Geometric min-Hashing: Finding a (Thick) Needle in a Haystack

Ondřej Chum, Michal Perďoch and Jiří Matas
CMP, Dept. of Cybernetics, Faculty of Elec. Eng., Czech Technical University in Prague
chum@cmp.felk.cvut.cz

Abstract

We propose a novel hashing scheme for image retrieval, clustering and automatic object discovery. Unlike commonly used bag-of-words approaches, the spatial extent of image features is exploited in our method. The geometric information is used both to construct repeatable hash keys and to increase the discriminability of the description. Each hash key combines visual appearance (visual words) with semi-local geometric information.

Compared with the state-of-the-art min-Hash, the proposed method has both higher recall (probability of collision for hashes on the same object) and lower false positive rates (random collisions). The advantages of Geometric min-Hashing approach are most pronounced in the presence of viewpoint and scale change, significant occlusion or small physical overlap of the viewing fields. We demonstrate the power of the proposed method on small object discovery in a large unordered collection of images and on a large scale image clustering problem.

1. Introduction

Algorithms based on hashing techniques are the core of methods that have produced impressive results for a range of computer vision problems, like matching of point sets [14], object recognition [8], image retrieval [27], duplicate detection and clustering in large image collections [5].

In the paper, we propose a novel hashing scheme – *the Geometric min-Hash* (GmH). The advantages of the Geometric min-Hashing have high impact in problems involving significant occlusion or small physical overlap of the viewing fields. In such cases the difference in recall and precision reach orders of magnitude compared to min-Hash [7] algorithm. Moreover, GmH is less sensitive to scale changes. The advantages of the min-Hash, *e.g.* compact representation and robustness, are preserved. The potential of the method is demonstrated on small object discovery in a large unordered collection of images (see Fig. 1).

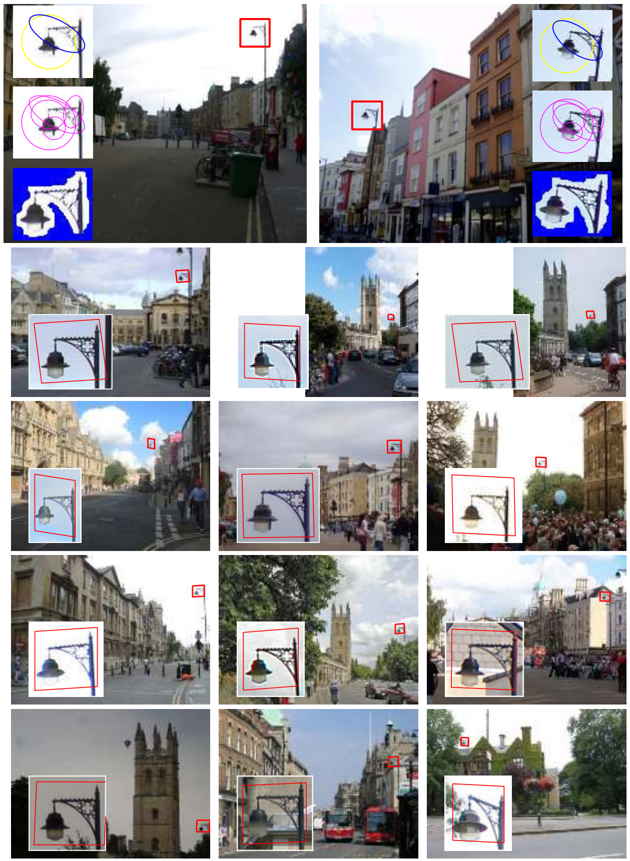


Figure 1. An automatically discovered object in an unordered collection of 100k images. The lamp is visible in approximately 0.014% of the images and it covers on average about 0.28% of pixels of those images. Top: the ‘seed’ image pair found by the Geometric min-Hash. The three close-ups show the colliding sketch, its geometric support and the co-segmentations, respectively. Bottom: other detections (with close-ups) obtained by object retrieval using the seed pair. Note that *all* bounding boxes are discovered, *not* drawn by the user.

The min-Hash describes images by selecting *independently* visual words as global descriptors, with the property that the higher number of common features in two images, the higher the probability of having the same min-Hash. A single min-Hash is not sufficiently discriminative to support indexing and thus min-Hashes must be grouped into s-tuples (*sketches*) for hashing. In order to find a pair of pos-

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

sibly related images (objects, windows), *all* s min-Hashes in a sketch must agree. Increasing s reduces both the number of random collisions and true hits exponentially. Due to different "half-times" in the exponential reduction of collisions, the ratio of true to false positives increases rapidly too. On the other hand, the "and" operation exponentially reduces the probability of retrieving all instances of the image or object.

Motivation. Geometric min-Hashing improves on the min-Hash by achieving higher recall, higher precision (lower false positive rate) simultaneously. The algorithmic change is motivated by the following observations:

Observation 1 (on uniqueness): with large vocabularies (such as 1M visual words), visual words in images usually do not appear more than once. Selecting a visual word from an image is typically equivalent to selecting a feature (visual word with spatial location and extent).

Observation 2 (on repeatability): s-tuples of features localized in space and scale have (much) higher repeatability than random s-tuples.

Both observations have a simple statistical explanation. Apart from images with repeated structures that violate the assumption of feature independence, Observation 1 is a consequence of the fact that the ratio of the number of features in an image ($\approx 10^3$) and the number of visual words (10^6) is small ($\approx 10^{-3}$). The number of occurrences of a visual word in an image (assuming independence) is thus well approximated by the Poisson distribution with a very small lambda and no or one occurrence is by far most likely.

Observation 2 captures the fact that repeatability is higher in compact local neighbourhoods mainly due to occlusions and the presence of occluding boundary and if the features have similar scale. Feature detectors operate well only for a certain range of scales and features with similar scale enter/exit the range "together". Similarly for occlusions – if a features is (is not) occluded, its neighbour is likely to be in the same state.

The two observations lead naturally to the following ideas: 1. Select first a min-Hash (a unique visual word - a feature) from the whole image and the rest of the sketch randomly from features close (in space and scale) to the first feature; and 2. Since all min-Hashes in the sketch must match, their mutual spatial position provides a geometric invariant.

We show both by theoretical considerations and empirically that GmH performance degrades very slowly with the reduction of overlap between images. GmH not only proposes a matching image pair, but also the image transformation that locally maps one image to another. Unlike the commonly used bag-of-words approaches, the spatial extent of image features is exploited in GmH. The geometric information is used both to construct repeatable hash keys and to increase the discriminability of the description. Each hash key combines visual appearance (visual words) with semi-local geometric information. The GmH representation has strong indexing ability, allowing discovery of small object in a collection of over 100k images (see Section 4).

From another perspective, the GmH can be viewed as an implementation of the following reasoning: if the first min-Hash in the s-tuple is not repeated, bad luck. However, if it is repeated, the selected feature provides information how to select a second min-Hash with much higher repeatability, while maintaining the discriminative power.

Related work. Hashing is a popular method of image retrieval due to its speed. Recently a lot of attention has been paid to the GIST descriptor [19] and fast retrieval of similar images [27, 15]. Jain *et al.* [12] introduced a method for efficient extension of Locally Sensitive Hashing scheme [11] for Mahalanobis distance.

Another popular hashing scheme is min-Hash [2], originally used for near duplicate detection of text documents or images [7]. This method is, to some degree, insensitive to occlusion, viewpoint and scale changes. However, if any of the aforementioned effects exceeds certain level, the probability of success of min-Hash decays rapidly. We build on min-Hash and we directly compare with it.

Geometric hashing approaches [14, 4] are used to match two (or a few) point clouds or images. The representation of the geometric hash is not very compact and it is difficult to scale to hundreds or thousands of images. Beside that, geometric min-Hash completely ignores the local visual appearance.

There is a limited literature on a *small* object discovery from *large* image collections. The closest work on this topic is [25] and [22], the databases used in this paper are an order of magnitude larger.

Following recent work on image retrieval [24, 18, 21, 13], affine invariant features and descriptors [17] are used, images are represented as bags (or sets) of words (vector quantized descriptors) [24]. In particular, we use hessian affine features and the SIFT descriptor [16] with gravity vector – details can be found in [20].

The structure of the paper is as follows. Section 2 gives a brief overview of the min-Hash algorithm, providing only the background necessary for understanding the rest of the paper. Section 3 presents the proposed approach, properties and implementation details are further discussed in Sections 3.1 and 3.2. Two applications, clustering of unordered image collections and large–scale small–object discovery, and experimental results are presented in Section 4.

2. The min-Hash algorithm overview

The min-Hash algorithm is a Locality Sensitive Hashing [11] for sets. A brief overview of the min-Hash algorithm follows; for detailed description see [2, 7]. For the purpose

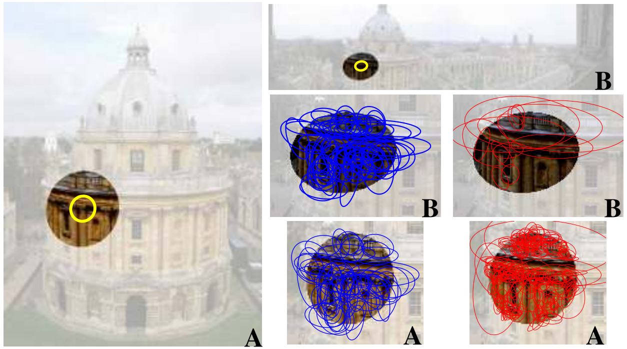


Figure 2. Generating a sketch by Geometric min-Hash. The *central* feature (left & top) is selected using min-Hash from all unique visual words in the images (independently). The remaining features in the sketch are selected from affine covariant neighbourhoods (highlighted in the images) of the central feature using independent min-Hash functions. The secondary min-Hash selects only from features with scale similar to the central feature (left column), too small or too large features are not considered (right column). Letters A and B are used to distinguish the images.

of min-Hashing, 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 a binary information (present or absent). A min-Hash is a function f that assigns a number to each set of visual words (each image representation). The function has a property that the probability of two sets having the same value of the min-Hash function is equal to their set overlap, *i.e.* the ratio of the intersection and union of their set representations

$$\mathrm{ovr}_1(\mathcal{A}_1, \mathcal{A}_2) = \frac{|\mathcal{A}_1 \cap \mathcal{A}_2|}{|\mathcal{A}_1 \cup \mathcal{A}_2|} \in \langle 0, 1 \rangle. \tag{1}$$

The set overlap similarity measure assumes that all words are equally important. It was shown in [7] that the similarity measure can be extended by a simple transformation to a weighted set overlap. Let $d_w \geq 0$ be an importance of a visual word X_w. The weighted set overlap of two sets $\mathcal{A}_1$ and $\mathcal{A}_2$ is

$$\mathrm{ovr}(\mathcal{A}_1, \mathcal{A}_2) = \frac{\sum_{X_w \in \mathcal{A}_1 \cap \mathcal{A}_2} d_w}{\sum_{X_w \in \mathcal{A}_1 \cup \mathcal{A}_2} d_w} \in \langle 0, 1 \rangle. \tag{2}$$

The weights d_w can be arbitrary non-negative numbers, one possible choice, inspired by text retrieval, is the tf-idf (term frequency – inverse document frequency) weighting scheme [1]. The weighted similarity overlap (2) has been shown to give better results than the plain set overlap (1). In the sequel, when speaking about set or word overlap, we always refer to the weighted version.

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

$$P\{f(\mathcal{A}_1) = f(\mathcal{A}_2)\} = \mathrm{ovr}(\mathcal{A}_1, \mathcal{A}_2).$$

To estimate the word overlap of two images, multiple independent min-Hash functions f_i are used. The fraction of the min-Hash functions that assigns an identical value to the two sets gives an unbiased estimate of the similarity of the two images. To efficiently retrieve images with high similarity, the values of min-Hash functions f_i are grouped into s-tuples called sketches. Similar images have many values of the min-Hash function in common and hence have high probability of having the same sketches. On the other hand, dissimilar images have low chance of forming an identical sketch. Identical sketches are efficiently found by hashing.

The recall of min-Hash is increased by repeating the random selection of s-tuples k times. A pair of images is a potential match when at least one sketch collision is encountered. The potential matches are typically further verified. The probability of a pair of images having at least one sketch out of k in common is a function of the word overlap

$$P\{\text{collision}\} = 1 - (1 - \mathrm{ovr}(\mathcal{A}_1, \mathcal{A}_2)^s)^k. \tag{3}$$

Randomized clustering with min-Hashing. We briefly review the approach of [5] to large scale image clustering. The clustering is based on min-Hash algorithm, and similar steps are also used in the application Section 4.

Typical values of the probability in (3) are close to one for near duplicate images, (very) close to zero for unrelated images, and 3 – 10% for images depicting the same object. Hence, min-Hash is not suitable for image retrieval (with the exception of near duplicates), because of the low recall. A cluster of images is defined as a set of images with related content. In such a cluster, there are many image pairs depicting the same object. Since each such a pair has certain (even as low as 3%) probability of retrieval, the probability that not a single image pair is retrieved from a cluster quickly drops with the size of the cluster. Retrieved image pairs are called cluster seeds. Clusters are then found by image retrieval as a connected component of related images.

3. Geometric min-Hash

In min-Hash, images are treated as sets of visual words and represented as ordered s-tuples of visual words. We observed that for large visual vocabularies, for most features there is a one-to-one mapping from visual word to a feature. In other words, that there is usually at most one feature in an image that is assigned to a particular visual word. This observation does not hold true for textures and repeated structures. However, statistically the observation holds: in the 100k database used in the paper, the average number of features per image is 2233.1 while the average number of features with unique visual word is 2131.9, *i.e.* more than 95%.

In min-Hash, the sketch is simply created as a ordered s-tuple of independent min-Hashes. In the proposed Geo-

1. Select a set F of image features that have a unique visual word within the image and have at least $v = 3$ features of similar scale in their neighbourhood. The set F is used for all sketches.
2. Use a random min-Hash function to select a central feature (= visual word) from F.
3. Find a set of features N that are (a) no closer than $d_{\min}$ and no further than $d_{\max}$ from the central feature, (b) the relative scale change to the central feature is not smaller than $c_{\min}$ and not larger than $c_{\max}$, and (c) no other feature with the same visual word satisfies the first two conditions.
4. Select secondary feature(s) using independent random min-Hash function(s) from features in N.

Table 1. The GmH sketch construction algorithm.

metric min-Hash, the sketch is divided into two parts: central feature and secondary features. The central feature is selected as in standard min-Hash, the difference is that only visual words with unique mapping to features in the image are considered. Using the unique mapping of visual words to a feature in the image, we obtain also spatial and scale information in the image. We claim, theoretically justify, and experimentally verify that using this additional information to guide the selection of the secondary feature(s) significantly improves the efficiency of the hashing procedure.

The algorithm of sketch generation is summarized in Table 1. For an example, look at Fig. 2. Yellow ellipses denote the central features, highlighted region around them shows the neighbourhood. In the close-ups, blue ellipses denote features considered for secondary feature, red ellipses features inside the neighbourhood violating the scale constraint.

There are four parameters involved in the procedure. Two parameters $d_{\min}$ and $d_{\max}$ governing the minimal and maximal distance of the secondary feature(s) from the central feature. To preserve affine invariant selection, the distance is measured as a Mahalanobis distance using the covariance matrix of the central feature. The values of the parameters are set to $d_{\min} = 0$ and $d_{\max} = 3$ in our experiments. The other two parameters $c_{\min}$ and $c_{\max}$ give minimal and maximal relative scale change of the secondary feature with respect to the central one. In our experiments, we set $c_{\max} = 1/c_{\min} = \sqrt{2}$.

3.1. Modeling the properties of GmH

In indexing and retrieval problems, the true positive and false positive rates are the key characteristics of a method. We first compare GmH and min-Hash in terms of their true positive rates. Then we discuss the false positive rates, *i.e.* the probability of retrieving incorrect matches by the two methods. The analysis assumes that the one-to-one mapping between features and visual words approximately holds, which is the case for the database of real images downloaded from web [10] and 1 million visual word vocabulary. Further terms considering textures and repeated structures can be injected into the analysis, we omit that for the sake of readability.

Recall of min-Hash and GmH. The success rate of min-Hash sketch is given by the probability P_{m} of a sketch collision. For better understanding, we decompose the probability into a product of visual r and geometric g terms

$$P_{\mathrm{m}} = \mathrm{ovr}(\mathcal{A}_1, \mathcal{A}_2)^s = r^s(\xi) g^s(\xi). \tag{4}$$

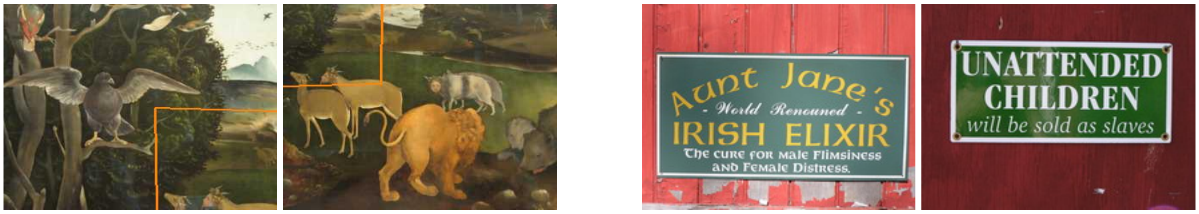
The vector parameter ξ encodes the acquisition conditions (including view point, illumination, compression level, etc.), s is the sketch size, $r(\xi)$ is the word overlap of standard min-Hash computed on the parts of the scene visible in both images, $g(\xi)$ is the fraction of features inside the region visible in both images. Some components of ξ affect r and g differently, as discussed later in the section. Note that such a decomposition is neither known nor required in the algorithm, it is used solely in the analysis. Similarly, the probability P_{g} of sketch collision of GmH is

$$P_{\mathrm{g}} = r(\xi) g(\xi) \rho^{s-1}(\xi) \gamma^{s-1}(\xi), \tag{5}$$

where $\rho(\xi)$ is overlap of words in the scale-spatial neighbourhood selected by GmH given the first min-Hash matched and $\gamma(\xi)$ is the fraction of features inside the neighbourhood that are also inside the geometric overlap of the images.

Equations (4) and (5) decompose the probability of a true hit into factors dependent on appearance change (r, ρ) and effects of visibility (g, γ), since their influence on GmH and standard min-Hash differs. There is a common factor $r(\xi)g(\xi)$ to both probabilities P_{m} and P_{g}. This factor represents the probability that the first (central) min-Hash is correct. Both methods fail if the central min-Hash fails, and this happens with equal probability. The difference between the two methods lies in the case when the central min-Hash is matching. For a min-Hash sketch, the words are drawn s times independently. The hit probability is thus reduced by $g(\xi)$ for each min-Hash. For GmH, the first word is selected as in min-Hash. However, the other $s - 1$ min-Hashes are selected from a scale-spatial neighbourhood. If the central feature is matching, then it lies inside the geometric overlap of the images. Under assumption of spatially localized occlusion or change of background, the probability $\gamma(\xi)$ that a feature in the neighbourhood of the central feature is inside the geometric overlap is high, often close to one, certainly $\gamma(\xi) > g(\xi)$.

Probabilities $r(\xi), \rho(\xi)$ express repeatability of visual words, taken at random from the overlapping part of the images. Both probabilities are similarly affected by the change of viewpoint and illumination conditions. However,



GmH 37 hits / min-Hash 4 hits

GmH 0 hits / min-Hash 4 hits

Figure 3. Comparison of Geometric min-Hash and min-Hash. An image pair with small spatial overlap (left) and a pair of unrelated images (right). The region of overlap is marked in the left pair. The numbers of sketch collisions (from 5000 independent sketches) for GmH/ min-Hash are shown under the image pairs. GmH has almost 10 times higher probability of retrieving the related image pair than min-Hash, and significantly lower probability of retrieving un unrelated pair. Note that the probability of min-Hash retrieving either of the two image pairs is the same.

the change in scale degrades r more than ρ, as feature detector as well as descriptor repeatability is significantly affected by scale [17]. Given that the central feature survived the scale change, features of similar scale are more likely to survive too than other (significantly smaller or larger) features.

We demonstrate the difference in the two approaches on two examples: scale change Fig. 2 and camera panning Fig. 3 (left). Consider the image pair in Fig. 2. Here, the word overlap $r = 0.0399$, while the geometric overlap is $g = 0.6391$. Combined, the probability of sketch of $s = 2$ min-Hashes is $P_{\mathrm{m}} = r^2 g^2 = 0.0255^2$. We have also computed, by exhaustive enumeration of neighbourhoods of all matching features, the product of $\rho\gamma = 0.1751$. Hence the probability $P_{\mathrm{g}} = rg\rho\gamma$ is 6.9 times higher than the probability of P_{m}. In a simulated experiment, 5000 GmH and min-Hash sketches were generated. For this image pair, 3 min-Hash collisions were encountered (expectation is 3.3) and 19 GmH collisions (expectation 22.3).

Camera panning Fig. 3 (left): for this image pair, $r = 0.1768$ and $g = 0.1736$, which gives $P_{\mathrm{m}} = 0.0307^2$ the probability of min-Hash success. For GmH, we measured $\rho\gamma = 0.2403$, which means 7.8 times higher success rate for GmH. In the simulated sketch generation, 4 min-Hash collisions were observed (expectation 4.7) and 37 GmH collisions (expectation 36.9).

Probability of random collisions. In the following analysis we show that, rather surprisingly, GmH has lower false positive rates than min-Hash, since it selects a hash from a smaller set. For simplicity, we will demonstrate the property on standard set overlap. To understand the effect, assume two unrelated images represented by two different sets of random visual words. Let $\mathcal{N}$ and $\mathcal{M}$ be random subsets of visual words from a vocabulary of size w, with $n = |\mathcal{N}|$ and $m = |\mathcal{M}|$. To estimate the mean overlap, we fix the set $\mathcal{N}$ and average the overlap over all possible sets $\mathcal{M}$ of size m. The probability that a visual word e is a member of random $\mathcal{M}$ is $P(e \in \mathcal{M}) = m/w$. Now, assuming $n, m \ll w$, the distribution of $|\mathcal{M} \cap \mathcal{N}|$ can be closely approximated by Binomial distribution $\mathrm{Bi}(n, m/w)$. We can also bound the size of the union by $\max(m, n) \leq |\mathcal{N} \cup \mathcal{M}| \leq 2\max(m, n)$. A lower bound on the expected overlap is given by $\mathbf{E}(|\mathcal{M} \cap \mathcal{N}|/|\mathcal{M} \cup \mathcal{N}|) \geq \mathbf{E}(|\mathcal{M} \cap \mathcal{N}|)/(2\max(m, n)) = nm/2\max(m, n)w = \min(m, n)/2w$. In the derivation, we have used $\mathbf{E}(\mathrm{Bi}(n, m/w)) = mn/w$ and $mn/\max(m, n) = \min(m, n)$. Similarly, an upper bound is given by $\min(m, n)/w$. Finally

$$\frac{\min(m,n)}{2w} \leq \mathbf{E}\left(\mathrm{ovr}_1(\mathcal{M}, \mathcal{N})\right) \leq \frac{\min(m,n)}{w}. \qquad (6)$$

The derivation shows that the expected value of overlap of the two sets, *i.e.* $\mathbf{E}(\mathrm{ovr}_1(\mathcal{N}, \mathcal{M}))$ is well approximated by a linear function of the size of the smaller set. Therefore, drawing secondary min-Hash(es) from a restricted neighbourhood (smaller set) has lower chance of a random collision than drawing from the whole image (larger set).

Lower probability of a random collision is demonstrated in Fig. 3, right image pair. The two images contain similar features (described by identical visual words) and hence min-Hash generates a number of sketch collisions. However, none of the features appearing in both images has any support in its neighbourhood – no Geometric min-Hash sketch collision are encountered.

A similar property has been empirically observed and used in a different problem of correspondence matching [23] and object retrieval [25]. The following semi-local tentative correspondence filtering was proposed in [23]. For a correspondence, look at nearby features. If there are no similar features in the neighbourhoods of the corresponding points, discard such a correspondence as an outlier. The procedure of selecting secondary feature(s) can be seen as an implicit randomized version of such a spatial verification test. The probability of a sketch collision depends on the set overlap of features in the neighbourhood of the central feature. The higher the overlap, the higher probability of a sketch collision. Therefore, a sketch collision guarantees with certain probability that there are identical features (at least one – the secondary min-Hash) in the neighbourhood of the central sketch feature.

3.2. Implementation details and discussion

Geometric invariants. Under the assumption of local near coplanarity of the observed surface, the mutual position of the central and secondary features provides additional affine geometric invariants. The mutual position can be either encoded directly into the hash or used to verify/reject the sketch collision. In our experiments we have used the latter.

While the standard min-Hash provides a hypothesis that two images may have similar content, the Geometric min-Hash also provides a hypothesis of image to image mapping. In practice, to perform the final global (semi-planar) spatial verification, it is sufficient to evaluate consensus for only a *single hypothesis*. The hypothesis is generated by the features participating in the colliding sketch.

Feature non-maxima suppression. Feature detectors perform non-maxima suppression to avoid multiple response at the same physical location. This procedure is always a compromise between having multiple feature detections and missing some important features.

Multiple feature detections can affect the performance of GmH, since features with non-unique visual word are not used. In fact, two detections of (almost) the same feature provide equally good localization in position and scale as a single detection and there is no need to remove such features from the GmH process. Therefore, we perform second round of non-maxima suppression. The suppression region is larger than in the stage of feature detection, however, only regions with identical visual word are now considered.

Textures and repeated structures. The concept of GmH can be extended to textures. Instead of looking at a neighbourhood of a single feature, neighbourhood of all features would be considered. This way, a semi-local textural descriptor can be generated. Such a descriptor, however, would provide only a weaker geometric localization. We do not address the min-Hash based textural descriptor in this work.

Note the difference between statistical texture and repeated structures: the discriminative features repeated a few times (*e.g.* two towers) are still used in GmH as secondary features. The secondary features are not required to be unique within the image, only within the neighbourhood of the central feature.

4. Applications

In this section, two applications of the GmH are presented. First, it is applied to large scale image clustering, where the method exceeds the state of the art results both in accuracy and in efficiency. Second, we extend the randomized clustering approach to (small) object discovery in large collections of images.

In the experiments, we use the 100k Oxford database of Flickr [9] images introduced in [21]. This dataset is a superset of the 5k Oxford dataset [10].

4.1. Image clustering

In image clustering the objective is to find image pairs with some (significant) visual content in common. Cluster is then defined as connected component of related images. Such clusters can be consequently used *e.g.* for geolocation or for automatic 3D reconstruction [26] of buildings, or even groups of buildings that are nearby and their mutual location is captured in the image collection.

We base our approach on the randomized clustering [5] reviewed in Section 2. The randomized clustering first detects seed image pairs via min-Hash collisions and then completes the search for connected components of the matching graph by image retrieval. Due to low recall and precision, the min-Hash clustering [5] generates $k = 512$ independent sketches consisting of $s = 3$ min-Hashes. For GmH, it was sufficient to set $k = 60$ and $s = 2$. Reduction of both k and s saves both memory and processing time.

The second improvement over the min-Hash clustering [5] involves interleaving cluster seed detection and connected component completion. Such approach has the advantage that there is no need to pre-generate a fixed number of min-Hashes. Instead, a few (or just a single) sketches are generated for each image. Sketch collisions are resolved and new cluster seeds are completed using image retrieval. The procedure is iterated as many times as necessary or as long as acceptable - in a real application this can be a background process discovering new small objects within the database in idle time. The clustering algorithm becomes an *anytime algorithm*, easy clusters (near duplicates and large overlaps, many images) are found early, “needles in the haystack” (small objects, just a few images) much later (on average).

Match verification. Since matches between relatively large fractions of images are considered in this task, an image pair is considered to be matching if there is sufficient (non-degenerate) support for a global geometry (at least 20 matches).

Results. In the Oxford 5k dataset, 11 landmarks were manually labelled. We measure the fraction of ground truth images of the same building that are assigned to the same cluster. We also measure the number of false positive (visually unrelated) images by visual inspection. Results are summarized in Table 2. We compare GmH results to [5] (last two columns). To give some insight, we show results of plain GmH with no image retrieval involved in column ‘seeds’. Partial results after a single sketch per image (equivalent to setting $k = 1$) followed by the connected component crawl are shown in the ‘iter1’ column; seven clusters are discovered after drawing a single geometric min-Hash.

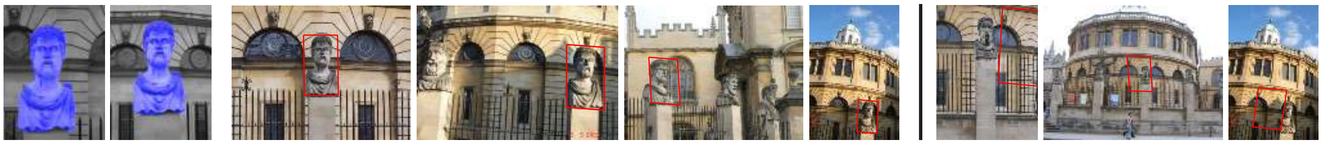


Figure 4. The role of co-segmentation (highlighted in blue on the two leftmost seed images). Correctly retrieved objects using features from the co-segmentation (middle group) and a 'leak' into the background when using bounding box (right group).

	seeds		iter 1		GmH		[5]	
	CR	fp	CR	fp	CR	fp	CR	fp
all souls	58.97	0	98.72	0	98.72	0	97.44	0
ashmolean	36.00	0	0.00	0	76.00	0	68.00	0
balliol	33.33	0	0.00	0	91.67	0	33.33	0
bodleian	91.67	0	100	0	100	0	95.83	1
christ ch	71.79	0	97.44	1	97.44	1	89.74	0
cornmarket	44.44	0	77.78	1	77.78	1	66.67	0
hertford	62.96	0	0.00	0	100	0	96.30	1
keble	71.43	0	0.00	0	100	0	85.71	0
magdalen	14.81	0	38.89	0	38.89	0	5.56	0
pitt rivers	100	0	0.00	0	100	0	100	0
radcliffe	97.29	0	99.55	0	99.55	0	98.64	0

Table 2. Image clustering results: 'CR' columns display recall for each cluster (percentage of ground truth images in it), 'fp' columns the number of its false positive images. Results are shown for connected components of seeds without applying completion by retrieval ('seeds'), a single sketch per image ('1 iter'), 60 sketches ('GmH'), and results from [5] for comparison.

The most noticeable result is for the 'Magdalen' landmark. It consists of images of a tower photographed from four different sides. GmH, even without the retrieval part, has achieved better results that [5]. This is due to significantly better recall of GmH– a larger cluster (different side of the tower) with less similar images was discovered.

Finally, we report the running time to achieve the results (measured on a 3GHz Linux machine). The first iteration takes 5min 53sec and the 60 iterations of the full clustering process take 15min 16sec. Interestingly, the first iteration takes more than 30% of the time required by 60 iterations. This is caused by the fact that the largest number of images is touched in the first iteration (most of the near duplicates and all very large clusters). The total clustering time per image in the database is 0.008 seconds.

The speed of the proposed method benefits from two important properties of the GmH: (i) the decrease in false positive rate and thus a lower number of unnecessary verification, and (ii) faster verification since image-to-image transforation is proposed together with a seed image pair.

4.2. Discovering small objects

Unlike in the previous application, the focus is paid to small objects that repeat in different images in an unordered large collection. A similar task was addressed by [25], where small objects were discovered in feature films, represented by approximately 3000 keyframes. We highlight the main differences to our approach: the algorithm in [25] is quadratic [5] and hence not scalable to very large databases; in our approach, the scale of the object can vary in different images, while in [25] it has to be roughly the same.

1. Generate object seeds (hypotheses) using GmH. Keep only those with support of at least $n = 5$ correspondences.
2. Apply co-segmentation in the manner of [3] using the matching features, reject the hypothesis if the co-segmentation fails.
3. Use features inside the co-segmentation to retrieve further exemplars of the object.

Table 3. Small object discovery algorithm using GmH (outline).

The approach is fairly similar to the randomized image clustering described above. Instead of trying to connect all related images and images related to those, the task is to focus on a single small object. In order to do that, extra attention has to be paid to the match verification as well as the retrieval part of the algorithm.

Match verification. Unlike in the image clustering task, neither sufficient support, nor non-degeneracy are useful for verification of a small object hypothesis. The objects to be discovered are sharing as few as five features. Also, since the objects are small, from the perspective of the whole image, the location of features often looks degenerate, collapsed to a single point. The verification thus exploits images at pixel level, and the hypotheses are ascertained using co-segmentation [3]. The co-segmentation not only helps to verify the correctness of the match, but a successful verification provides a segmentation of the object at pixel level, obtaining a precise model of the object. In the retrieval part, namely query expansion [6], only features that fall inside the co-segmented region are used to query for further examples of the object. The situation is depicted in Fig.4.

Results and discussion. The focus of the experiments was on small objects. We show some of the detected objects that have between 5 to 40 matching features. Figs. 1, 5, and 6 show some of the detected objects. In Figs.1 and 5, the seed sketches are shown in large images, with the three stages shown: the sketch, matching features, and the co-segmentation.

Note that, unlike other clusters, the face cluster in Fig. 5 is not a rigid object, but rather an object category. The seed was generated on two different, but sufficiently similar faces. Since the query expansion [6] builds a simple generative model, it is possible to discover many instances (1000 faces were discovered, where 1000 was the cut-off

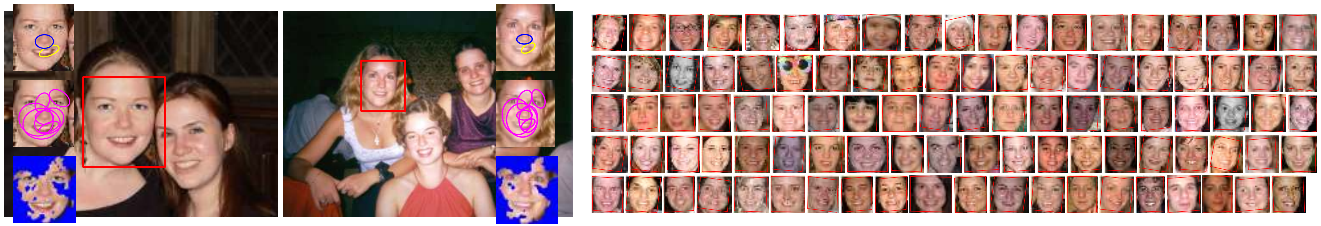


Figure 5. An automatically discovered face cluster. Two similar faces of different people have been found as a seed by the GmH (left two images). The sketch features, supporting features and co-segmentation are shown in close-up on the image sides. Sample of further detections from the cluster are shown on the right.

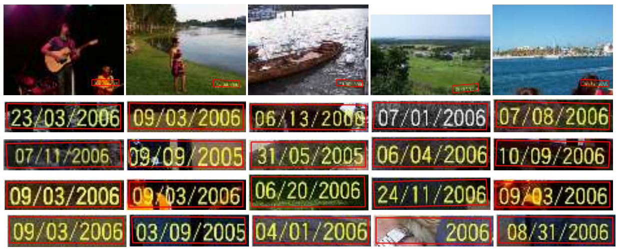


Figure 6. The discovered visual entities that repeat over the database are not always meaningful objects, such as in the 'date' cluster shown.

threshold) of this category. We do not claim, that the proposed method is directly applicable to categorization tasks (certainly not with the current features).

Failure cases. Unfortunately, not all discovered subimages represent true objects, as in Fig.6. Images of text are in general difficult to deal with in retrieval, and our approach is not completely overcoming the problem. A number of clusters similar to Fig. 6 has been discovered.

5. Conclusions

Geometric min-Hashing is an efficient hashing scheme that combines visual appearance and geometric interaction of image features. The high indexing ability – high recall and low false positive rate – makes it an powerful tool for a number of applications, such as image retrieval, large scale image clustering, and automatic small object discovery in large collections of images.

References

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, ISBN: 020139829, 1999.

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

[3] J. Cech, J. Matas, and M. Perdoch. Efficient sequential correspondence selection by cosegmentation. In *Proc. CVPR*, 2008.

[4] O. Chum and J. Matas. Geometric hashing with local affine frames. In *Proc. CVPR*, 2006.

[5] O. Chum and J. Matas. Web scale image clustering. Technical Report 15, CMP, CTU in Prague, May 2008.

[6] 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 *Proc. ICCV*, 2007.

[7] O. Chum, J. Philbin, and A. Zisserman. Near duplicate image detection: min-hash and tf-idf weighting. In *Proc. BMVC.*, 2008.

[8] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *Proc. ICCV*, 2005.

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

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

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

[12] P. Jain, B. Kulis, and K. Grauman. Fast image search for learned metrics. In *Proc. CVPR*, 2008.

[13] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Proc. ECCV*, 2008.

[14] Y. Lamdan and H. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *Proc. ICCV*, pages 238 – 249, 1988.

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

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

[17] 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.

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

[19] A. Oliva and A. Torralba. Building the gist of a scene: The role of global image features in recognition. *Visual Perception, Progress in Brain Research*, 155, 2006.

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

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

[22] T. Quack, V. Ferrari, and L. Van Gool. Video mining with frequent itemset configurations. In *Proc. CIVR*, 2006.

[23] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *PAMI*, 19(5):530–535, May 1997.

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

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

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

[27] A. Torralba, R. Fergus, and Y. Weiss. Small codes and large databases for recognition. In *Proc. CVPR*, 2008.

The authors were supported by Czech Science Foundation Project 102/09/P423 and by EC project ICT-215078 DIPLECS.