac{\partial^2 I}{\partial x^2}, \frac{\partial^2 I}{\partial y^2}, \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial x}\right)^2} \right\} \quad (1)$$

where C_1 C_2 and C_3 are the three color channels such as RGB or LMS. The salient features are those $F(x, y)$'s with a corresponding saliency score $S(x, y)$ greater than a threshold

$$SF(x,y) = F(x,y) \;\; if \;\; S(x,y) \geq T \quad (2)$$

The covariance matrix of the salient features is defined as

$$SC = \frac{1}{|\mathcal{S}|-1} \sum_{SF(x,y)\in\mathcal{S}} (SF(x,y) - \mu_{SF})(SF(x,y) - \mu_{SF})^T \quad (3)$$

where, μ_{SF} is the mean of features in the salient region,

$$\mu_{SF} = \frac{1}{|\mathcal{S}|} \sum_{SF(x,y)\in\mathcal{S}} SF(x,y) \quad (4)$$

3. RIEMANNIAN METRICS AND COARSE TO FINE NEAR-DUPLICATE IMAGE SEARCH

3.1 Affine Invariant Riemannian Metric

We denote $S(n)=\{S\in R^{dxd},\ S^T=S\}$, the space of all $n\times n$ symmetric matrices and denote $P(n)=\{P(n)\in S(n),\ P>0\}$ the set of all $n\times n$ symmetric positive defined (SPD) matrices. For $X, Y \in S(n)$, in order to carry out computations with these objects, one need to define distance between these tensors. As the $S(n)$ is part of vector space of square matrices, the easiest way of defining a distance is the Frobenius norm, which is equivalent to considering the $n\times n$ matrices as a $n\times n$ vector in Euclidean space. However, this ruins the intrinsic structure of the manifold. The symmetric, positive semi-defined matrices in $P(n)$ forms a Riemannian manifold. The space is not closed under manipulation with negative scalars. According to [8], an affine-invariant Riemannian metric is given,

$$< y, z >_X = tr\left(X^{-\frac{1}{2}} y X^{-1} z X^{-\frac{1}{2}}\right) \quad (5)$$

Using exponential and logarithm map, the distance between two points X and Y is,

$$\delta(X,Y) = \left\| \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) \right\|_F \quad (6)$$

Furthermore, (6) is equivalent to

$$\delta(X,Y) = \sqrt{\sum_{k=1}^{d} ln^2 \lambda_k(X,Y)} \quad (7)$$

where $\lambda_k(X,Y)$, k =1, 2, …, d are the joint eigenvalues of Σ_i and Σ_j computed as $\lambda_k \Sigma_i x_k - \Sigma_j x_k = 0$ and $x_k \neq 0$ are the generalized eigenvectors [8].

3.2 Log-Euclidean Riemannian Metric

The affine-invariant Riemannian metrics have theoretically excellent properties but lead to complex algorithm. The affine-invariant distance computation involves intensive use of matrix inverse, square roots and logarithms or generalized eigenvalues, thus the computation burden is high which is essentially due to the curvature of the Riemannian space [5].

The log-Euclidean Riemannian metric is another family of metrics which is much simpler to compute. It is defined as follows [5].

$$\delta(X,Y) = \|\mathrm{Log}(X) - \mathrm{Log}(Y)\|_F \quad (8)$$

where Log(X) is the matrix logarithm which converts the manifold into vector space.

As the fast nearest neighbor (NN) searching algorithm in vector space is well-developed, it is relatively easy to find a query's Nearest Neighbors which lies in Riemannian manifold through this matrix logarithm. For example, a hashing (such as LSH [9]) scheme can be used for NN-searching.

3.3 Coarse to Fine Image Search Strategy

It should be noted that the conversion in subsection 3.2 is only approximate rather than exactly transform, because in general there is no such mapping that globally preserves the distance between the points on the manifold. The advantage of (8) is that it can be computed relatively faster than (7), but (7) is more accurate than (8). Based on this observation, we have developed a coarse-to-fine searching strategy to improve efficiency and at the same time maintain the accuracy for near duplicate image detection in the non-vector space of SCOV descriptors. The strategy consists of following steps.

Step1: Convert the Riemannian manifold into vector space using matrix logarithm.

Step2: Use Frobenius norm, which has many fast searching algorithms, e.g. LSH, as similarity measure, and return a relatively large set of potential targets (we found 3000 to be sufficient for our database of 100,000 images)

Step3: Use affine-invariant Riemannian metric to search the set of potential targets returned in Step 2 for near-duplicates of the querying image.

4. EXPERIMENT

We have evaluated the two Riemannian metrics and a coarse-to-fine strategy in near duplicate image detection.

4.1 Datasets

The evaluation is performed on two testing datasets and one distracting dataset:

- ✧ INRIA Copydays dataset [3]
- ✧ SYSU_Test, A set of images randomly chosen from my image dataset.
- ✧ 25,000 Flickr images and another 36,480 images (totally 61480 images) as distracting image dataset.

The INRIA Copydays dataset contains 157 original high resolution images containing a variety of scene types, such as natural, man-made, water, sky, *etc*. Our testing image dataset SYSU_Test contains 977 images randomly choosing from my image collections that contains different image types. 25,000 Flickr images were downloaded from the internet and another 36,480 images are collected from various source.

In original INRIA Copydays dataset [3], there are three main kinds of transformations:

- ➢ resizing plus jpeg compression
- ➢ cropping image surface.
- ➢ strong transformations including print & scan, contrast change, blur, *etc*

The first two kinds of transformations are conducted on each test images, and the authors only produced 229 transformed imaged for the strong transformation.

In this paper, in order to get a comprehensive performance evaluation of our algorithm, besides the attacks mentioned in [3], we extend the transformation types to the two testing image dataset, such as additive noise (salt & pepper, Gaussian), flipping, rotate, blur, illumination change, combination attacks, *etc*. For details of transformations, please refer to Table 1.

Transformation	Transformation parameters
crop	10%,20%,30%,40%,50%,60%,70%,80%
rotate	10,20,30,40,50,60,70,80,90,180 (degree)
flip	horizontal , vertical
salt & pepper	0.05,0.1,0.15,0.2
Gaussian noise	5,8,10,15,20,25 (psnr)
illumination	0.6,0.7,0.8,0.9,1.2,1.3,1.4
jpeg compression	5,8,10,15,20,30,50,75
combination attack	rotate 45,crop 45, resize 0.45, compression quality factor 0.45, median filter 3×3, illumination 1.2, salt & pepper noise 0.09

Table 1: The transformations and transformation parameters

4.2 Evaluation Criterion

ROC. The well-known ROC curve is employed to evaluate the overall performance. The true positive rate (TPR) and false positive rate (FPT) are defined as following.

$$\text{TPR} = \frac{\#\text{ of true recognized near} - \text{duplicates}}{\text{Total }\#\text{ of near} - \text{duplicates}} \times 100\%$$

$$\text{FPR} = \frac{\#\text{ of false recognized near} - \text{duplicates}}{\text{Total }\#\text{ of near} - \text{duplicates}} \times 100\%$$

mFP is the mean of FPR of query images while mTP is the mean of TPR of query images. In this paper, there are a total of 46 transformed copies in the database. These transformations have generated 157×46 = 7222 images for INRIA Copydays dataset and 977×46 = 44942 images for our dataset. All transformed images are then embedded in distracting databases. So the total number of images for Copydays testing experiment is 7222+25000=32222, 7222+61480=68702 and total number of images for SYSU_Test is 44942+61480=106422. The aim is to use the original image as query to retrieve those transformed copies of the image.

mAP. For each query q, there are N copies of q in the database. For a total of Q queries, we measure the average precision as follows

$$AP = \frac{1}{N}\sum_{j=1}^{i}\frac{r_j}{i} \qquad (9)$$

Where r_j is 1 if document j is relevant to the topic.

mAP = Expectation(AP) is the mean average precision. This metric is also sometimes referred to geometrically as the area under the Precision-Recall curve.

4.3 Result Analysis

Fig.1 and Fig.2 shows the ROC performances of affine invariance Riemannian metric, log-Euclidean metric and the coarse to fine strategy. It is seen that log-Euclidean was less accurate than affine invariant. For the coarse to fine strategy, when returning 3000 images in the coarse stage, it almost achieved the same accuracy of the affine invariant, demonstrating the effectiveness of the proposed technique.

Table 2 shows the mAP performances of different metrics and different database sizes. It is again seen that affine invariant is more accurate than log-Euclidean, while it is only necessary for the coarse to fine strategy to return 3K image in the coarse level to match the full search results based on affine invariant.

As the affine-invariant Riemannian metric involves intensive use of matrix inverse, square roots and logarithms, it is very time consuming. While in vector space, the Frobenius norm just involves of simple addition and multiplication, thus it is relatively efficient.

We performed our experiments using Matlab R2009a on a server with Intel Xeon processor (8 cores, 2.13 GHz) and 8 GB memory. Table 3 shows the average time per query for various metrics and strategies for different database sizes. It is seen that both log-Euclidean and the coarse to fine strategy scaled very well with database sizes. The speed of affine invariant metric was 4 time slower as the database size increased from 32K to 106K. As returning 3K will enable the coarse to fine strategy achieve the same performance as full search affine invariant metric, the new strategy was 20 times, 10 times and 5 times faster for

database sizes of 106K, 86K and 32K respectively. The larger the database is, the higher the speedup will be.

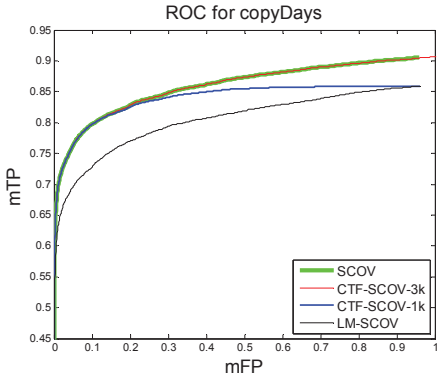

Fig. 1: ROC comparison of affine-invariant Riemannian metric(SCOV), log-Euclidean metric (LM-SCOV) and coarse to fine strategy (CTF-SCOV-3K, CTF-SCOV-1K) where 3K and 1K refer to the number of images returned in the coarse stage.

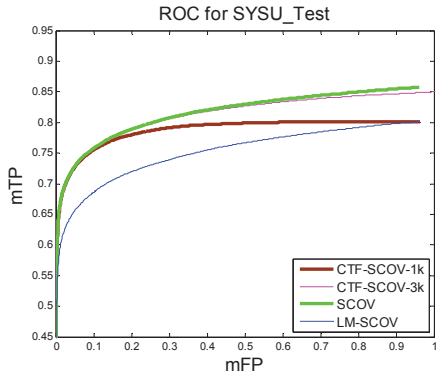

Fig. 2: ROC comparison of affine-invariant Riemannian metric(SCOV), log-Euclidean metric (LM-SCOV) and coarse to fine strategy (CTF-SCOV-3K, CTF-SCOV-1K) where 3K and 1K refer to the number of images returned in the coarse stage.

Database size	32k	68k	106k
Affine-invariant	0.86	0.82	0.77
Log-Euclidean	0.79	0.71	0.65
Coarse to fine(3k)	0.86	0.815	0.75

Table 2 the mAP for different database size and different Riemannian metrics. Coarse to fine (3k) means the mAP is the result when 3k images are returned in coarse searching stage.

5. CONCLUSION

In this paper, we evaluated the two major Riemannian metric for salient covariance (SCOV) based near duplicate image detection. Both Riemannian metrics have weakness, the affine-invariant Riemannian metric is very time demanding while log-Euclidean Riemannian metric has bad detection accuracy. We have presented a simple but practical coarse to fine strategy for salient covariance based near-duplicate image detection, which can greatly reduce the searching time complexity while achieving almost the same results.

Database size	32k	68k	106k
Affine-invariant	8.48s/query	18.4s/query	32.3s/query
Log-Euclidean	0.29 s/query	0.3 s/query	0.36 s/query
Coarse to fine(3k)	1.5s/query	1.56s/query	1.68s/query
Coarse to fine(1k)	0.45 s/query	0.47s/query	0.50 s/query

Table 3 The average time for per query image

6. REFERENCES

[1] A. joly, O. Buisson, and C. Frelicot, "Content based copy retrieval using distortion-based probabilistic similarity search," *IEEE Transactions on multimedia*, pp.293-306, 2007.

[2] J. Law-To, L. Chen, A. Joly, I. laptev, O. Buisson, V. Gouet-Brunet, N. Boujemass, and F. Stentiford, "Video Copy Detection: a comparative study," In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pp.371-378, 2007

[3] M. Douze, H. Jegou, H. Sandhawalia, L. Amsaleg and C. Schmid, "Evaluation of Gist descriptors for web-scale image search," In *Proceeding of the ACM International Conference on Image and Video Retrieval*, pp.19.1-19.8, 2009.

[4] L. G. Zheng, G. P. Qiu, J. W. Huang and H.Fu, "Salient Covariance for Near Duplicate Image and Video Detection," In *Proceedings of International Conference on Image Processing*, pp.2585-2588, 2011.

[5] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Fast and simple calculus on tensors in the log-Euclidean framework," *Medical Image Computing and Computer-Assisted Intervention* (2005) Volume: 8, Issue: Pt1, Publisher: Springer, pp. 115-122

[6] C. Kanan, and G. Cottrell, " Robust classification of objects, faces, and flowers using natural image statistics," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.2472-2479, 2010

[7] W. Förstner, B. Moonen, F. Gdp and C. F. Gauss, "A Metric for Covariance Matrices," *Technical report*, Dept. of Geodesy and Geoinformatics, Stuttgart University (1999)

[8] Xavier Pennec, Pierre Fillard, and Nicholas Ayache, "A Riemannian Framework for Tensor Computing," *International Journal of Computer Vision*, Volume 66 Issue 1, January 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.

[9] Piotr Indyk and Rajeev Motwani, "Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality," In *Proceedings of the thirtieth annual ACM symposium on Theory of computing (1998)*, pp.604-613.