

# Propagation of geotags based on object duplicate detection

Peter Vajda, Ivan Ivanov, Jong-Seok Lee, Lutz Goldmann and Touradj Ebrahimi

Multimedia Signal Processing Group – MMSPG  
Institute of Electrical Engineering – IEL  
Ecole Polytechnique Fédérale de Lausanne – EPFL  
CH-1015 Lausanne, Switzerland

## ABSTRACT

In this paper, we consider the use of object duplicate detection for the propagation of geotags from a small set of images with location names (IPTC) to a large set of non-tagged images. The motivation behind this idea is that images of individual locations usually contain specific objects such as monuments, buildings or signs. Therefore, object duplicate detection can be used to establish the correspondence between tagged and non-tagged images. Our recent graph based object duplicate detection approach is adapted for this task. The effectiveness of the approach is demonstrated through a set of experiments considering various locations.

**Keywords:** Object duplicate detection, Tag propagation, Geotags, EXIF, IPTC

## 1. INTRODUCTION

The popularity of social networks, digital photography and web-based personal image collections has resulted in a continuously growing volume of publicly available photos. Moreover, a recent trend is also to "geotag" them by labeling them with geographic information or marking the location they were taken.

To manage a large number of photos, tagging is one of the popular methods, which enables us to search our personal photo storages with words. However, tagging by hand for a lot of photos is a time-consuming task. Therefore, robust and efficient algorithms for automatic tagging (or tag propagation) are desirable to help people organize and browse large collections of personal photos in a more efficient way.

The geographical metadata embedded in the image file itself usually consists of location names and GPS coordinates, but may also include altitude, viewpoint, etc. Two of the most commonly used metadata formats for image files are EXIF (Exchangeable Image File Format) and IPTC (International Press Telecommunications Council). In this paper, we consider the existing IPTC schema and introduce a hierarchical order for a subset of the available geotags, namely: city (name of the city where image was taken) and sublocation (area or name of the object).

The growing need for automatic metadata generation, search and retrieval has boosted research efforts towards these directions. Several web-based photo-sharing applications visualize locations on a map where pictures were taken: Picasa, Flickr, Panoramio, Zoomr. A similar web-based application GeoPix, built for sharing pictures taken with mobile phones, has been developed by Carboni *et al.*<sup>1</sup> Hays and Efros<sup>2</sup> propose an algorithm called IM2GPS for estimating a probability distribution over the Earth's surface of a single image using a purely data-driven scene matching approach. They show that the most helpful visual features in disambiguating between different geographic locations are GIST features, color histogram and texton histogram. Their idea differs from ours in a way that they do not recognize specific objects but rather geographic areas. Naaman *et al.*<sup>3</sup> propose a method to automatically organize personal photo collection by combining time and location metadata. Their application PhotoCompass uses time-based event detection techniques and a temporal-geographical clustering algorithm trying to mimic the way people think when organizing their photo collections into groups. Kennedy and Naaman<sup>4</sup> present a method to search representative landmark images from a large collection of geotagged images by making use of global and local visual features. They cluster landmark images into visually similar

---

Further author information: (Send correspondence to Peter Vajda)  
Peter Vajda: E-mail: peter.vajda@epfl.ch

groups and generate links between those images that contain the same visual objects. This method is proved to be effective to extract representative image sets of selected landmarks, but it cannot be applied to images that are not tagged with the landmark name, which limits its applicability. Hoashi *et al.*<sup>5</sup> propose a method for automatic landmark identification of geotagged photos by performing geographical clustering to obtain points-of-interest (POI) and then for each extracted POI they retrieve landmark candidates from the Web based on the geo-location. A query photo is compared with retrieved candidates by making use of location and content-based features. The very recent work of Zheng *et al.*<sup>6</sup> finds frequently photographed landmarks automatically from a large collection of geotagged photos. They perform clustering on GPS coordinates and visual texture features from the image pool, and extract landmark names as the most frequent tags associated with the particular visual cluster. Additionally, they extract landmark names from the travel guide articles, such as Wikitravel\*, and visually cluster photos gathered by querying Google Image Search<sup>†</sup>. However, the test set they use is quite limited – 728 total images for a 124-category problem, or less than 6 test images per landmark. While they focused on mining landmark names and photos, we perform recognition of landmarks. Crandall *et al.*<sup>7</sup> also considered the problem of estimating the geographic locations of photos. In addition to the visual features, they used the spatial distribution of popular places where photos were taken considering GPS coordinates. They found representative images for popular cities and landmarks by matching the SIFT interest points among the photos and considering temporal information, as photos taken within a short period of time are often different shots of the same landmark. In contrast, we target landmark matching by using a graph model, which imposes spatial constraints between SIFT features and thus improves the accuracy of the image matching. Another application that combines textual and visual techniques has been proposed by Quack *et al.*<sup>8</sup> They developed a system that crawls photos on the internet and identifies clusters of images referring to a common object (physical items on fixed locations), and events (special social occasions taking place at certain times). The clusters are created based on the pair-wise visual similarities between the images, and the metadata of the clustered photos are used to derive labels for the clusters. Finally, Wikipedia<sup>‡</sup> articles are attached to the images and the validity of these associations is checked. Gammeter *et al.*<sup>9</sup> extends this idea towards object-based auto-annotation of holiday photos in a large database that includes landmark buildings, statues, scenes, pieces of art, with help of external resources such as Wikipedia. In both Quack *et al.*<sup>8</sup> and Gammeter *et al.*,<sup>9</sup> GPS coordinates are used to pre-cluster objects which may not be always available.

In this paper, we propose to use object duplicate detection for the propagation of geotags. The main motivation of using object duplicate detection comes from its robustness in accepting the same objects and discarding similar objects. New images are automatically annotated based on the detection of the same objects from a small set of training images with associated geotags. The performance of our approach is evaluated through a set of experiments on a comprehensive dataset of unique landmarks. We take into consideration objects like castles, churches, bridges, towers and statues.

The remaining sections of this paper are organized as follows. Section 2 describes our approach for tag propagation. Experiments and results are shown in Section 3. Finally, Section 4 concludes with a summary and some perspectives for future work.

## 2. ALGORITHM

In this section, we present a solution for geotag propagation between images. The main novelty of the proposed method is to use object duplicate detection for geotag propagation, which is precise and robust enough to identify the same object. The system architecture is illustrated in Figure 1 and it contains two functional modules, each of which has a specific task: object duplicate detection and tag propagation.

### 2.1 Object duplicate detection

The goal of the object duplicate detection is to detect the presence of a target object based on some training images containing that object. Such objects may vary from their perspective, be of different sizes and scales, or

---

\*<http://www.wikitravel.com>

†<http://images.google.com>

‡<http://www.wikipedia.org>

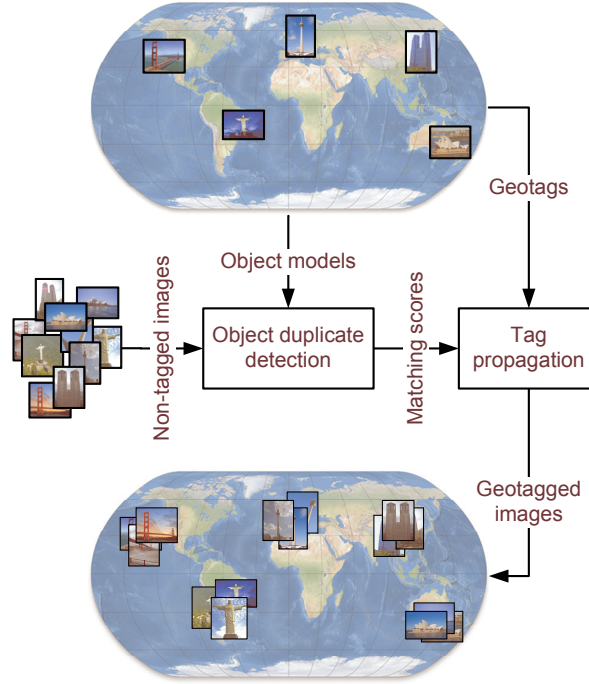


Figure 1. System architecture of the geotag propagation.

modified versions of the original object after minor manipulations, malicious or not, as long as these manipulations do not change their perceptual content.

The algorithm is described in.<sup>10</sup> To resolve the localization problem efficiently, we use sparse features which are extracted from the training and the testing images. The features are robust to arbitrary changes in viewpoints. Spatial graph model matching is applied to improve the accuracy of the detection, which considers the scale, orientation, position and neighborhood of the features.

## 2.2 Tag propagation

Our system supports two operation scenarios as shown in Figure 2. In the *closed set problem*, each test picture is assumed to correspond to one of the known (trained) locations. Therefore, the test image gets assigned the most probable model of the trained location and the corresponding tag is propagated to the test image which can be seen as an identification task. However, in the *open set problem* the test picture may correspond to an unknown location. This scenario is comparable to a watchlist task where the goal is to distinguish between known and unknown locations and propagate the tags only for the known ones.

In more detailed way, the output of the object duplicate detection module is a match score matrix which represents the pair-wise comparison of the trained locations and the testing images. In the closed set problem, we find the maximum score for each test image and propagate the geotag of the corresponding object model. In the open set problem the tag propagation is only done if the corresponding score exceeds a predefined threshold.

## 3. EXPERIMENTS

In this section, we assess the performance of the developed geotag propagation method on a comprehensive dataset and explore the influence of the object type on the accuracy.

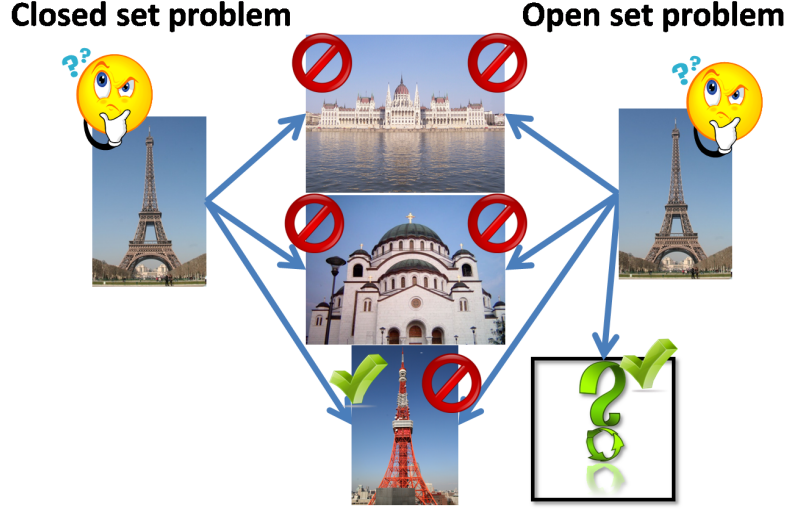


Figure 2. The closed and the open set problem.

### 3.1 Database

The experiments are based on a novel dataset which is split into a training and a testing subsets. We are interested in images that depict some geographically unique objects. For instance, pictures taken by tourists are ideal because they often focus on the unique and interesting objects of a place. The dataset is obtained from Google Image Search and Flickr by querying the associated tags for famous landmarks. This dataset consists of 1320 images: 22 cities (Barcelona, Beijing, Istanbul, London, Paris, Sydney, Moscow, San Francisco, Rio de Janeiro, etc.) and 3 sublocations for each of them (objects or areas in those cities, such as Sagrada Familia, Tiananmen, Hagia Sofia, Big Ben, Eiffel Tower, Opera House, Golden Gate Bridge, etc.). Fig. 3 shows a single image for a single landmark from each of the 22 considered cities, while Fig. 4 provides several images for 3 selected landmarks (e.g. Berlin - Reichstag, San Francisco - Golden Gate Bridge and Paris - Eiffel Tower). Training images are chosen carefully so that they provide a frontal wide angle view of those monuments without other dominating objects. Each sublocation has only one training image. For the testing, 19 images with a large variety of view points and distances are considered for each sublocation leading to an overall testing set of 1254 images.

For the processing, all images are downsized to maximum dimension of  $500 \times 500$  pixels and JPEG compressed. Reducing the image dimensions makes our approach more computationally feasible.

### 3.2 Evaluation

The *open set problem* is evaluated as a detection task where an image has to be classified as known or unknown. Given the ground truth and the predicted labels a confusion matrix is computed, which contains the number of true positives ( $TP$ ), true negatives ( $TN$ ), false positives ( $FP$ ) and false negatives ( $FN$ ). For the evaluation of detection problems, precision recall (PR) curve can be derived from this confusion matrix. PR curves plot the recall ( $R$ ) versus the precision ( $P$ ) with

$$R = \frac{TP}{TP + FN} \quad (1)$$

$$P = \frac{TP}{TP + FP} \quad (2)$$

The F-measure is calculated to determine the optimum thresholds for the object detection. It can be computed as the harmonic mean of the  $P$  and  $R$  values:

$$F = \frac{2 \cdot P \cdot R}{P + R} \quad (3)$$



Figure 3. Examples from our dataset.

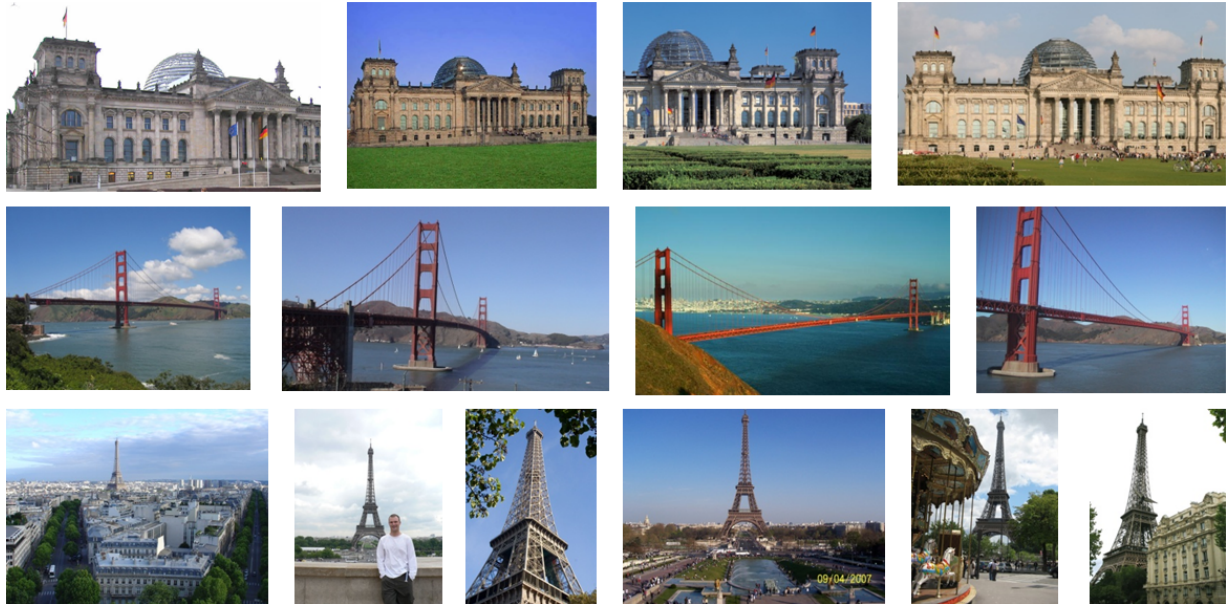


Figure 4. Three landmarks examples from our database.

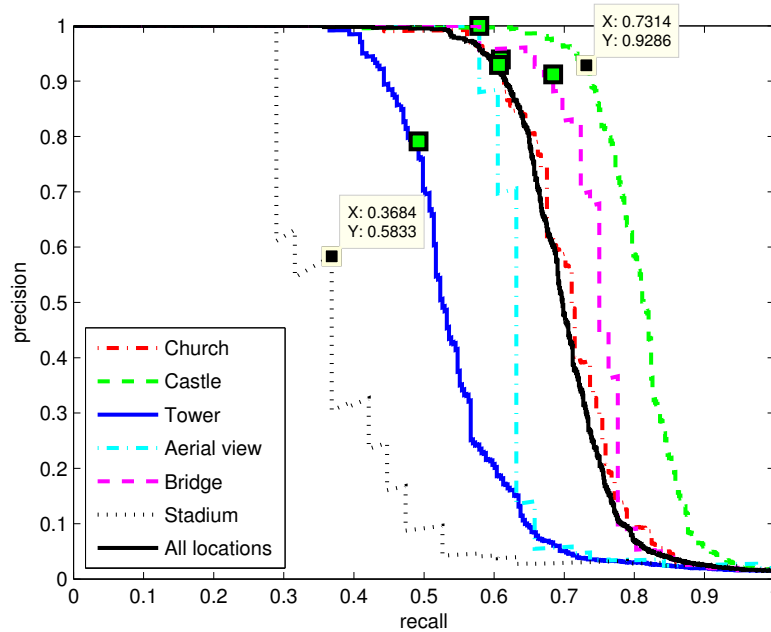


Figure 5. Precision versus recall curve. Markers show the optimal precision and recall, considering the F-measure.

It considers precision and recall equally weighted.

The *closed set problem* is evaluated as a recognition task using the recognition rate ( $RR$ ). It is defined as the ratio between correctly suggested tags  $T$  and overall samples  $A$ :

$$RR = \frac{T}{A} \quad (4)$$

### 3.3 Results

In order to analyze the performance of the geotags propagation, negative and positive test pictures should be selected for each training image. Positive images for each object are 19 pictures in the test subset of the dataset. Negative images (images which do not contain the ground truth object) depict all other objects (65 monuments) which are not related with the training image.

In the analysis below, we separate objects into several groups based on their visual properties. The locations within the dataset can be grouped into 6 categories: castles, churches, bridges, towers/statues, stadiums and aerial view.

At first, the *open set problem* is evaluated as a detection task through the PR curves shown in Figure 5, that provide a good visualization of the opposing effects (high precision versus high recall) which are inherent to any detection task.

The PR curve shows a significant difference between the different groups. The proposed method performs well with castles or buildings which have more salient regions. In case of towers, it performs worse because the object does not have enough discriminative features. However, in case of stadiums, the performance is low due to the large variety of viewpoints.

In order to determine the optimal threshold of object duplicate detection for general applications and compare performances of groups, the F-measures for the different thresholds are calculated and threshold for the maximum F-measure is chosen. We conclude that the optimal threshold does not vary much depending on locations (standard deviation of 13%). Fig. 6 shows the F-measures across the different groups.

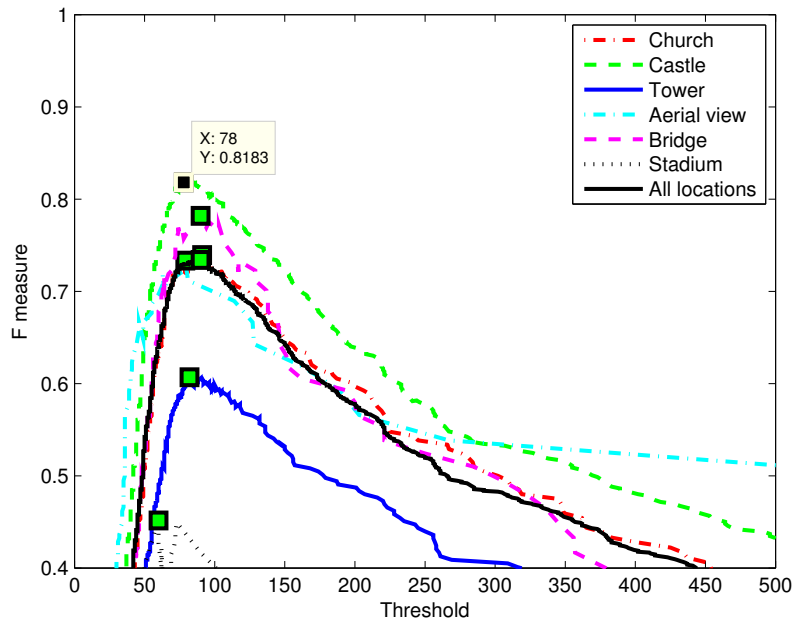


Figure 6. F-measure versus threshold.

The *closed set problem* is evaluated as a recognition task. The *RR* for all test data is 71% and the *RR* for each location is in the range of [20%, 100%]. In the further analysis we focus on the *RR* of each group (Figure 7). Our approach performs best for the group of castles, while stadiums show the worst results. The results of the closed set problem verify the conclusion we made for the open set problem.

#### 4. CONCLUSION

In this paper, we present a system for geotag propagation by making use of object duplicate detection. The main motivation of using object duplicate detection comes from its robustness in accepting the same objects and discarding similar objects. The system is successfully tested with real images of famous landmarks taken from different photo-sharing websites. We get large number of images that are labeled with geographic information, namely object or area name, and city name (IPTC). As a final result, labels from initially tagged images are propagated to the same object found in the test dataset. The open and the closed set problems are evaluated as detection and recognition tasks for tag propagation, respectively. The comparison between different groups shows that castles are more likely to be easily detected due to their large number of salient regions, while towers and stadiums are the most difficult to recognize. Considering the F-measure, the optimal threshold of the object duplicate detection is also determined.

The automatic geotagging system has the potential to be improved in many ways. As a future study, we will explore approaches for incorporating such a system into a standard web-based image search engine.

#### ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation Grant "Multimedia Security" (number 200020-113709), the Swiss National Foundation for Scientific Research in the framework of NCCR Interactive Multimodal Information Management, and partially supported by the European Network of Excellence PetaMedia (FP7/2007-2011).

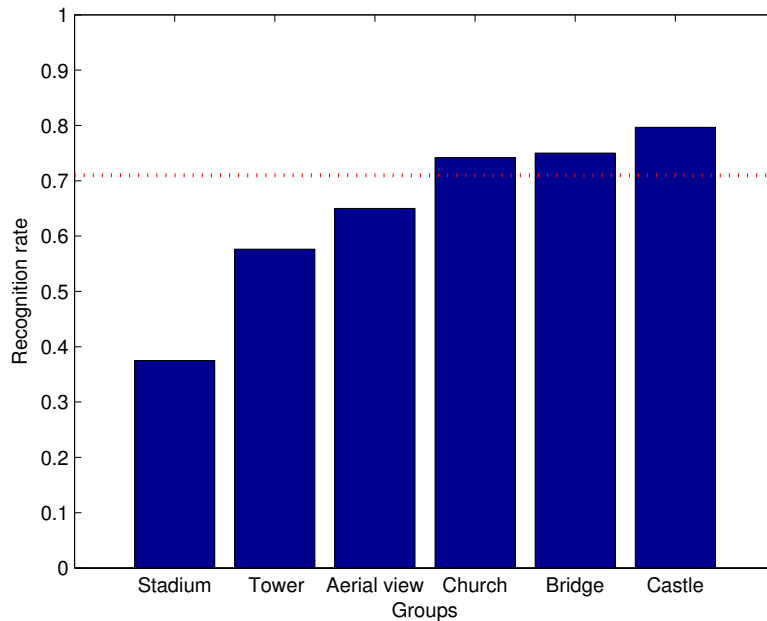


Figure 7. The recognition rate across groups in the closed set problem (bars) and the recognition rate of all locations (dashed line).

## REFERENCES

- [1] Carboni, D., Sanna, S., and Zanarini, P., “GeoPix: image retrieval on the geo web, from camera click to mouse click,” in *[Proc. Int. Conf. Human-Computer Interaction with Mobile Devices and Services]*, 169–172 (2006).
- [2] Hays, J. and Efros, A. A., “IM2GPS: estimating geographic information from a single image,” in *[Proc. Int. Conf. Computer Vision and Pattern Recognition]*, 1–8 (2008).
- [3] Naaman, M., Song, Y. J., Paepcke, A., and Garcia-Molina, H., “Automatic organization for digital photographs with geographic coordinates,” in *[Proc. Int. Conf. Digital Libraries]*, 53–62 (2004).
- [4] Kennedy, L. S. and Naaman, M., “Generating diverse and representative image search results for landmarks,” in *[Proc. Int. Conf. World Wide Web]*, 297–306 (2008).
- [5] Hoashi, K., Uemukai, T., Matsumoto, K., and Takishima, Y., “Constructing a landmark identification system for Geo-tagged photographs based on Web data analysis,” in *[Proc. Int. Conf. Multimedia and Expo]*, 606–609 (2009).
- [6] Zheng, Y., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T., and Neven, H., “Tour the World: Building a web-scale landmark recognition engine,” in *[Proceeding of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)]*, 1085–1092 (2009).
- [7] Crandall, D., Backstrom, L., Huttenlocher, D., and Kleinberg, J., “Mapping the world’s photos,” in *[Proceedings of the 18th International Conference on World Wide Web (WWW 2009)]*, 761–770 (2009).
- [8] Quack, T., Leibe, B., and Van Gool, L., “World-scale mining of objects and events from community photo collections,” in *[Proceedings of the IEEE International Conference on Content-based Image and Video Retrieval (CIVR 2008)]*, 47–56 (2008).
- [9] Gammeter, S., Bossard, L., Quack, T., and Van Gool, L., “I know what you did last summer: Object level auto-annotation of holiday snaps,” in *[Proceedings of the 20th International Conference on Computer Vision (ICCV 2009)]*, (2009).
- [10] Vajda, P., Dufaux, F., Ha, T. M., and Ebrahimi, T., “Graph-based approach for 3D object duplicate detection,” in *[Proc. Int. Work. Image Analysis for Multimedia Interactive Services]*, 254–257 (2009).