RE-RANKING USING COMPRESSION-BASED DISTANCE MEASURE FOR CONTENT-BASED COMMERCIAL PRODUCT IMAGE RETRIEVAL

Lunshao Chai[*1], *Zhen Qin*[*2], *Honggang Zhang*[1], *Jun Guo*[1], *Christian R. Shelton*[2]

[1]Pattern Recognition and Intelligence Systems Lab, Beijing University of Posts and Telecommunications
[2]Computer Science and Engineering Department, University of California, Riverside

ABSTRACT

With the prevalence of E-Commerce sites such as eBay, Content-based Commercial Product Image Retrieval (CBCPIR) has become an emerging application-oriented field of Content-based Image Retrieval (CBIR). Though a number of traditional CBIR techniques and evaluation criterions have been applied directly or with minor modifications, they tend to neglect one critical factor that greatly affects user experience: users usually care about the exact ranks of the results, especially few top ones, which should share very high similarity with the query image. In this work, we propose a novel two-stage retrieval framework that uses a compression-based re-ranking method and a new subjective retrieval evaluation criterion to address such a problem. More specifically, we extend the state-of-art texture descriptor Campana-Keogh (CK) method from data mining in several aspects and validate the superiority of our framework via extensive experiments and real-world user feedback. We also make our code and CBCPIR dataset publicly available. The number of images of the latter is much larger than current freely accessible ones and better represents real-world commercial product images.

Index Terms— image retrieval, image matching, video compression, image databases

1. INTRODUCTION

CBCPIR has drawn much attention recently as an application-oriented sub-field of CBIR, owning to its practical prospects as more and more people tend to shop online, as well as its lowered technical difficulties (such as easiness of segmentation) from the favorable image properties in general (centered product, simple background, etc.). A number of traditional CBIR theories and methods have been introduced into this new research area directly or with minor modifications [15][6][4]. However, the central role that user experience plays in this area is largely neglected by these traditional methodologies, which results in a research-practice gap. In practice (as shown in Section 5.4 and 5.5), users care little about the number of the resulted top N (say N=20) images

---

*Lunshao Chai and Zhen Qin contributed equally for this paper.

Fig. 1. An example of our re-ranking method. The 1st row is the initial result and the 2nd row is the re-ranked result.

in the same category as the query one (what precision/recall measures), but do care about the exact ranks of the resulted images, especially the first few candidates. One bad result for these candidates may severely degenerate user experience. To the best of our knowledge, this gap has not been discussed in the CBCPIR literature. In this paper, we address this problem for both the retrieval method and the evaluation criterion.

In the retrieval method, we suggest applying a re-ranking stage using a compression-based distance measure for the top ranked results returned from the traditional CBIR process. Such "search within result" framework guarantees fast speed (as for the first stage) and good performance when more sophisticated (but possibly slower) descriptors are used for re-ranking. Almost all existing unsupervised re-ranking algorithms only aim at improving the precision/recall. Some apply clustering in post-processing, which is too time-consuming for real-time systems [2][10], and others need nontrivial offline training [8][11][13]. We show that our augmentation of the CK method [3] borrowed from data mining community works extremely well in this scenario, meeting both user experience and time requirements. We modify the CK method in three ways: First, we extend it to measure color similarity as well as texture similarity. Second, we preprocess the images to achieve shift and scale invariance of the products. Third, we use the MPEG-4 encoder instead of MPEG-1 encoder as the video compressor for better performance.

As for the evaluation criterion, a novel subjective retrieval result measure called Commercial Product Retrieval Rank (CPRR) is proposed. Two features make CPRR a better measure than other alternatives: First, comparing to the precision/recall rate, this measure takes the exact ranks of ranked results into consideration. Second, different from Av-

eraged Normalized Modified Retrieval Rate (ANMRR)[12] and Mean Average Precision (MAP) which also consider ranks of results, CPRR is not limited to situations where the number of images in each category must be known in advance, but preserves the approximately linear coincidence with the results of subjective evaluation about retrieval accuracy.

2. A REVIEW OF THE CK METHOD

As noted in the previous section, for re-ranking we intend to augment the state-of-art Campana-Keogh (CK) method [3], which is a Kolmogorov complexity-based similarity measure for texture. Its idea is to exploit video compression for measuring the similarity of two images, simply by creating a synthetic "video" which is comprised of them. If they are indeed similar, the inter-frame compression utilized in the video compressor should be able to exploit that and produce a smaller file size. The method is shown to be robust and efficient in many applications, such as nematodes classification, breast cancer detection [3] and historical manuscripts recognition [9], where results were stated as "unexpected good" and we find this also true in our focused area. In Fig. 2 we demonstrate the effectiveness of CK for CBCPIR. As we can see, our augmented CK4 distance measure can recognize subtle changes for product images and the clustering result coincides with human perception.

Fig. 2. Some product images are clustered using our augmented CK4 distance measure (details in Section 3) with complete linkage hierarchical clustering.

3. EXTENDED CK METHOD FOR RE-RANKING

3.1. Intuition behind our Method

We choose the CK distance measure for re-ranking over dozens of alternatives for four main reasons. First, it is a *parameter-free* (by viewing the video compression method as a "black box") measure which is more likely to be adopted by customers (researchers as well) who will be unable or unwilling to carefully tune parameters. Second, compression algorithms explore the similarity of details between compared

images and preserve the spatial structure information along the process, which tend to produce small CK distances between images that are indeed similar. Third, as shown by our experiments, our CK4 measure is *very* accurate and fast, the latter of which benefits from the highly optimized MPEG-4 encoding. Fourth, this measure is robust to distortions on images, which meets our need in real-world application scenarios such as users take a query image casually by phones under imperfect environment conditions.

3.2. Preprocessing of the image

We propose a simple yet effective preprocessing step to meet the invariances required by CK to guarantee performance. CK achieves superior performance on images with centered objects of similar sizes [3], which is not the case for real-world images. Two images of the same object at very different image positions would result in a big CK distance given that MPEG only compensates for small motions. Fortunately, this technical difficulty is greatly alleviated in commercial product images because in practice the product is usually centered in the image with a simple background [6]. Thus we first use the method in [4] to extract the main object using location and size information, and then we normalize the product region to a preset size. Experiments show that this normalization strategy is effective. We select 24 images in *Heraldic shields* and *Insect* datasets [3] and use each one as the query image to search in the others. Normalization makes the average inter-set distance bigger and average intra-set distance smaller, and the ranking performance also improves as shown in Fig. 3.
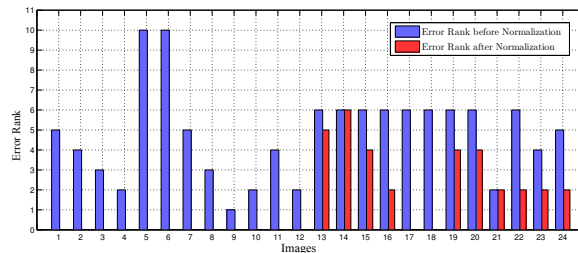

Fig. 3. Error Rank before/after normalization. Each positive result ranked after negative one(s) adds 1 to Error Rank.

3.3. CK4 Distance Measure

Originally the authors of [3] employed an MPEG-1 encoder and named it the CK1 distance measure. We further augment the original CK1 measure to take advantage of the MPEG-4 encoder which has many favorable features and call it the CK4 distance measure following the same convention. The MPEG-4 image and video coding algorithms give an efficient representation of visual objects of arbitrary shape and use Variable Block-size Motion Compensation (VBMC) instead of Block Motion Compensation (BMC). These two features measure similar objects more effectively. Empirically

we observe that CK4 and CK1 have similar re-ranking performance, but the running time of CK4 scales better with the size of compared images.

Also, color features have been shown to be useful in CBCPIR and users do care about color (notice that CK1 so far only handles grey-scale images). Thus, we create color video using MPEG-4 encoding to incorporate color information instead of grey-scale video by CK1 to take color similarity into consideration.

4. CPRR-A NOVEL RETRIEVAL RESULT MEASURE

With the same precision/recall, different rankings of the results may lead to very different user experience. To overcome this drawback, ANMRR [12] and MAP take into account the actual ranks of the retrieval results. But for ANMRR, the number of images in the same category as the query needs to be known, which may not be practical. While MAP doesn't give approximately linear results coincided with the results of subjective evaluation about retrieval accuracy, so users can not judge the retrieval performances directly from MAP values. Our proposed CPRR solves or mitigates both problems.

In CPRR, only the top-K results that may be checked by the user are considered. For the kth retrieved result, we define a specific Rank(k) as

$$Rank(k) = \begin{cases} k & \text{if the kth result is positive,} \\ K+1 & \text{if the kth result is negative,} \end{cases} \quad (1)$$

where "positive" means the result is in the same category as the query image and "negative" the opposite meaning. Then the AVerage Rank (AVR) for query q is defined as

$$AVR(q) = \frac{1}{K} \sum_{k=1}^{K} Rank(k), \quad (2)$$

which we Modify and Normalize to become

$$MNAVR(q) = \frac{AVR(q) - 0.5(1+K)}{0.5(1+K)}, \quad (3)$$

Based on this, our CPRR is defined as

$$CPRR = \frac{1}{K} \sum_{q=1}^{Q} MNAVR(q), \quad (4)$$

which gives one number indicating the retrieval quality over all queries Q. CPRR takes on values between 0 (when all K results are positive) and 1 (when all K results are negative).

5. EXPERIMENTS

5.1. A Sanity Check

We begin with two sets of results in Fig. 4, from which human perception can easily judge the effectiveness of our method. For the initial retrieval stage, we use the histogram-based global color-texture feature CEDD [5]. Then the top-50 resulted images are re-ranked according to their CK4 distances with the query image. Overall, our re-ranking algorithm shows impressive performance and we show detailed results in Section 5.3 and 5.4.
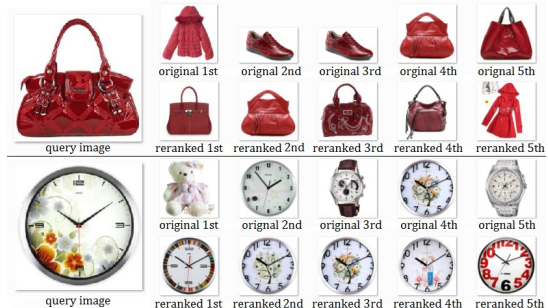

Fig. 4. Two examples of initial retrieval results and their re-ranked version, showing top-5 results for each case.

5.2. Database

At present, there are no widely acknowledged image galleries for CBCPIR. Existing ones such PI100 [15] suffer from low resolution and poor representation of real world commercial product images. Hence we have built a new database *CPImage*. We collected over *4 hundred thousand* images and selected *1 hundred thousand* images in 25 categories from the most popular E-commerce site in China, Taobao (www.taobao.com). We will make this dataset as well as the source code of this work publicly accessible [1] to encourage research in this area.

5.3. Statistical Results

In our experiments, the result of our method is compared with the highly optimized SIFT from VLFeat [14] and SURF in OpenSURF [7] as competitive re-ranking algorithms, *without* the Bag-of-Words framework (which would degenerate the performance). We conduct experiments on CPImage as well as PI100. Experimental results (10 categories for each) are shown in Fig. 5 and Fig. 6.
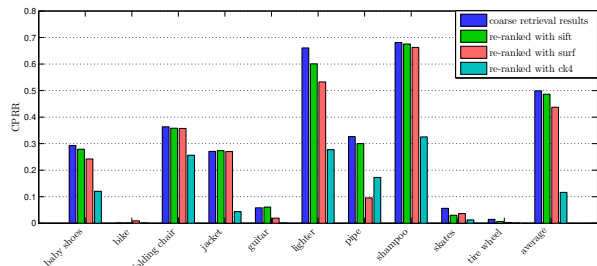

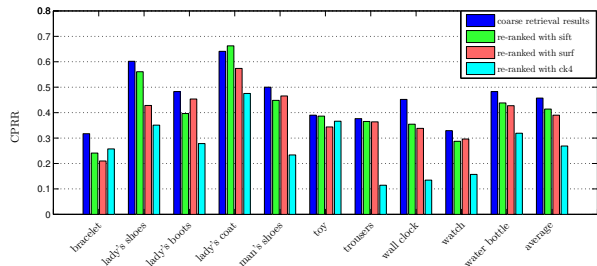**Fig. 5**. Re-ranking performance comparison on PI100 dataset.

Fig. 6. Re-ranking performance comparison on the new CPImage dataset.

Our method outperforms SIFT and SURF in terms of re-ranking performance in most categories. We also observe that CK4 is way faster than SIFT and SURF, especially when the size of image scales. We further observe that the running time of CK4 scales slightly better than CK1, benefiting from the MPEG-4 technique.

5.4. User Study

To further verify our claim that the user experience is also enhanced, we conducted a user survey with 47 volunteers who were asked to score ten sets of randomly chosen results from 0 (totally unsatisfied) to 10 (very satisfied). User study results in Table 1 verify the superiority of our approach in terms of user experience that we focus on.

Table 1. Comparison based on user feedback.

Method	Average Point	% as best
Without Re-ranking	4.85	9.8%
SIFT	5.45	17.8%
SURF	5.68	18.9%
CK4	**6.93**	**53.4%**

5.5. CPRR and User Experience

Space limitation was one reason why we only reported CPRR performance in Section 5.3. More importantly, we compared CPRR, precision and MAP with regard to user scores gathered in Section 5.4. From Fig. 7 we see that though they behave similarly, CPRR coincides better (in terms of linearity) with user experience. Also, the average variance is 0.2634 for CPRR, 0.2940 for MAP, and 0.3312 for precision, meaning CPRR is a more consistent measure. So we feel it is reasonable to show the effectiveness of our method by only reporting CPRR performance in this paper.

6. CONCLUSION

In this work, we augment the state-of-art compression-based CK method from data mining community and embed it into a novel "search within result" framework for CBCPIR. This framework is very accurate and fast, the effectiveness of which is supported by both extensive experiments and real-world user study. Besides, we also propose a novel retrieval
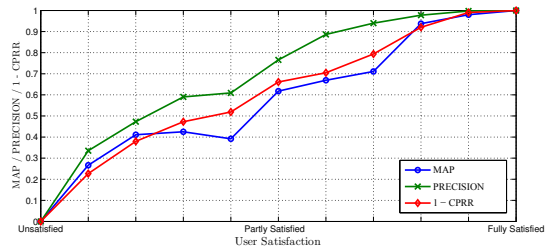

Fig. 7. Average CPRR/MAP/precision over user satisfaction. We report one minus CPRR (thus big values indicate better performance, same as precision and MAP) for clearer comparison. Notice that large precision may relate to plain user satisfaction.

result measure and build a commercial product image gallery for public research.

7. REFERENCES

[1] http://dl.dropbox.com/u/57435211/CPImage_code.rar.

[2] T. Berg and A. Berg. Finding iconic images. In *2nd Internet Vision Workshop of CVPR*, 2009.

[3] B. Campana and E. Keogh. A compression based distance measure for texture. In *SDM*, 2010.

[4] L. Chai, H. Zhang, Z. Qin, J. Yu, and Y. Qi. Multi-feature content-based product image retrieval based on region of main object. In *ICICS*, 2011.

[5] S. Chatzichristofis and Y. Boutalis. CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In *Computer Vision Systems*, 2008.

[6] Z. Chen, L. Duan, C. Wang, T. Huang, and W. Gao. Generating vocabulary for global feature representation towards commerce image retrieval. In *ICIP*, 2011.

[7] C. Evans. Notes on the OpenSURF library. Technical Report CSTR-09-001, University of Bristol, January 2009.

[8] M. Fritz and B. Schiele. Decomposition, discovery and detection of visual categories using topic models. In *CVPR*, 2008.

[9] B. Hu, T. Rakthanmanon, B. Campana, A. Mueen, and E. Keogh. Image mining of historical manuscripts to establish provenance. In *SDM*, 2012.

[10] L. Kennedy and M. Naaman. Generating diverse and representative image search result for landmarks. In *WWW*, 2008.

[11] W. Liu, Y. Jiang, J. Luo, and S. Chang. Noise resistant graph ranking for improved web image search. In *CVPR*, 2011.

[12] B. S. Manjunath, J. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. In *IEEE Trans. CSVT*, 2001.

[13] D. Pedronette and R. S. Torres. Exploiting contextual information for rank aggregation. In *ICIP*, 2011.

[14] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. *Design* 3(1), 1-4, 2010.

[15] X. Xie, L. Lu, M. Jia, H. Li, F. Seide, and W. Ma. Mobile search with multimodal queries. In *Proc. IEEE*, 2008.