EVALUATING KEYPOINT METHODS FOR CONTENT-BASED COPYRIGHT PROTECTION OF DIGITAL IMAGES

Larry Huston[1] *Rahul Sukthankar*[1,2] *Yan Ke*[2]

[1]Intel Research Pittsburgh, [2]Carnegie Mellon University

larry.huston@intel.com rahuls@cs.cmu.edu yke@cmu.edu

ABSTRACT

This paper evaluates the effectiveness of keypoint methods for content-based protection of digital images. These methods identify a set of "distinctive" regions (termed keypoints) in an image and encode them using descriptors that are robust to expected image transformations. To determine whether a particular image were derived from a protected image, the keypoints for both images are generated and their descriptors matched. We describe a comprehensive set of experiments to examine how keypoint methods cope with three real-world challenges: (1) loss of keypoints due to cropping; (2) matching failures caused by approximate nearest-neighbor indexing schemes; (3) degraded descriptors due to significant image distortions. While keypoint methods perform very well in general, this paper identifies cases where the accuracy of such methods degrades.

1. INTRODUCTION

The increasing availability of digital multimedia content creation and storage has led to an explosive growth in large collections of digital imagery. The ease with which digital images can be copied (without licensing) motivates content providers to seek schemes to detect unauthorized distribution and use of this data. A common approach to the problem is digital watermarking [3], where a small amount of data is imperceptibly embedded into each image. Such a watermark, when extracted from an image, should unambiguously identify the image's source. Unfortunately, typical watermarking schemes can be defeated in several ways; some watermarks can be detected and removed by knowledgeable users, other watermarks may be destroyed if the image is slightly cropped or compressed using a lossy scheme. This has stimulated research in a complementary approach to copyright protection: identification of potential violations on the basis of similarity in image content.

There has been significant recent research on techniques relevant to content-based copyright protection of image databases, particularly in the context of near-duplicate detection and sub-image retrieval [2, 7, 9, 11, 12]. These systems typically employ a traditional content-based image retrieval (CBIR) framework, which can suffer from the following two problems. First, many techniques calculate and store global statistics for each image, which is efficient but insufficiently accurate: recall can suffer when a significant transformation perturbs global statistics; and precision can be poor because global statistics, such as histograms (that are robust to geometric transforms), tend to generate many false positives. Second, those systems that compute local statistics of an image (e.g., by partitioning an image into smaller pieces) can suffer from low precision [9, 11]. This is because such systems typically concatenate all of an image's local statistics into a single feature vector describing the image. The matching threshold must be loose in order to allow vectors derived from a portion of an image to match those generated from the entire image. This design consideration necessitates a considerable loss in the filter's specificity. Approaches that employ distinctive local features of the image, termed *keypoints* [8], are believed to be robust to these problems because: (1) they match images using only a small subset of features; and (2) the keypoints are resistant to image distortions [**?**]. However, naive implementations of keypoint methods scale very poorly with image collection size because they require a prohibitive number of nearest-neighbor searches. This has led to systems [1, 7] where fast approximate nearest-neighbor algorithms, such as locality-sensitive hashing [5], are employed to enable efficient search of large image databases. This paper evaluates the performance of keypoint methods for content-based protection of large image databases under extreme conditions.
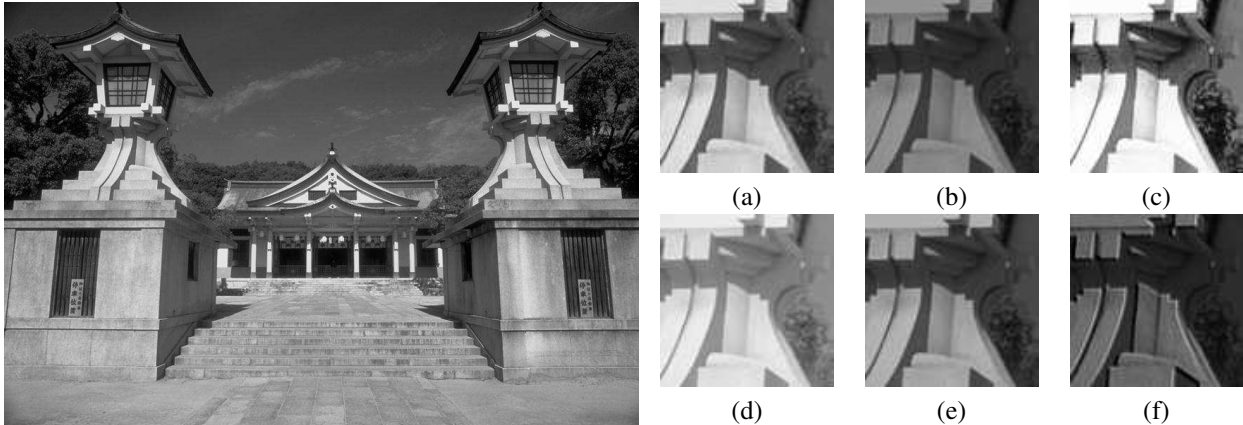
Figure 1: Six examples of queries derived from an original: (a) 100x100 crop; (b) colorized (R=95/G=10/B=0); (c) histogram equalized; (d) gamma corrected (1.7/2.3/0.6); (e) modulated (B=80/S=119/H=100); (f) simulated shade (light at azim=$0°$/elev=$45°$).

2. BACKGROUND

Keypoint methods have recently demonstrated success across a variety of applications including panoramic image stitching, robot localization, object recognition and image retrieval. The basic idea can be summarized as follows: (1) identify distinctive regions in each image using an interest operator in scale-space; (2) represent the image patch surrounding each keypoint using a local descriptor that is robust to expected transformations; (3) store each keypoint descriptor in a manner that enables efficient matching.

We employ the SIFT interest operator [8], which identifies a few thousand keypoints in each image, requiring our system to store almost a billion keypoints for a realistic image collection. We use the PCA-SIFT local descriptor [6] to generate a 36-dimensional vector for the local patch surrounding each keypoint. Experiments have demonstrated that keypoints are reasonably robust to a variety of common image transformations, such as rotation, scaling, affine warps, compression and some lighting changes [?]. To identify matching images, we find keypoints that appear in both images and perform geometric verification to ensure that the corresponding keypoints appear in the same relative locations. The use of highly-distinctive local descriptors combined with geometric verification allows us to treat a small number of consistent keypoint matches as a positive image match, thus making our system highly robust to cropping, occlusion and certain forms of distortion.

Our system requires matching query keypoints with a large database of potential keypoints in a high-dimensional space. Since traditional indexing methods are infeasible in this domain, we employ locality-sensitive hashing (LSH) for efficient approximate near-neighbor queries [4, 5]. LSH uses multiple hash tables, each keyed by a different random subset of the high-dimensional vector. To perform a near-neighbor query, LSH performs probe in each of the hash tables and returns the union of these results. This search is very efficient but only returns a subset of the near-neighbors for a given query.

This paper builds on previous work [7] that employs a similar scheme for performing sub-image retrieval. Given a collection of copyrighted images, we build an LSH database of the keypoints in the original images. To query the database, we extract keypoints from the query image, search the LSH database and perform geometric verification using the resulting keypoints.

The goal of this paper is to evaluate the applicability of our system for detecting potential copyright violations. Specifically, we consider three problems that reduce the keypoint matching accuracy: (1) the loss of keypoints due to small query images (either from extreme scaling or significant cropping); (2) the failure of approximate near-neighbor searches (LSH) to find all of the keypoints within the desired distance of the query; (3) the distortion of query keypoints due to image modification, as shown in Figure 1.

3. EVALUATION

This section evaluates the effectiveness of keypoint methods under different operating scenarios to iden-

Size	Precision	Recall
Scaled		
max 200	1003/1292 (77.6%)	1003/1014 (98.9%)
max 100	960/1039 (92.4%)	960/1014 (94.7%)
max 50	570/578 (98.6%)	570/1014 (56.2%)
Cropped		
200×200	927/1152 (80.5%)	927/1014 (91.4%)
100×100	724/762 (95.0%)	724/1014 (71.4%)
50×50	306/308 (99.4%)	306/1014 (30.2%)

Table 1: Results on scaled and cropped query images.

tify conditions where they can fail. These experiments use the 30,273 photographs from a commercial collection [10] as the set of data to be protected. The images span a large variety of subjects including texture, animals, buildings, people and natural scenery. We extract the keypoints from these images (approximately 758 million) and place them into LSH hash tables. These experiments use l=20 hash-tables and k=450 bits for each LSH key. Our system rejects image matches that contain fewer than four geometrically-consistent keypoint correspondences.

We randomly select 1,014 images from the collection to serve as the source images for copyright violations. We perform several experiments to examine the system's failure characteristics for each of the problems described above. Each experiment modifies the source images to generate a set of queries as described below.

3.1. Sub-Image Queries

These experiments examine the case where the number of keypoints in the query images are reduced through extreme scaling or cropping. Scaled images are generated at various sizes (by fixing the maximum dimension of the thumbnail) and cropped images are created by choosing an arbitrary location in the original image and extracting square sub-images of various dimensions. In contrast to earlier work [7], this paper focuses exclusively on small queries as they prove to be the most challenging. Table 1 presents the results of these experiments.

We observe that, for larger query images we get good recall, but as the query image decreases in size (the number of keypoints decreases), the recall drops. On the other hand, the number of false positives decreases with the size of the query image, boosting precision.

To further examine the effects of LSH and query image size behind these results, we show a cumulative dis-
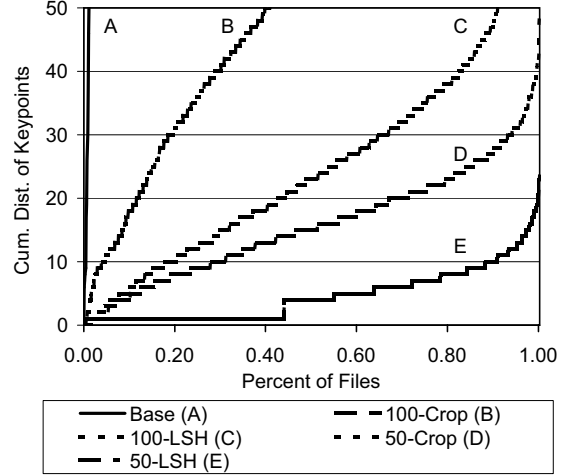

Figure 2: Keypoint distribution for scaled queries

tribution of the keypoints for a selection of these queries (see Figure 2). Each line in the graph shows the percentage of images that contain at least the given number of keypoints. The first line (A) shows that the majority of unmodified images contain well over 50 keypoints. The remaining lines show the relation between the number of keypoints in the scaled query image (lines B & D) and the number of points found using LSH (lines C & E). We see that scaling greatly reduces the number of keypoints available compared to the baseline (A), with the smaller image having significantly fewer available keypoints. We see that LSH further drops a significant fraction of the keypoints; in particular, a comparison between lines (D) and (E) shows that LSH cannot find the minimum keypoints (4) required for an image match for over 40% of the images, even though most of these keypoints could be found through exhaustive search.

3.2. Distorted Queries

These experiments examine the effects of an added complication: modifying the image content in addition to its size through cropping. The robustness of keypoints to several image transformations has been demonstrated in other work [?,7]. Here, we focus on transforms that perform non-uniform modifications to pixel values to stress the keypoint representation. We modify each original query image (using ImageMagick) and crop it to the desired size. Examples of distorted queries are shown in Figure 1.

Table 2 summarizes these results. Recall drops dramatically on the smaller query images and the distor-

Crop Size	Distortion	Precision	Recall
200×200	Colorize	81.4%	86.9%
200×200	Equalize	80.2%	65.6%
200×200	Gamma	78.9%	87.4%
200×200	Modulate	79.6%	90.0%
200×200	Shade	95.1%	47.4%
100×100	Colorize	94.1%	60.0%
100×100	Equalize	95.5%	35.3%
100×100	Gamma	94.6%	58.7%
100×100	Modulate	95.0%	65.9%
100×100	Shade	91.8%	4.4%
50×50	Colorize	98.5%	18.8%
50×50	Equalize	100.0%	5.6%
50×50	Gamma	99.3%	13.4%
50×50	Modulate	99.2%	23.3%
50×50	Shade	0.0%	0.0%

Table 2: Precision and recall for distorted query images.

Figure 3: Keypoint distribution for distorted queries.

tions further reduce the recall (compare to Table 1). Equalize and shade have the most significant impact. To derive greater insight into these results, Figure 3 depicts the cumulative distribution of keypoints for a selected set of distortions. Line (A) is the baseline; line (B) shows the distribution after cropping alone; Lines (C) and (E) show the impact of the gamma and shade distortions, respectively. Even with exhaustive search, it is clear that shade makes it very difficult to find any matching keypoints. LSH exacerbates the problem by failing to find many of the existing keypoints, as shown by Lines (D) and (F), As before, exhaustive search would significantly improve the accuracy of the system under these extreme conditions, at the cost of running time.

4. CONCLUSIONS

For typical content-protection scenarios (with reasonably-sized query images), our experiments confirm that the combination of keypoints and LSH gives excellent results. When the experiments explore the extreme conditions (small query images and challenging image modifications), we find that these techniques do not perform quite as well. However, we know of no existing techniques that would perform better on these scenarios. The use of aggressive LSH parameters exacerbates the results by making it difficult to match the few existing keypoints. Further exploration of LSH parameters and other near-neighbor techniques could improve accuracy.
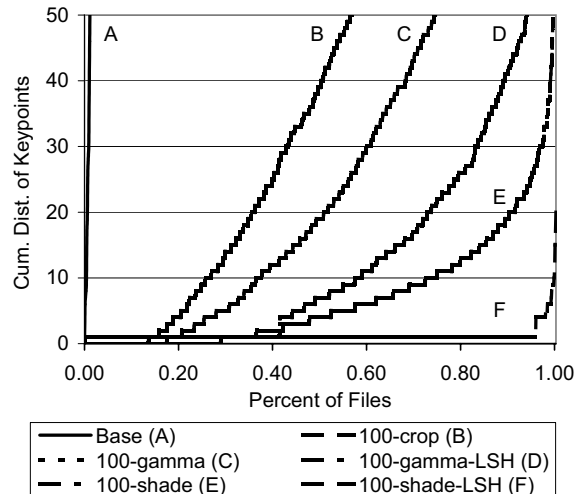
5. REFERENCES

[1] S. Berrani, L. Amsaleg, and P. Gros. Robust content-based image searches for copyright protection. In *Proc. ACM Workshop on Multimedia Databases*, 2003.

[2] E. Chang, J. Wang, C. Li, and G. Wiederhold. RIME: A replicated image detector for the world-wide web. In *Proc. SPIE*, 1998.

[3] I. Cox, J. Bloom, and M. Miller. *Digital Watermarking: Principles and Practice*. Morgan Kaufmann, 2001.

[4] A. Gionis, P. Indyk, and R. Motwani. Similarity search in high dimensions via hashing. In *Proc. Intl. Conf. on Very Large Databases*, 1999.

[5] P. Indyk and R. Motwani. Approximate nearest neighbor—towards removing the curse of dimensionality. In *Proc. Symposium on Theory of Computing*, 1998.

[6] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Proc. IEEE Computer Vision and Pattern Recognition*, 2004.

[7] Y. Ke, R. Sukthankar, and L. Huston. Efficient near-duplicate and sub-image retrieval. In *Proc. ACM Conf. on Multimedia*, 2004.

[8] D. Lowe. Distinctive image features from scale-invariant keypoints. *Intl. Journal of Computer Vision*, 2004.

[9] J. Luo and M. Nascimento. Content based sub-image retrieval via hierarchical tree matching. In *Proc. ACM Workshop on Multimedia Databases*, 2003.

[10] Media Graphics International. 270,000 Multimedia Graphics Pack, 1998.

[11] N. Sebe, M. Lew, and D. Huijsmans. Multi-scale sub-image search. In *Proc. ACM Conf. on Multimedia*, 1999.

[12] D. Zou, C. Wu, G. Xuan, and Y.-Q. Shi. A content-based image authentication system with lossless data hiding. In *Proc. Intl. Conf. on Multimedia and Expo*, 2003.