A Novel Approach to Retrieval of Similar Patterns in Biological Images

Andrzej Sluzek
Khalifa University, Abu Dhabi

Abstract. Novel descriptors of keypoints are proposed for matching (primarily) biological images. The descriptors incorporate characteristics of limited-size neighborhoods of keypoints. Descriptors are quantized into small vocabularies representing photometry of images (SIFT words) and geometry of their neighborhoods, so that significant distortions can be tolerated. In order to keep precision at a high level, Harris-Affine and Hessian-Affine detectors are independently applied. The retrieval results are accepted only if confirmed by both techniques. Using several test datasets, we preliminarily show that the method can retrieve semantically meaningful data from unknown and unpredictable images without any training or supervision. Low computational complexity of the method makes it a good candidate for scalable analysis of biological (e.g. zoological or botanical) visual databases.

1 Introduction

Visual inspection of random collections of images is one of the most tedious tasks in biological and biomedical practice. Algorithms performing automatic analysis of images are continuously developed for selected applications, where the objective is to identify well-defined objects/phenomena, e.g. cell counting, MRI segmentation, etc. However, in case of vaguely defined problems (e.g. searching for similarly looking components in images of plants, detection for similar phenomena in microscopic images of bacteria, etc.) images might be too diversified or too unpredictable for such specialized algorithms. The tasks are particularly difficult if the correspondence between semantics of image contents and their pictorial representations is not straightforward.

A prospective solution for the above challenges is to apply data mining, which in the visual domain is often referred to as content-based visual information retrieval (CBVIR).

Reported CBVIR applications generally focus on three areas, i.e.

i. Retrieval of near-duplicate images (including partial near-duplicates, e.g. the same objects).

ii. Retrieval of (semantically) the same category objects.

iii. Retrieval of (semantically) the same category scene.

In (i), the expected result is a collection of image pairs depicting physically identical scenes or objects (subject to deformations caused by the viewpoint change,
camera quality, visibility conditions, partial occlusions, etc.). No training or supervision (regarding image semantics) is usually required. In (ii) and (iii), however, training is indispensable to generalize visual characteristics of semantic categories/objects, because individual images from the same category may be (visually) very different. In all the above areas, keypoint-based approaches are one the most fundamental tool (in particular visual words, BoW, and other word-based techniques, e.g. semantic topics built upon visual words).

Unfortunately, many types of biological images do not fit any of these areas. First, biological objects (even if considered identical) usually have diversified forms (and, subsequently, their images are only approximately near-duplicate). Secondly, in many problems the numbers and types of the semantic categories are not specified (e.g. the contents of images might not be fully classified).

The objective of this paper is to partially fill the gap between (i) and (ii)/(iii). We propose a scheme (based on the general concept of keypoint description and matching) to preliminarily identify images (e.g. biological images) which may contain meaningfully similar contents in spite of a wide range of deformations within these contents. No training (or pre-existing knowledge about the images) is assumed so that the scheme can quickly switch from on application to another. Moreover, the scheme is computationally efficient; we use a novel affine-invariant keypoint description where individual matches indicate the presence of visually similar fragments. The main difference, compared to typical descriptors like SIFT or SURF, is that our descriptions represent not only keypoints but also their limited-size neighborhoods.

The method is proposed as a tool for preliminary search and analysis in collections of (primarily) biological images with random and unknown contents (although the general scope of these collections should be known for easier interpretation of the results). As proof-of-concept examples, we use three small (but diversified) datasets containing images of butterflies, leaves and viruses.

In Section 2 of the paper we overview pre-existing tools and techniques contributing to the proposed method. Its description is provided in Section 3. Results for the selected datasets are presented in Section 4, while Section 5 summarizes the paper.

2 Typical Keypoint-Based Tools and Techniques

Affine-invariant keypoint detectors are the low-level tools of the method. Two popular detectors, i.e. Harris-Affine and Hessian-Affine, [8], are used because of their mutually supplementing characteristics. Harris-Affine highlights corner-like saliences, while Hessian-Affine returns saliences corresponding to blobs.

Keypoints are represented by SIFT descriptor, [7], quantized into visual words. A relatively small vocabulary $SV$ of 2000 words is used. Such a vocabulary is able to accept significant photometric distortions of image contents, but the discriminative power of individual words is very low (e.g. comments in [10], [13]). Nevertheless, as shown in Section 3, we actually use a 3D Cartesian product of vocabularies so that the practical resolution of descriptions is much higher.
In general, keypoints can be matched using either O2O (typically *mutual nearest neighbor*) or M2M (typically *the same visual word*) schemes. However, regardless the scheme (and regardless the vocabulary size) none of the schemes based on individual keypoint correspondences can reliably distinguish between actually similar image fragments and random locally similar contents. Fig. 1 shows examples of matching in a pair of images sharing the same object and in an unrelated pair. For all schemes, there is no qualitative difference between the results for both pairs.

![Fig. 1. Exemplary keypoint matching results using: (A) O2O scheme over Harris-Affine keypoints, (B) M2M (2000 words) over Hessian-Affine keypoints, (C) M2M (2^{20} words) over Harris-Affine keypoints](image)

Partial near-duplicates are usually detected by the verification of configuration constraints for unspecified groups of preliminarily matched keypoints (e.g. [1, [3], [9], [14]]). However, this is a computation-intensive operation which is not fully scalable to large databases. In particular, if numerous matches are found using a small vocabulary (e.g. Fig. 1B) the configuration verification can be prohibitively costly. Moreover, it is generally believed (e.g. [13]) that small vocabularies excessively reduce precision of retrieved results.

Unfortunately, in biological images we need both small vocabularies and relaxed configuration constraints (because of highly diversified appearances of nominally the same objects – e.g. individual butterflies of the same species). Thus, the available methods and techniques are generally unable to handle such flexibility, unless they are trained using sufficiently diversified positive and (sometimes) negative examples to classify/recognize images or objects of predefined categories, e.g. [2], [5], [15].
3 Principles of the Method

Our objective is to develop a method which can overcome the difficulties highlighted in Section 2 for a wide range of biological images of diversified contents. In particular, our intentions are:

1. To bypass the configuration analysis for preliminarily matched keypoints by incorporating affine-invariant descriptions of keypoint neighborhood geometry into descriptors of the keypoints themselves. Descriptors of geometry are quantized into another small-size vocabulary so that significant geometric distortions can be tolerated.

2. To improve precision of image retrieval (when small vocabularies are used) by combining independently obtained results for Harris-Affine and Hessian-Affine keypoints. Only images retrieved in both operations are accepted (subject to additional details discussed below).

3.1 Extended Description of Keypoints

To represent geometry of keypoint neighborhoods, we propose to use a modification of the method discussed in [12]. First, we build limited-size neighborhoods of extracted keypoints. The neighborhoods consist of a limited number (not more than 20) keypoints of similar sizes (between 50% and 150% of the area of the central keypoint) and within a limited distance (between 70% and 200% of the Mahalanobis distance defined by the ellipse of the central keypoint) from the keypoint of interest; see Fig. 2A for illustration.

Fig. 2. A keypoint and its neighbors (A); the size of ellipses for two neighbors is shown as an illustration. Trapezoids built within a triplet of elliptic keypoints (B).

Subsequently, triplets of the keypoints are formed and their ellipses are represented by trapezoids shown in Fig. 2B. Then, the least complex affine-invariant moment expression

\[ \text{Inv} = \frac{M_{20} M_{02} - M_{11}^2}{M_{00}^3}. \]  

is used to characterize each of the trapezoids.
The range of *Inv* invariant is actually quantized into 12 values (words) so that geometry of the whole triplet of ellipses is described by a small vocabulary *GV* of $12^3 = 1728$ words. Altogether, a triplet of keypoints can be represented by SIFT words from a vocabulary of 2000 words (visual properties of individual keypoints) and one word from 1728 words of *GV* vocabulary (geometry of the whole triplet).

Given a keypoint *K* and its neighbors $\{K_1, K_2, ..., K_N\}$, a number of $\{K, K_i, K_j\}$ triplets can be built around *K* as explained above. Our experiments show that the most typical numbers of such triplets are between 50 to 80 only (because we exclude triplets for which the triangles are too narrow). There are, nevertheless, some keypoints around which no triplets are found (e.g. large keypoints surrounded by very small keypoints only).

Altogether any keypoint *K* and its neighborhood are represented by the SIFT word $SV(K)$ and a set of 3-word phrases $SoP(K)$:

$$SoP(K) = \bigcup \{SV(K_i), SV(K_j), GV(K, K_i, K_j)\}$$

from the Cartesian product $SV \times SV \times GV$.

### 3.2 Matching Keypoints and Images

Using descriptions proposed in Section 2.1, matching keypoints (i.e. actually matching their neighborhoods as well) is straightforward. Two keypoints *K* and *L* match if

$$SV(K) = SV(L) \quad \text{and} \quad SoP(K) \cap SoP(L) \neq \emptyset$$

i.e. the keypoints are visually similar and their neighborhoods are similar (at least partially) both visually and geometrically.

Compared to the correspondences obtained by matching only the keypoints themselves, i.e. $SV(K) = SV(L)$ (see the second row of Fig. 1) the results are more meaningful. The results in Fig. 3 show relatively few correspondences (and many of them are correct in the global context). Note that no verification of configuration constraints is needed; the results are obtained by simple keypoint matching.

Nevertheless, some correspondences are between locations which are semantically rather different. This is primarily because small *SV* and *GV* vocabularies can tolerate significant photometric and geometric distortions.

Thus, as the final verification step, we combine Harris-Affine and Hessian-Affine correspondences. A pair of matched Harris-Affine keypoints ($K_{HA}, L_{HA}$) is accepted only if there is another pair of Hessian-Affine keypoint ($K_{HE}, L_{HE}$) overlapping it (and another way around). Formally, “overlapping” means that the Mahalanobis distances (defined by the shapes of keypoint ellipses) between the corresponding keypoints are below the threshold. Fig. 4 provides an illustrative example (using 140% of the corresponding Mahalanobis unit distances as the threshold; this value is used in the paper).

Eventually, a pair of images is retrieved (i.e. is it assumed that both images contain partial near-duplicate of noticeable size) if at least one pair of keypoint correspondences is retained after the final verification. The coordinates of those keypoints indicate the approximate locations of the partial near-duplicates in the matched images.
Fig. 3. Keypoint correspondences obtained by the proposed method for Harris-Affine keypoints (A) and Hessian-Affine keypoints (B). The same images are used in Fig. 1.

Fig. 4. A pair of matched Harris-Affine keypoints (blue) and Hessian-Affine keypoints (red) in close proximity.

Fig. 5. Keypoint correspondences after the final verification. The same images are used in Figs 1 and 3.

It should be noted that each pair of keypoint correspondences additionally indicates an unspecified number of matches between keypoints from both neighborhoods.

The results after the final verification for exemplary images from Figs 1 and 3 are given in Fig. 5. The same roadsign is correctly detected as a partial near-duplicate in Fig. 5A. However, one of the keyboard keys is also sufficiently similar (because of small vocabularies) to the roadsign and recognized as a partial near-duplicate as well (Fig. 5B). This can be considered a disadvantage, but in biological or biomedical images such relatively weak similarities should be usually retrieved as potential candidates for semantically meaningful partial near-duplicates.
4 Experiments

The proposed approach is currently being tested on diversified databases of biological and biomedical images. The objective is to evaluate the practical significance of the results (relevance to the human interpretation of the images, retrieval of unspecified “hidden” semantics, etc.).

Three small-scale examples are discussed in this paper. In all examples only grey-level images are used (so that the visual analysis is more difficult).

4.1 Images of Butterflies

A part of the Ponce Research Group butterfly dataset\(^1\) has been used. The same dataset was analyzed in [6] with similar concepts of keypoint matching. However, both training and geometric consistency verification were used. Other reported works on automatic recognition of butterflies (e.g. [4], [11]) also heavily rely on the specific properties of butterfly images.

The presented method achieves comparable results without any preliminary knowledge about the image domain and by using only individual keypoint matches (note that complexity of the final verification illustrated in Fig. 4 is negligible).

The total number of matched image pairs is 7,140 of which 1,140 are the semantic ground truth (i.e. showing butterflies of the same species). Our method retrieved 1,578 image pairs from which:

- 743 pairs contain butterflies of the same species (examples in Fig. 6A);
- 696 pairs show butterflies with similar fragments of wings (see Fig. 6B);
- 47 pairs indicate similarities between background plants (examples in Fig. 6C);
- 87 pairs indicate similarities between butterflies and the background plants (good camouflage?);

and

- 5 pairs show random similar fragment.

Altogether, recall of the method (in retrieval of image pairs containing butterflies of the same species) is 65.2%, while precision is nominally 47.1%. However the actual precision is 94.2% because both similarly looking wing fragments and similarly looking background plants are also semantically correct partial near-duplicates (especially because the domain of images is not specified and we should not distinguish between partial near-duplicities within different types of objects).

Additionally, we have built similarity graphs between the dataset images. It was found that its \(K\)-connected sub-graphs consist of nodes representing images of the same species butterflies. We do not discuss details of this issue in the paper. However, it is again mentioned in the following Section 4.2.

\(^1\) http://www-cvr.ai.uiuc.edu/ponce_grp/data/
Fig. 6. Exemplary matches between the same species butterflies (A), between similar fragments of butterfly wings (B) and between plants in the image backgrounds (C).

4.2 Other Images

The other two datasets contain images of plant leaves\textsuperscript{2} and viruses (low-resolution images collected from the web).

The leaf images are of rather low quality (see examples in Fig. 7). The total number of image pairs to be matched (images showing only contours of leaves were deleted) is 2,775. However, the ground truth is not available (even though preliminary grouping is provided).

The algorithm has retrieved 83 pairs of images containing partial near-duplicates. Surprisingly, the similarity graph of the matched images is strongly connected (the similarity graph is 5-connected) so that the retrieved pairs of images – with only 14 images contributing to these pairs - can prospectively define some properties of the retrieved images. Seven of those images are given in Fig. 8, and it can be clearly seen that all of them have jagged (at least partially) borders. Thus, we can conclude that this is the (semantic) property of leaves automatically identified by the algorithm.

\textsuperscript{2}http://www.imageprocessingplace.com/root_files_V3/image_databases.htm
Fig. 7. Exemplary images from the dataset of leaves

Fig. 8. Seven examples of leaves with (partially) jagged borders

Fig. 9. Exemplary images of viruses (A) and two pairs of images retrieved by the system (B)

In the third test, 528 pairs of images (some are given in Fig. 9A) showing viruses of various diseases have been matched. Eventually, partial near-duplicates have been found in only two pairs of images (Lassa fever virus sharing partial near-duplicates with rubella virus, and SARS virus with West Nile fever virus) – see Fig. 9B.

In general the numbers of images in these two datasets are too small to reliably evaluate performances (e.g. recall and precision) especially because the ground truth classification is missing for some images.

Nevertheless, the visual inspection of the results confirms that the retrieved visual partial near-duplicates actually have some semantic significance.

5 Summary

We show that very simple and general tools, i.e. (1) keypoint matching using small vocabularies representing visual and geometric properties of keypoints and their neighborhoods and (2) superposition of results obtained for Harris-Affine and Hessain-Affine keypoints, can be often used to retrieve meaningful similarities
(approximated by partial near-duplicates) from visual databases semantically. No training or domain-specific approaches are required.

Because of small vocabularies used, the method can tolerate significant photometric and geometric distortions so that it is particularly suitable for processing biological/biomedical images.

The presented experiments preliminary confirm the practicality of the method. Currently, experiments are conducted on diversified and much larger databases of biomedical images.

References

1. Chum, O., Perdoch, M., Matas, J.: Geometric min-hashing: Finding a (thick) needle in a haystack. In: Proc. IEEE Conf. CVPR 2009, pp. 17–24 (2009)
2. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: Proc. IEEE Conf. CVPR 2003, vol. 2, pp. 264–271 (2003)
3. Jegou, H., Douze, M., Schmid, C.: Improving bag-of-features for large scale image search. Int. J. Comp. Vision 87, 316–336 (2010)
4. Kang, S.-H., Jeon, W., Lee, S.-H.: Butterfly species identification by branch length similarity entropy. J. Asia-Pacific Entomology 15, 437–441 (2012)
5. Lampert, C., Nickisch, H., Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer. In: Proc. IEEE Conf. CVPR 2009, pp. 951–958 (2009)
6. Lazebnik, S., Schmid, C., Ponce, J.: Semi-local affine parts for object recognition. In: Proc. BMVC 2004, pp. 779–788 (2004)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Com. Vision 60, 91–110 (2004)
8. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. Int. J. Comp. Vision 60, 63–86 (2004)
9. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: Proc. IEEE Conf. CVPR 2007, pp. 1–8 (2007)
10. Romberg, S., August, M., Ries, C.X., Lienhart, R.: Robust feature bundling. In: Lin, W., Xu, D., Ho, A., Wu, J., He, Y., Cai, J., Kankanhalli, M., Sun, M.-T. (eds.) PCM 2012. LNCS, vol. 7674, pp. 45–56. Springer, Heidelberg (2012)
11. Silveira, M., Monteiro, A.: Automatic recognition and measurement of butterfly eyespot patterns. Biosystems 95, 130–136 (2009)
12. Śluzek, A.: Large vocabularies for keypoint-based representation and matching of image patches. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 229–238. Springer, Heidelberg (2012)
13. Stewénius, H., Gunderson, S.H., Pilet, J.: Size matters: Exhaustive geometric verification for image retrieval accepted for ECCV 2012. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 674–687. Springer, Heidelberg (2012)
14. Wu, Z., Ke, Q., Isard, M., Sun, J.: Bundling features for large scale partial-duplicate web image search. In: Proc. IEEE Conf. CVPR 2009, pp. 25–32 (2009)
15. Yang, K., Wang, M., Hua, X.-S., Yan, S., Zhang, H.-J.: Assemble new object detector with few examples. IEEE Im. Proc. 20, 3341–3349 (2011)