## Scalable Visual Instance Mining with Instance Graph

Wei Li<sup>1</sup> liwei2658@gmail.com Changhu Wang<sup>2</sup> chw@microsoft.com Lei Zhang<sup>3</sup> leizhang@microsoft.com Yong Rui<sup>2</sup> yongrui@microsoft.com Bo Zhang<sup>1</sup>

dcszb@mail.tsinghua.edu.cn



Figure 1: The proposed framework for visual instance mining.

Visual instances are the basic re-occurring information in the visual world. In this paper we address the problem of visual instance mining. Specifically, the goal is to automatically discover frequent visual instances from a collection of images. We are interested in mining specific instances, for example, Air Force One and Monarch Butterfly, which are different from high-level visual categories like plane and butterfly. This task plays a fundamental role in many applications such as multimedia summarization [7] and image annotation [6].

The problem of visual instance mining brings two challenges. The first one is the mining of small instances that have large variations. In many cases, visual instances may only cover very limited image areas, which sometimes can hardly be noticed even by humans. Different types of variations, such as scale, rotation, and occlusion, especially the variations resulted from out-of-plane rotation and non-rigid objects, require highly robust algorithms. The second challenge is the scale of the dataset. A large-scale database is essential for discovering practically useful visual instances. However, large-scale databases usually contain more noises that significantly affect the performance of the methods designed on small databases. Also, the need of high efficiency rules out complex and non-parallelizable methods.

Most existing work focused on large instance mining, or image clustering. An inspiring work by Philbin and Zisserman [4] constructed a matching graph by searching with every image as a query in the dataset and then finds its dense sub-graphs. Their matching graph is a powerful way to represent image-level context properties, but it suffers from a low recall due to the whole-image search with spatial verification. Among a few methods that target for scalable small instance mining, geometric min-hashing [1] and thread of features [9] are two representative works. These methods are capable of discovering small instances via instancelevel local matches, but they do not consider the structure of connections between images.

To tackle these two challenges, we propose a novel method that is robust and scalable from the graph perspective for mining both large and

- <sup>1</sup> State Key Lab of Intelligent Technology and Systems, TNList, Department of Computer Science and Technology, Tsinghua University Beijing 100084, China
- <sup>2</sup> Microsoft Research No. 5 Danling Street, Haidian District, Beijing 100080, China
- <sup>3</sup> Microsoft Research One Microsoft Way, Redmond WA 98052, USA





small instances. We argue that a graph structure constructed by local matches not only helps to capture the similarities between instances, but also reflects image-level context properties via indirect paths, thus combining the advantages of [4] and [9]. Figure 1 shows the framework.

First of all we extract local SIFT [3] features, quantize them [5], and augment each local feature with the Hamming Embedding [2] binary code of its feature vector and neighbor features. Then we build a weighted and undirected instance graph with images as vertices. The similarity scores between the augmented local features, which combine the HE weighting scheme [2] and the Jaccard similarity of neighboring visual word sets, contribute to the weights of edges between images. As an example, Figure 2 demonstrates an instance graph of PartialDup [8] dataset. From the sparse instance graph we are able to efficiently discover instance clusters by the proposed greedy breadth-first search algorithm. Experiments show that our method outperforms the state-of-the-art ToF method on both clustering performance and running speed.

- [1] O. Chum, M. Perdoch, and J. Matas. Geometric min-hashing: finding a (thick) needle in a haystack. In *CVPR*, 2009.
- [2] H. Jegou, M. Douze, and C. Schmid. Improving bag-of-features for large scale image search. *IJCV*, 2010.
- [3] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.
- [4] J. Philbin and A. Zisserman. Object mining using a matching graph on very large image collections. In *ICVGIP*, 2008.
- [5] J. Sivic and A. Zisserman. Video google: a text retrieval approach to object matching in videos. In *ICCV*, 2003.
- [6] X.-J. Wang, L. Zhang, M. Liu, Y. Li, and W.-Y. Ma. Arista image search to annotation on billions of web photos. In *CVPR*, 2010.
- [7] X.-J. Wang, Z. Xu, L. Zhang, C. Liu, and Y. Rui. Towards indexing representative images on the web. In ACM MM, 2012.
- [8] Z. Wu, Q. Ke, M. Isard, and J. Sun. Bundling features for large scale partial-duplicate web image search. In CVPR, 2009.
- [9] W. Zhang, H. Li, C.-W. Ngo, and S.-F. Chang. Scalable visual instance mining with threads of features. In ACM MM, 2014.