České vysoké učení technické v Praze
Fakulta elektrotechnická

Czech Technical University in Prague
Faculty of Electrical Engineering

Mgr. Ondřej Chum, Ph.D.

Objevování prostorově souvisejících obrázků ve velkých obrazových databázích

Large-Scale Discovery of Spatially Related Images

Summary

We present a randomized data mining method that finds clusters of spatially overlapping images. The core of the method relies on the min-Hash algorithm for fast detection of pairs of images with spatial overlap, the so-called cluster seeds. The seeds are then used as visual queries to obtain clusters which are formed as transitive closures of sets of partially overlapping images that include the seed. We show that the probability of finding a seed for an image cluster rapidly increases with the size of the cluster.

The properties and performance of the algorithm are demonstrated on datasets with 10^4, 10^5, and $5 \cdot 10^6$ images. The speed of the method depends on the size of the database and on the number of clusters. The first stage of seed generation is close to linear for databases sizes up to approximately $2^{34} \approx 10^{10}$ images. On a single 2.4GHz PC, the clustering process took only 24 minutes for a standard database of more than hundred thousand images, i.e. only 0.014 seconds per image.

Souhrn

V této práci představujeme metodu randomizovaného vytěžování dat, který nachází shluky obrázků s překrývajícím se obsahem. Metoda je založena na algoritmu min-Hash pro rychlou detekci párů obrázků s překrývajícím se obsahem, takzvaná semínka shluků. Semínka jsou následně použita jako vizuální dotazy k získání shluků. Ty jsou vytvořeny jako komponenty souvislosti v grafu, kde vrcholy jsou obrázky a hrany představují částečný překryv. Ukazujeme, že pravděpodobnost nalezení semínka shluku obrázků rychle roste s velikostí shluku.

Vlastnosti a chování algoritmu jsou demonstrovány na množinách 10^4, 10^5 a $5 \cdot 10^6$ obrázků. Rychlost metody záleží na velikosti databáze a na počtu shluků. První část algoritmu – generování semínek – je téměř lineární pro databáze velikosti přibližně $2^{34} \approx 10^{10}$ obrázků. Na jednom standardním 2.4GHz PC trvá nalezení shluků v databázi o velikosti více než sto tisíc obrázků pouze 24 minut, tedy přibližně 0.014 sekundy na obrázek.

Klíčová slova: shlukování obrazů, vytěžování obrazových dat, vyhledávání obrázků, neorganizované množiny obrázků, min-Hash

Keywords: image clustering, image data mining, image retrieval, unorganized image sets, min-Hash

Contents

Chapter 1

Introduction

Collections of images of ever growing sizes are becoming common both due to commercial efforts, such as Google street view, and as a result of photo and video sharing of individual people (e.g. Flickr [8]). Structuring and browsing large images databases is a challenging problem. Developments like Photo Tourism [22] show that access to images based on the 3D acquisition location or on the spatial overlap of the scenes they depict is intuitive and has high user acceptability. Commonly, the sets of relevant spatially related images are obtained using manual annotations. We propose a method for discovering spatial overlaps using image content only via image retrieval techniques.

We formulate the task of discovery of spatially related images as finding connected components in a graph. Vertices of the graph represent images. Two images are related (connected by an edge) if they depict the same scene. From the point of view of the fast clustering algorithm, we adopt a pragmatic definition: a pair of images depicts the same scene if they can be matched by some robust matching method. An example of a matching graph is shown in Figure 1.1.

While the vertices of the graph are defined as the image database, the edge structure is not known a-priori and has to be discovered by the clustering algorithm. An image-retrieval system (e.g. [16]) can be thought of as an efficient method that, given one vertex (an image), returns all edges to related images. In most current retrieval systems, a query has complexity linear in the number of images in the database, but is many orders of magnitude faster than actually attempting to match every single image to the query image.

The min-Hash, on the other hand, is a hashing method for *fast* retrieval of edges. However, the price paid for the efficiency of the method is a low recall: each edge is only discovered with a certain probability. The probability is proportional to the image pair similarity based on the fraction of common visual words shared by the images. The probability is high (close to one) only for near duplicate images, which is the domain where the min-Hash has been used so far [1, 3]. The complexity of this approach is linear in the number of images in the database.

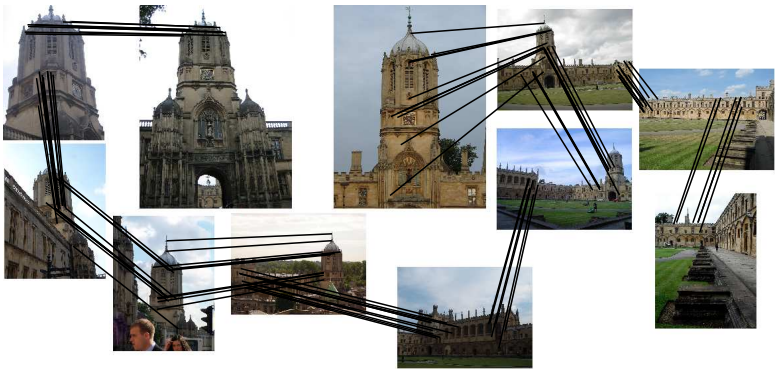


Figure 1.1: Visualization of a part of a cluster of spatially related images automatically discovered from a database of over 100K images (Oxford 105K dataset). Overall, there are 113 images in the cluster, all correctly assigned. A sample of geometrically verified correspondences is depicted as links between images. Note that the images show the tower from opposite sides.

The properties and performance of the algorithm are demonstrated on datasets with 10^4, 10^5, and $5 \cdot 10^6$ images. The speed of the method depends on the size of the database and on the number of clusters. The first stage of seed generation is close to linear for databases sizes up to approximately $2^{34} \approx 10^{10}$ images. On a single 2.4 GHz PC, the clustering process takes only 24 minutes for a standard database of more than hundred thousand images, or equivalently, only 0.014 seconds per image.

1.1 Related work on unsupervised object and scene discovery

The problem of matching (organization) of an unordered image set was introduced by Schaffalitzky and Zisserman in [19] for sets of tens of images. The approach closest to ours is [21] by Sivic and Zisserman who aimed at unsupervised discovery of multiple instances of particular objects in feature films. The is quadratic in the number of images and hence not suitable for large scale clustering. Large scale clustering has been recently demonstrated by Quack et al. in [18], who use the GPS information to reduce the large scale task down into a set of smaller tasks. Further related methods are described and discussed in detail in [7].

Chapter 2

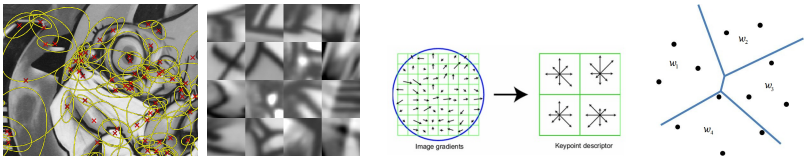
Background Review

This chapter briefly reviews the adopted image representation and the basic min-Hash algorithm that are necessary to understand the topic and the contributions presented in this thesis.

2.1 Image representation

Recently the majority of image retrieval systems has adopted the bag-of-words image representation [20]. The details vary from method to method, but most of the approaches follow a common scheme: 1) detection of local features, 2) feature description, 3) vector quantization of the feature descriptors. The three steps are visualized in Figure 2.1.

First, local regions of interest are detected [14]. Typically, detectors covariant with a similarity transformation (DoG [12], or with an affine transformation (MSER [13], Harris-affine [14], or Hessian-affine [16]) are used. Features covariant with an affine transformation are more general. Such features are designed to be repeatedly detected under a projective transformation of a close-to-planar surface. Such a mapping is locally well-approximated by an affine transformation. A comprehensive comparison of different feature detectors is provided in [14].



(a) (b) (c) (d)

Figure 2.1: Construction of the bag of words for image representation: (a) affine co-variant features, (b) examples of features transformed into a canonical frame, (c) the SIFT descriptor – histogram of gradients over the canonical frame – image taken from [12], (d) k-means is used to construct a visual vocabulary – each cell represents a different visual word w_i.

In feature description stage, each image feature is assigned a vector from a vector space called the descriptor space. The feature similarity is then defined as a distance (typically L_2) in the descriptor space. The invariance of the descriptor to a geometric transformation is achieved by geometric normalization into a canonical frame. Any descriptor computed over the canonical frame can be used, the most popular being the SIFT descriptor [12].

To reduce the memory footprint and to enable a fast access to features with similar descriptors, the descriptor space is vector-quantized into a visual vocabulary. Instead of the full descriptors, only (the identifier of) the vector quantized prototype for visual word is kept. Different vector quantization methods have been proposed for visual vocabulary construction: k-means [20], hierarchical k-means [15], approximate k-means [17] and others.

The images are represented as bags (multi-sets) of visual words. Such a representation has been shown to be well-suited for large-scale image search and analysis. The local-feature approach introduces robustness to occlusions and viewpoint change. The vector-quantized descriptors are a reasonably compact representation that allows for efficient search using techniques adopted from large-scale text-retrieval community.

2.2 Introduction to min-Hash

The min-Hash algorithm, introduced in [1], is a Locality Sensitive Hashing [9] for sets. In the min-Hash method, images are represented as sets of visual words. This is a weaker representation than a bag of visual words since word frequency information is reduced into binary information (present or absent). However, it was shown that for large vocabularies the set-of-words and bag-of-words representations are almost identical [2].

There is a number of equivalent definitions of the min-Hash. It will be convenient to use the definition exploiting ordering of the vocabulary by a random permutation π. Let $\mathcal{W}$ be the set of visual words, $N = |\mathcal{W}|$ be the size of the vocabulary and

$$\pi(i) : \{1 \dots N\} \to \{1 \dots N\}$$

a permutation of N elements. Let $p(i)$ be an inverse function to $\pi(i)$. That is, $\pi(i)$ gives the rank of visual words $w_i \in \mathcal{W}$, while $p(i)$ is the index of ith smallest visual word in the ordering induced by π. The random permutation π is often implemented as a hash function $f(w_i)$, so that $\pi(i) < \pi(j)$ iff $f(w_i) < f(w_j)$.

A min-Hash signature of a set $\mathcal{A} \subset \mathcal{W}$ is defined as $h(\mathcal{A})$, where

$$h(\mathcal{A}) = \min_{i:w_i \in \mathcal{A}} \pi(i). \tag{2.1}$$

Such a function has the property that the probability of two sets having the same value of the min-Hash signature for a random permutation π is equal to their set overlap [1, 3], i.e. the ratio of the cardinalities of the intersection and union of the two sets. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be sets of visual words. To simplify the notation and terminology, in connection with min-Hash, we use the term 'similarity' for the set overlap:

$$\text{sim}(\mathcal{A}_1, \mathcal{A}_2) = \frac{|\mathcal{A}_1 \cap \mathcal{A}_2|}{|\mathcal{A}_1 \cup \mathcal{A}_2|} \in [0, 1]. \tag{2.2}$$

The probability of two images having the same min-Hash signature is then

$$P\left[h(\mathcal{A}_1) = h(\mathcal{A}_2)\right] = \text{sim}(\mathcal{A}_1, \mathcal{A}_2).$$

To estimate the similarity of two images, multiple independent min-Hash functions h_j (i.e. independent permutations π_j of the vocabulary) are used. The fraction of the min-Hash functions that assigns an identical min-Hash signature to the two sets is an unbiased estimate of the similarity of the two images.

Retrieving similar images. So far, a method to estimate the similarity of two images was discussed. To efficiently retrieve images with high similarity, the values of min-Hash functions h_i are grouped into s-tuples called sketches. Similar images have many values of the min-Hash function in common (by the definition of similarity), and thus have a high probability of having the same sketches. On the other hand, dissimilar images have a low chance of forming an identical sketch. Identical sketches are efficiently found by hashing [10].

The probability of two sets having at least one sketch out of r in common is

$$P_C(\mathcal{A}_1, \mathcal{A}_2) = 1 - (1 - \text{sim}(\mathcal{A}_1, \mathcal{A}_2)^s)^r. \tag{2.3}$$

The probability depends on the similarity of the two images and on the two parameters of the method, which are the size of the sketch s, and the number of (independent) sketches r. Figure 2.2 visualizes the probability of collision plotted against the similarity of two images for fixed $s = 3$ and $r = 512$.

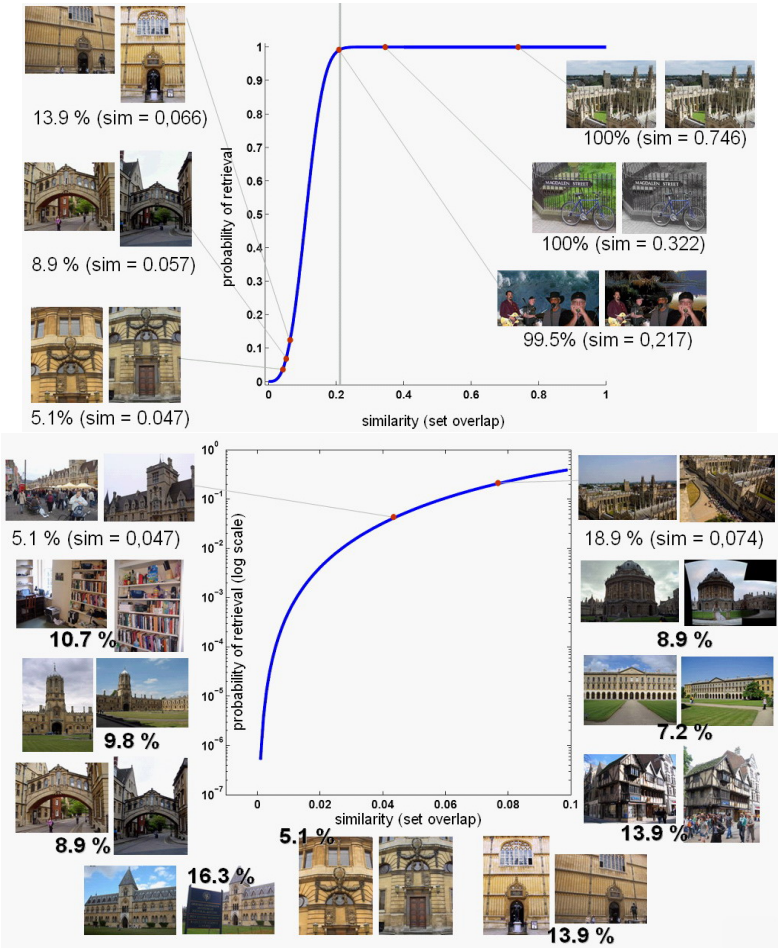


Figure 2.2: The probability of at least one sketch collision for two documents plotted against their similarity; with $r = 512$ sketches, $s = 3$ min-Hashes per sketch. Image pairs of different similarities are added to relate to the 'visual similarity'. The bottom plot shows a close-up of the bottom left corner of the left plot. Note the logarithmic vertical axis.

Chapter 3

Randomized Clustering

3.1 Cluster seed generation

In this section, a randomized procedure that generates seeds of possible clusters of images is described. Let us first look at the plot of the probability of sketch collision as a function of image similarity depicted in Fig. 2.2. The sigmoid-like shape of the curve is important for the near duplicate detection task [3]. Image pairs with high similarity are retrieved with a probability close to one. The probability drops rapidly – through similar image pairs (typically images of the same object from a slightly different viewpoint) that are occasionally retrieved to unrelated image pairs (with similarity below 1%) that have close to zero probability of being retrieved.

Now, for the purpose of data mining, let us focus on the bottom left corner of the graph. According to eqn. (2.3), an image pair with similarity $\mathrm{sim} = 0.05$ has probability 6.2% to be retrieved (using 512 sketches of size 3). Such a poor recall is certainly below acceptable level for a retrieval system. However, we do not aim at retrieving all relevant images from the image clusters in a single step. The task is to quickly detect seeds from the clusters – it is sufficient to retrieve a single seed per cluster, and we are fortunate that the importance of a cluster is related to its size in the database.

The probability that not a single image pair (seed) is found by the min-Hash depends on two factors – the similarity of the images in the cluster and the number of image pairs that actually observe the same object. In the following analysis, which demonstrates an approximate lower bound on this probability, we assume that a particular object or landmark is seen in v views. The probability that none of the pairs $(\mathcal{A}_i, \mathcal{A}_j)$ of v views is retrieved is approximated by

$$P\{\mathrm{fail}\} = \prod_{i \neq j} 1 - P_C(\mathcal{A}_i, \mathcal{A}_j) = (1 - \varepsilon)^{\frac{v(v-1)}{2}}. \tag{3.1}$$

Here, ε stands for an "average" collision probability. The "average" cluster similarity is then defined by eqn. 2.3. The plot in Fig. 3.1 shows that for pop-

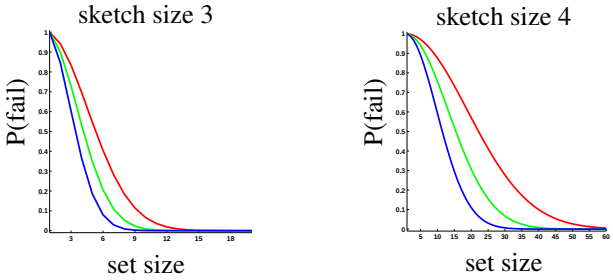


Figure 3.1: Probability of failure to generate a seed in a set of images depicting the same object using min-Hash with 512 sketches of size 3 (left) and 4 (right); note the different scales on the horizontal axes. The three curves show the dependence for different 'average' similarity equal to 7% (lowest curve), 6% (middle) and 5% (highest).

ular places (i.e. those where photos are often taken from) the probability of failure to retrieve an image pair vanishes. There are three plots for similarities 5%, 6% and 7%. Since the similarity is defined as a ratio of the size of the intersection over the size of the union, the difference between similarity 6% and 5% is substantial. Going from 6% to 5% similarity means removing 17.5% of elements that were in the intersection.

It is important to point out that the probability of finding a seed depends on the image similarities and the number of views and is completely *independent* of the size of the database. The v views have the same chance to be discovered in a database of 5000 images as in a database of several millions of images without any need to change the method parameters or re-hash. This is not true for many topic discovery approaches.

Time complexity. The method is based on hashing with a fixed number M of bins. The number of bins is based on the size of the vocabulary which cannot be infinitely increased without splitting descriptors of the same physical region. Assuming the uniform distribution of the keys, the number C of keys that fall into the same bin is a random variable with a Poisson distribution where the expected number of occurrences is $\lambda = D/M$ (the number of documents divided by the number of bins in the hashing table). The expected number of key pairs that fall into the same bin (summed over all bins) is

$$\sum_{i=1}^{M} \mathbf{E}(C^2) = \sum_{i=1}^{M} \left(\lambda^2 + \lambda\right) = \frac{D^2}{M} + D. \tag{3.2}$$

The asymptotical time complexity is $\mathcal{O}(D^2)$ for D, i.e. size of the image database, approaching the infinity. However, for finite databases of sizes up

to $D \leq M$, the method behaves as linear in the number of documents since $D^2/M + D \leq 2D$. In the min-Hash algorithm, the number of keys depends on the size of the vocabulary w and the size of the sketch s and is proportional to $M = w^s$. In the experiments in this paper, we used $w = 2^{17}$ and $s = 3$ or $s = 4$. This gives the number of different hash keys $M = 2^{51}$ and $M = 2^{68}$. We believe that this number is sufficient to conveniently deal with web scale databases.

3.2 Growing the seed

We build on the query expansion technique [4] to increase the recall. The idea is as follows: an original query is issued and the results are then used to issue new query. Not all results are used, only those that have the same spatial feature layout (for more details on spatial verification see the following section). The spatial verification prevents the query expansion from so-called topic drift, where an unrelated image is used to expand the query.

Time complexity. Each query is linear in the number of images in the database. Hence, the time complexity of completing the connected components is $\mathcal{O}(DV)$, where D is the size of the database and V is the number of images in all clusters. The worst case behaviour of this step is thus quadratic, when every image is assigned to one of the clusters. In practice, however, we observe that $V \ll D$, which brings immense computational savings.

Further reduction of the time complexity can be achieved by the following observation. The number of images of one object (say the Colosseum in Rome) will typically grow with the size of the dataset, but the number of different viewpoints gets saturated after certain amount of images is exceeded. Grouping images into similar viewpoints (based on a global descriptor) has been used in [11]. In the proposed approach, for very large clusters (over 500 images), we exclude all images with large number of matches (more than 50) from the query expansion step. This does not have a significant impact on the recall, since well matching images usually do not carry sufficient amount of new information to be used in the enhanced query. It also reduces the time complexity to $\mathcal{O}(DL)$, where L is the number of clusters rather than the number of images in all clusters.

3.3 Experimental Evaluation

We have conducted two experiments. The first one checks whether the probability of seed generation is sufficiently high on real data as predicted by theoretical estimates presented in Section 3.1. In the second experiment, clusters of spatially related images are discovered in a database of 100K images.

3.3.1 Seed generation success rate

To evaluate the success rate of the seed generation stage on real data, we use a standard image retrieval benchmark dataset (the University of Kentucky dataset) introduced in [15]. This database contains 10200 images; a group of 4 images depicts the same object / scene, i.e., there are 2550 clusters of size four. The dataset provides images, detected image features and SIFT descriptors. The provided features and descriptors were used.

The objective is to measure for how many clusters (all of size four) the proposed method generates at least one seed. For this experiment, we have used a visual vocabulary of 2^{17} visual words. For each image, 512 independent random min-Hash functions were evaluated and grouped into 512 sketches of size 3 (individual min-Hashes were used multiple times). With this setting, there are 11556 pairs of images with at least one common sketch value (a sketch collision) of which 3553 passed the similarity test at 0.045 (step 2 of the clustering procedure); out of the 3553 seeds 3210 were within a ground-truth defined group of four images. The number of clusters of four images for which at least one pair was suggested by the hashing is 1196 (out of 2550 possible clusters). In other words, a seed for a cluster of size four is generated with a probability of 46.9%, which is very close to the expected value of failure, see Fig. 3.1, left plot. The approximately 50% probability of detecting a cluster might seem low, but a cluster of four images is much smaller than typical clusters in image collections containing $10^5 - 10^7$ images. The experiment shows performance of the algorithm for the smallest practical cluster size.

In Fig. 3.2, we compare the predicted success / failure rate (from eqn. (3.1)) and the empirical failure rate. In the experiment, the "average" collision probability ε was computed (exactly) for each cluster by enumerating all image pairs within the cluster. For each cluster, we also observe whether a seed has been generated in the cluster or not. Fig. 3.2 plots the frequency of observed seed generation success rate for different levels of predicted success rate. The histogram closely follows the grey identity line. We conclude that the prediction given in eqn. (3.1) is precise for the Kentucky dataset.

3.3.2 Clustering on the 100K Oxford Landmark Database

The experiment was conducted on a large database of images downloaded from Flickr [8]. This database contains $5,062$ images from publicly available Oxford Landmark Database[1] and $99,782$ from *Flickr1* dataset[2] used in [17]. Both sets are composed of high resolution images (1024×768). The dataset

[1] http://www.robots.ox.ac.uk/∼vgg/data/oxbuildings/

[2] Courtesy of VGG, University of Oxford

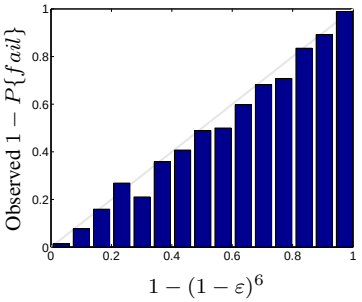


Figure 3.2: Histogram of observed success rate plotted against the expected success rate on the Kentucky dataset.

consists of images, as well as detected features with SIFT descriptors – these standard features and descriptors were used in the experiment. Together, there are $104,844$ images with $294,105,803$ features (2805 features per image on average). The SIFT descriptors of the features occupy 35GB. In this dataset, images of 11 landmarks were manually labelled. Presence of each landmark in an image is characterized by one of four labels: (i) Good – a nice, clear picture of the object, (ii) OK – more than 25% of the object is clearly visible, (iii) Junk – less than 25% of the object is visible, or there is a very high level of occlusion or distortion, and (iv) Absent – the object is not present.

As in the previous experiment, we used a vocabulary with 2^{17} visual words for min-Hash seed generation and with 1M words for seed growing. The Oxford Landmark Database contains clusters with $10^2 - 10^3$ images. To show the potential of the method, we used 512 min-Hashes grouped into 512 sketches of size three. These settings allow to discover even small clusters of several images with reasonable probability and are the same as in the University of Kentucky database experiment. On average, the min-Hash generated 38.4 sketch collisions per image. These were reduced to 1.23 potential seeds per image by thresholding the estimated similarity at 0.045 – this corresponds to $129,341$ seeds. Out of those, 3103 images were found to have an exact duplicate in the database (the same image was downloaded under different user tags), and 289 images were found to have a near duplicate. Both exact and near duplicates were dropped and the remaining potential seeds were subject to spatial verification, leaving 441 verified seeds. This number is an upper bound on the number of clusters, since typically there are multiple seeds per cluster. The seed growing by query expansion discovered 354 distinct clusters covering $2,643$ images.

Table 3.1 summarizes the results on objects with ground truth information. For each landmark, we found cluster containing the most positive

	Good	OK	sketch 3	unrelated	sketch 4
All Souls	24	54	97.44	0	97.44
Ashmolean	12	13	68.00	0	0
Balliol	5	7	33.33	0	0
Bodleian	13	11	95.83	1	95.83
Christ Church	51	27	89.74	0	89.74
Cornmarket	5	4	66.67	0	0
Hertford	35	19	96.30	1	0
Keble	6	1	85.71	0	0
Magdalen	13	41	5.56	0	1.85
Pitt Rivers	3	3	100.00	0	0
Radcliffe Camera	105	116	98.64	0	98.46

Table 3.1: Results for annotated images in the Oxford Building Dataset. The first two columns show the number of ground truth images labelled 'Good' and 'OK' respectively. The column 'sketch 3' displays the percentage of labelled images that were clustered into a single cluster using min-Hash with sketches of size three, 'unrelated' gives an absolute number of unrelated images in that cluster. The column 'sketch 4' presents results for sketches of size four.

(Good and OK) images of that landmark and computed the fraction of positive ground truth images in this cluster. Also, the absolute number of unrelated images is reported by eye-balling these clusters. Other buildings that appear in the same cluster are not considered unrelated if images linking these objects exist. For example, images of All Souls and the Radcliffe Camera are all in one cluster – they are right next to each other and appear together on several images.

Clusters corresponding to all ground-truth objects were successfully discovered with the exception of the Magdalen Tower. The percentage of images assigned to the relevant cluster is consistent with the retrieval results in [17, 4] and is related to the 'difficulty' of each landmark. This also holds for the 'Magdalen' – reported retrieval results were by far the worst for this landmark. In our experiment, three images of the tower were discovered and the method was unable to spatially verify and grow to any other image.

Setting sketch size to three is suitable for demonstrating the method on a database of 100K images. It allows retrieving even small, perhaps uninterestingly small, clusters. These settings will not be acceptable for web scale database size of more 10^7 images or more. To simulate real conditions, we have also used 512 sketches of size four, which is suitable for very large databases, but returns with acceptable probability only larger clusters. Still, the size of discovered clusters is comparable (or smaller) than the size of clusters used in Photo Tourism [22]. The four largest clusters from the Oxford Landmark ground truth were discovered (together with other larger clusters

22,440 : sagradafamilia; sagrada; familia; gaudi; lasagradafamilia

10,682 : colosseum; coliseum; colosseo; coliseo; colloseum

4,539 : oldtownsquare; czechrepublic; czech; astronomicalclock; prague

Figure 3.3: Sample of large clusters discovered in the 5M database. Size of the cluster and the five most discriminative Flickr tags are shown beneath the images. Note the variety in scale, viewpoint, and illumination conditions.

that are not included in the ground truth).

Timing. The seed generation took 7 min 47 sec and the seed growing took 16 min 20 sec on a 2.4GHz PC using a single processor (MATLAB / MEX implementation). The complete processing of the database took thus slightly more than 24 minutes (the time does not include the feature extraction, SIFT computation, vector quantization, nor database indexing), which corresponds to 0.014 seconds per processed image. Note that all steps of the proposed method are easy to parallelize.

3.3.3 Large-scale clustering of 5 million images

We have executed the clustering on a database of 5 million Flicker images. In this experiment we have used: Hessian affine features [16], a vocabulary of 1M visual words, sketch size $s = 4$, and $k = 512$ sketches. The clustering took slightly under 28 hours on a single machine (3.0GHz PC, 64GB memory, using a single core), which is 0.020 seconds per image. Out of the 5M images, 474434 were assigned to 16957 clusters. Fig. 3.3 shows samples of some detected clusters together with the five most discriminative user tags for that particular cluster.

3.4 Conclusions

We have proposed a method for discovering spatially-related images in large scale image databases. Its speed depends on the size of the database and is very fast in practice and close to linear for database sizes up to approximately $2^{34} \approx 10^{10}$ images. The success rate of cluster discovery is dependent on the cluster size and the average similarity within the cluster and is independent

of the size of the database. The properties and performance of the algorithm were demonstrated on datasets with 10^4, 10^5, and $5 \cdot 10^6$ images.

References

[1] A. Broder. On the resemblance and containment of documents. In *SEQS: Sequences '91*, 1998.

[2] O. Chum, M. Perdoch, and J. Matas. Geometric min-hashing: Finding a (thick) needle in a haystack. In *CVPR*, 2009.

[3] O. Chum, J. Philbin, M. Isard, and A. Zisserman. Scalable near identical image and shot detection. In *CIVR*, 2007.

[4] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *ICCV*, 2007.

[5] O. Chum and J. Matas. Matching with PROSAC - progressive sampling concensus. In *CVPR*, June 2005.

[6] O. Chum, J. Matas, and S. Obdržálek. Enhancing RANSAC by generalized model optimization. In *ACCV*, 2004.

[7] O. Chum and J. Matas. Web scale image clustering. *IEEE PAMI*, 32:371–377, 2010.

[8] http://www.flickr.com/.

[9] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *of Symposium on Theory of Computing*, 1998.

[10] D. E. Knuth. *The Art of Computer Programming, Volume III: Sorting and Searching*. Addison-Wesley, 1973.

[11] X. Li, C. Wu, C. Zach, S Lazebnik, and J.-M. Frahm. Modeling and recognition of landmark image collections using iconic scene graphs. In *ECCV*, 2008.

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC.*, 2002.

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IJCV*, 65(1/2):43–72, 2005.

[15] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *CVPR*, 2006.

[16] M. Perdoch, O. Chum, and J. Matas. Efficient representation of local geometry for large scale object retrieval. In *CVPR*, 2009.

[17] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007.

[18] T. Quack, B. Leibe, and L. Van Gool: World-scale mining of objects and events from community photo collections. In*CIVR* 2008.

[19] F. Schaffalitzky and A. Zisserman: Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". In *ECCV*, 2002.

[20] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, 2003.

[21] J. Sivic and A. Zisserman: Video data mining using configurations of viewpoint invariant regions. In *CVPR*, 2004.

[22] N. Snavely, S. Seitz, and R. Szeliski. Photo Tourism: exploring photo collections in 3D. In *ACM SIGGRAPH*, pages 835–846, 2006.

Mgr. Ondřej Chum, Ph.D.

cmp.felk.cvut.cz/∼chum
Born on 11th November 1976 in Nymburk, Czech Republic

since 2007	Senior researcher at CMP, Dept. of Cybernetics, Faculty of EE, Czech Technical University in Prague
2006–2007	Postdoc at the VGG, University of Oxford, United Kingdom
2005–2006	Researcher at CMP, Dept. of Cybernetics, FEE, CTU in Prague
06–12 2002	University of Surrey, Guildford, United Kingdom
2001-2005	PhD at CMP, Dept. of Cybernetics, FEE, CTU in Prague
1996–2001	Master at the Faculty of Mathematics and Physics, the Charles University, Prague

Awards:

- The runner up award for the "2012 Outstanding Young Researcher in Image & Vision Computing" by the Journal of Image and Vision Computing for researchers within seven years of their PhD
- Winner of VOC PASCAL 2007 challenge detection task at the ICCV 2007, Rio de Janeiro, Brazil.
- A member of CMP team that came second at the ICCV 2005 Localization Contest, Beijing, China.
- British Machine Vision Association award for the best scientific paper at British Machine Vision Conference 2002, Cardiff, United Kingdom

Currently, O. Chum is co-supervising two PhD students and leading a number of master students. He is lecturing Processing of Medical Images and a graduate course Understanding State of the Art Methods, Algorithms, and Implementations. He was/is leading a team of 4–6 master and 2–6 PhD students under CTU SGS grants. He has (co-)acquired a number of grants, including 2 GAČR grants, Google research award, and Microsoft Live Labs grant. In 2012, his ERC proposal was evaluated as fundable.

O. Chum served as an area chair at international conferences ICCV 2011, BMVC 2008 and 2012. He regularly reviews for major computer vision journals and conferences. He has co-organized Vision and Sports Summer School 2012 in Prague, 25 Years of RANSAC workshop and tutorial in conjunction with CVPR 2006, New York, USA, and Computer Vision Winter Workshop 2006 in Telč, CR.

O. Chum has co-authored 6 international high-impact journal papers and over 25 international peer-reviewd conference papers. His work has been cited over 1500 times (excluding autocitations).