Detection of Object-centered Near-duplicates

Jan Erik Solem
Polar Rose AB
Anckargripsgatan 3
Malmö, Sweden
janerik@polarrose.com

ABSTRACT

This paper describes the use of object specific descriptors to search large image sets for near-duplicates and derivative works. We will focus on using descriptors for faces but the general procedure is completely generic and can be applied to any object class, as long as the descriptor is sufficiently strong.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing - Indexing methods; I.4.9 [**Image Processing and Computer Vision**]: Applications—*Image Processing and Computer Vision*: Applications

1. INTRODUCTION

Finding images that are modified versions of, or in other ways derivative works of, an original image is of great interest to many applications. Our main motivation comes from grouping search results in an image search engine for web images but there are other applications such as e.g. photo-organization software.

Finding images that are resized versions of an original is a relatively simple task and can e.g. be solved with global image descriptors e.g. [4]. Finding cropped versions is a much harder problem and requires some local description of image content, e.g. by detecting invariant regions [5] and extracting local descriptors like [3] for those regions. An even harder case is when portions of the cropped image are modified, e.g. by changing the background or adding new foreground objects. The top row of Figure 1 shows some examples and illustrates a difficult case where a persons face is resized, cropped and placed against different backgrounds. In this paper we show how such "near-duplicates" can be detected using object-centered descriptors.

In this context we define near-duplicates as images where a (central) region of an original image is kept and the remainder modified by any combination of scaling, rotating, cropping, blending (including replacing foreground/background). We also allow scaling and rotation of the (central) region.

2. BACKGROUND & NEAR-DUPLICATES

For a general overview of prior work related to near-duplicate detection and image retrieval see [2]. Due to the nature of these near-duplicates, global descriptors like [4] are not very useful. Instead, regional descriptors [3, 5] are preferable.

The problem addressed in this paper has many similarities with problems addressed within object recognition and differs mainly in the sense that the object region in the image is identical (up to scale and rotation). This means that perspective effects don't have to be considered.

Recently, there has been lots of activity around the use of bag of feature representations for mining large image collections or video. In [7] reoccurring objects are found in video using vector quantized local descriptors tracked over consecutive frames. Later work [6] also adds spatial consistency checks of local features on the top matching images.

In [1] geometric hashing was used on local affine image frames to find similar objects, like e.g. logotypes, in image collections. The method is robust to affine transformations and background and could with some modifications be used for detecting near-duplicates.

We propose to use class-specific descriptors optimized for a particular object class (e.g. faces) instead of the more generic object descriptors described above due to the fact that this gives higher discriminability. In the next section we describe our implementation and then show results from a live implementation running on a very large data set.

3. FINDING NEAR-DUPLICATES USING CLASS-SPECIFIC DESCRIPTORS

As noted above, things that will not work are global image representations, e.g. quantized bag of features, histogram of color or texture. We instead propose to use object class specific descriptors for detecting near-duplicates. The idea being that when images are cropped or otherwise modified, there is a central object mostly unchanged.

Here we will focus on faces, but in principle any object class could be used. All images are processed using a face detector and then each face described using a compact in-house de-

Figure 1: Some samples of automatically detecting near-duplicates in the form of modified versions of an image in a corpus of 60+ million images.

scriptor. Descriptor vectors are small enough for all vectors (∼20 million in the experiment below) to be held in 8GB RAM on a single server.

When implementing in our production setup we had hard constraints on response times (<100ms) for a descriptor lookup. This was solved by binning all face descriptor vectors using a kd-tree. We found that using around 10 levels and $k = 5$ left the number of descriptors in each leaf node small enough to be searched exhaustively. The response-time constraint does not allow for checking nearby leaf nodes so there is no guarantee that all duplicates are found.

4. EXAMPLES AND RESULTS

The method described above was implemented in an image search engine with images collected using a browser plugin installed by users who also manually annotate parts of the faces detected. The total number of images was just over 60 million with about 20 million faces.

Some results are shown in Figure 1 (first five duplicates shown for selected searches on http://search.polarrose.com/). Each row contains five duplicates based on face recognition. For more examples of near-duplicate results we refer to the website http://search.polarrose.com/ (click the duplicates link above images on the search results page).

5. CONCLUSIONS

We proposed a method for finding near-duplicate images in collections of millions of images using object-specific descriptors. A large-scale deployment using face recognition was discussed and sample results shown. More results are publicly available on the website. These results show that the procedure is very robust to the image modifications considered here and that it is possible to pick out difficult examples among millions of images.

6. REFERENCES

[1] O. Chum and J. Matas. Geometric hashing with local affine frames. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:879–884, 2006.

[2] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):1–60, April 2008.

[3] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[4] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):703–715, 2001.

[5] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, 2005.

[6] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[7] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 1470–1477, Oct. 2003.