Multimed Tools Appl (2012) 56:155–177
DOI 10.1007/s11042-010-0570-7

# Geotag propagation in social networks based on user trust model

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

Published online: 18 July 2010

**Abstract** In the past few years sharing photos within social networks has become very popular. In order to make these huge collections easier to explore, images are usually tagged with representative keywords such as persons, events, objects, and locations. In order to speed up the time consuming tag annotation process, tags can be propagated based on the similarity between image content and context. In this paper, we present a system for efficient geotag propagation based on a combination of object duplicate detection and user trust modeling. The geotags are propagated by training a graph based object model for each of the landmarks on a small tagged image set and finding its duplicates within a large untagged image set. Based on the established correspondences between these two image sets and the reliability of the user, tags are propagated from the tagged to the untagged images. The user trust modeling reduces the risk of propagating wrong tags caused by spamming or faulty annotation. The effectiveness of the proposed method is demonstrated through a set of experiments on an image database containing various landmarks.

**Keywords** Tag propagation · Social networks · Object duplicate detection · Geotags · User trust model · IPTC

I. Ivanov (✉) · P. Vajda · J.-S. Lee · L. Goldmann · T. Ebrahimi
Multimedia Signal Processing Group (MMSPG), Institute of Electrical Engineering (IEL),
Ecole Polytechnique Fédérale de Lausanne (EPFL),
1015 Lausanne, Switzerland
e-mail: ivan.ivanov@epfl.ch

P. Vajda
e-mail: peter.vajda@epfl.ch

J.-S. Lee
e-mail: jong-seok.lee@epfl.ch

L. Goldmann
e-mail: lutz.goldmann@epfl.ch

T. Ebrahimi
e-mail: touradj.ebrahimi@epfl.ch

## 1 Introduction

The past few years have witnessed an increasing popularity of social networks, digital photography and web-based personal image collections. A social network service typically focuses on building online communities of people who share interests and activities, or who are interested in exploring the interests and activities of others. Most social network services are web-based and provide a variety of ways for users to interact. They have become also a popular way to share and disseminate information, e.g. users upload their personal photos and share them through online communities asking other people to comment or rate their content. This has resulted in a continuously growing volume of publicly available photos, e.g. Flickr[1] contains over 3.6 billion photos [26] and more than 2 billion photos are uploaded to Facebook[2] each month [7]. Moreover, a recent trend is also to "tag" them. Tags are short textual annotations used to describe photos, in order to provide additional information to other users who are interested in those images. Tags are also essential in resolving user queries targeting shared photos.

As the popularity of social networking is on a constant rise, new uses for the technology are constantly being observed. To manage a large number of photos, tagging is one of the popular methods, which enables us to search our photo collections with keywords. However, tagging a lot of photos by hand is a time-consuming task. Users typically tag a small number of the shared photos only, leaving most of the other photos with incomplete metadata. This lack of metadata seriously impairs search, as photos without proper annotations are typically much harder to retrieve than correctly annotated photos. 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.

Considering the most popular tags from photo sharing sites, such as Flickr, tags are mostly related to either persons, objects, events or locations. In a large scale analysis of users' tagging behavior and the information provided by tags, Sigurbjörnsson and van Zwol [22] found that 28% of the tags in a random set of 52 million photos from Flickr corresponded to the location type of WordNet [8] categories. For a large portion of images, the association to their geographical locations provides a powerful cue for grouping and indexing. This is especially true for the large number of images depicting famous places from all over the world. Usually, the most salient region in the image corresponds to a specific landmark or object. When users annotate such images, they link a geotag to the object depicted in the image. Therefore, we propose to use object duplicate detection for the propagation of geotags since it is robust in detecting the same object and discarding similar objects. Untagged images are automatically annotated based on the detection of the same object from a small set of training images with associated tags.

Since the initial tags are provided by humans, it cannot be taken for granted that they are always correct. A user may put wrong tags just by mistake or even on purpose. Consequently, the trend to collaboratively attach any, theoretically unrestricted, free-form keywords (tags) to multimedia content may produce a danger

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

[2] http://www.facebook.com

that wrong or irrelevant tags may finally be propagated to other photos and prevent users from the benefits of annotated photos. Therefore, we propose to consider user trust information derived from users' tagging behavior for the tag propagation.

The remaining sections of this paper are organized as follows. We introduce related work in Section 2. Section 3 describes our approach for geotag propagation and discusses two application scenarios. Experiments and results are shown in Section 4. Finally, Section 5 concludes the paper with a summary and some perspectives for future work.

## 2 Related work

The proposed system is related to different research fields including visual analysis, geographic information systems, social networking and tagging. Therefore, the goal of this section is to review the most relevant work in the following fields: (a) joint analysis of visual content and geographical context, (b) human tagging and user trust modeling.

### 2.1 Combined analysis of geographical context and visual content

Google Image Search[3] is a popular image retrieval system, which is based on keywords extracted from the filename of the image, the link text pointing to the image, and text adjacent to the image. These keywords may contain geographical locations such as places, streets or cities. Beside this contextual information, support for content based image search based on visual features, such as color or image size, was recently added.

Most of the photo sharing websites (e.g. Flickr, Picasa,[4] Panoramio,[5] Zoomr[6]), provide information about where images were taken in form of maps or groups. This information is either provided by an external GPS sensor and stored as image metadata (Exchangeable Image File Format (EXIF) [23], International Press Telecommunications Council (IPTC) [12]) or manually annotated via geocoding. Our goal is to derive this information from the content of the image by comparing it to a small set of already tagged images.

Carboni et al. [5] have developed a web-based application called *GeoPix* that supports photo sharing through mobile phones. It allows mobile users equipped with camera phones to capture images, annotate and finally upload them on the GeoPix web page. Important information about the locations assigned to images is provided by a GPS devices or mobile phones.

Kennedy and Naaman [14] presented a method to search representative landmark images from a large collection of geotagged images. This method uses tags and the geographical location representing a landmark. Then, it analyzes the visual features

[3] http://images.google.com

[4] http://picasa.google.com

[5] http://www.panoramio.com

[6] http://www.zooomr.com

(global color and texture features, as well as Scale Invariant Feature Transform (SIFT) to cluster landmark images into visually similar groups. This method has been proven to be effective to extract representative image sets of selected landmarks, but it cannot be applied to untagged images, which limits its applicability.

The very recent work of Zheng et al. [28] 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,[7] and visually cluster photos gathered by querying Google Image Search. 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.

A pioneering paper in this area by Hays and Efros [10] proposed an algorithm called *IM2GPS* to estimate the locations of a single image using a purely data-driven scene matching approach. Given a test image, their approach finds the visual nearest neighbors in the database and estimates a geolocation of the image from those GPS tagged nearest neighbors. The estimated image location is represented as a probability distribution over the Earth's surface. However, the IM2GPS approach showed low recognition accuracy due to low-level features. While IM2GPS uses a set of more than 6 million training images, its general applicability is inconclusive because the performance was verified only on 237 hand-selected test images. Another drawback is limited availability of GPS coordinates associated to images. Our approach differs from their method in the way that we focus more on recognizing specific locations (landmarks), and not geographic scenes and areas (such as forest or savanna). We aim at achieving high recognition rates considering the problem as a high-level object duplicate detection task, rather than a low-level image matching task used in IM2GPS.

Crandall et al. [6] 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.

Cao et al. [3] proposed to infer the correlation between images based on low-level color and SIFT features, GPS labels and time stamps. They employed a probabilistic Bayesian algorithm for tag propagation based on a pair-wise similarity within a photo collection. As a final result, tags are propagated to the untagged images if the similarity measure exceeds a predefined threshold, meaning that multiple tags can be assigned to one photo. While the general idea of [3] is the same as ours, we use a spatial information of visual features, which makes our approach more robust, and consider a user trust modeling.

[7] http://www.wikitravel.com

Wu et al. [27] proposed the semi-automatic tag recommendation approach which requires users to provide a few initial tags and then conducts the recommendation accordingly. The tag recommendation is based on multi-domain relevance, such as tag co-occurrence, tag-to-tag similarity and image-conditioned tag similarity. Another interesting approach is presented by Cao et al. [4]. GPS locations of photos are clustered and the Logistic Canonical Correlation Regression (LCCR) method is applied for each cluster which maximizes the correlation between heterogeneous visual and tag features and an annotation lexicon. They use this approach to enhance tags for general web images and predict the geographical region where an image was taken. On the other hand, the approach presented in this paper considers images of specific landmarks rather than general categories, such as holiday, vacation, music, fun, birthday.

Another application that combines textual and visual techniques has been proposed by Quack et al. [20]. 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[8] articles are attached to the images and the validity of these associations is checked. Gammeter et al. [9] 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 [20] and [9], GPS coordinates are used to pre-cluster objects which may not be always available.

Most of the recent systems rely on GPS coordinates to derive geographical annotation, which is not available for the majority of web images and photos, since only a few camera models are equipped with GPS devices. Furthermore, a GPS sensor in a camera provides only the location of the photographer instead of that of the captured landmark, which may be up to several kilometers away. Therefore, the GPS coordinates alone may not be enough to distinguish between two landmarks within a city. Describing landmarks through location names rather than GPS coordinates is not only more reliable but also more expressive. A recent study by Hollenstein and Purves [11] indicated that geotagging should follow the way people actually describe locations, i.e. it is more convenient to use: Belgrade - Church of Saint Sava, rather than: latitude 44.798083, longitude 20.46855. Therefore, there is growing interest in the research community to infer geographic locations of the scenes in photos based on visual and text features.

### 2.2 User trust modeling

In a system with user contributions, the reliability of the users is critical to the performances of that system. In general, user provided tags may be either correct or wrong. There are several reasons to assign wrong tags to images. First of all, users are human beings and make mistakes. On the other hand, it is also possible to provide wrong tags on purpose for advertisement or increase of the rank of a particular tag

[8] http://www.wikipedia.org

in automatic search engines. These users are considered as spammers in analogy to email spammers.

A well known email spam filtering method is the Naive Bayesian classifier proposed by Sahami et al. [21]. The system learns the probability of each word within a training set of spam and non-spam emails. Then, these probabilities are used to calculate the probability that a test email is spam, which depends on whether it contains some particular words or not. If the probability is higher than a user-defined threshold, the email is considered as a spam.

In case of a user trust model, the same idea can be applied. To make an analogy with email spam filtering, users in the trust model correspond to emails and image tags correspond to words inside emails. The probability that a user is a spammer is calculated separately for each user. However, the computation of the spamming probability for users differs from email, as shown in Section 3.3.

Computational trust models have recently drawn the attention of researchers as more online communities are available to Internet users. The common ways of determining trust are through reputation [13, 16]. A user can measure the reputation of another user based on the past interactions between the former and the latter (personal experience), as well as the interactions between the latter and other users (recommendations). Another approach is to allow users to express the levels of their trust in other users. There are several research directions in this area. One of the most famous examples is the *Google* page rank system where each page is ranked depending on the number of links to it from different sites [2], where pages are comparable to users in a user trust model.

Massa and Avesani [17] compared and evaluated user trust models on their web-based application presented online at *Epinions*.[9] They considered two kinds of user trust models, global and local trust. The global trust model assigns a trust value to each user, independently of who is evaluating the other users. However, the local trust model considers the point of view of the evaluating user. They analyzed the differences in accuracy and coverage of local and global trust models, by focusing on controversial users—users who are judged in very different ways by other users.

## 3 System overview

In this section, we present our solution for geotag propagation between images. The main innovation is the combination of object duplicate detection and user trust modeling for accurate and reliable geotag propagation. The system architecture is illustrated in Fig. 1. It contains three functional modules, each of which has a specific task: object duplicate detection, user trust modeling evaluation and tag propagation. Since the object duplicate detection has been described previously [24, 25] the focus here is on the latter two modules.

The system takes a small set of training images with associated geotags to create the corresponding object (landmark) models. These object models are used to detect object duplicated in a set of untagged images. As a result, matching scores between the models and the images are obtained. According to the scores, the tag propagation

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

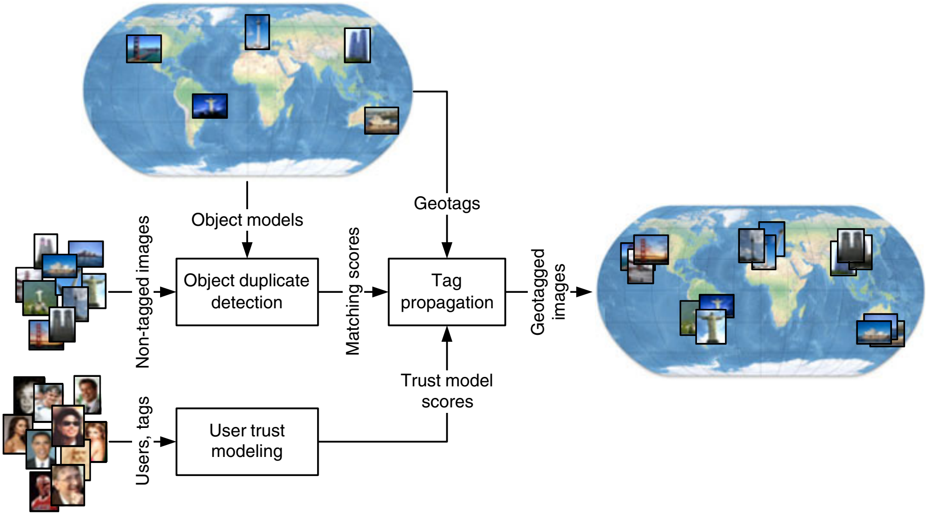


**Fig. 1** Overview of the system for geotag propagation. The object duplicate detection is trained with a small set of images with associated geotags. The created object (landmark) models are matched against non-tagged images. The resulting matching scores serve as an input to the tag propagation module, which propagates the corresponding tags to the untagged images. Given a user trust model only the tags from reliable users are propagated

module makes decisions about which geotags should be propagated to the individual images. Given a user trust model which describes the tagging reliablity of each user, only the tags from the users who are trusted are propagated to the photos in the dataset.

### 3.1 Object duplicate detection

The goal of the object duplicate detection module is to detect the presence of a target object in an image based on an object model created from training images. Duplicate objects may vary from their perspective, have different size, or be modified versions of the original objects after minor manipulations, as long as such manipulations do not change their identity. The basic idea of applying object duplicate detection for geotag propagation is that images of places typically depict distinctive landmarks (buildings, mountains, bridges, etc.) which can be considered as object duplicates.

Training is performed as follows: given a set of images, features are extracted and a spatial graph model describing the object, i.e. landmark, is created for each of the landmarks. In our case, one training image per landmark is used to create a graph model. First, regions of interest (ROIs) in an image are extracted using the Hessian affine detector [18] and each of these regions are described using SIFT features [15]. These features are robust to arbitrary changes in viewpoints. Then, hierarchical k-means clustering [19] is applied to the features, to group them based on their similarity. The result of the hierarchical clustering is used for the fast approximation of the nearest neighbor search, to efficiently resolve feature matching. Finally, a spatial graph model is constructed to improve the accuracy of the feature matching, which considers the scale, orientation, position and neighborhood of features. The nodes of the graph are the features of the training images. The edges of the graph connect features with their spatial nearest neighbors. The attributes of edges are the distance and orientation of the neighbors. These attributes are important for the matching step in the test phase.

To detect the presence of the landmark within a test image, the features are extracted from the image in the same way described above. These features are matched to those in the graph model derived from the training images using a one-to-one nearest neighbor matching. Considering only matched features and their positions, a spatial graph model of the query image is constructed in the same way described in the training phase. Then, graph matching is applied between two graph models to identify the local correspondences between regions in the training and the test image. Finally, for the global object matching and matching score computation, the general Hough transform [1] is applied on the nodes of the matched graph. The matching scores represent the pair-wise comparison of training and test images.

More details about the proposed object duplicate detection approach are presented in [24, 25].

### 3.2 Tag propagation

The goal of the tag propagation module is to propagate the geotags from the tagged to the non-tagged images according to the matching scores, provided by the object duplicate detection module. As a result, labels from the training set are propagated to the same object found in the test set.

The geographical metadata (geotags) embedded in the image file usually consist of location names and/or GPS coordinates, but may also include altitude, viewpoint, etc. Two of the most commonly used metadata formats for image files are EXIF and IPTC. 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 landmark).

Our system supports two application scenarios as shown in Fig. 2. In the *closed set problem*, each test image is assumed to correspond to exactly one of the known (trained) landmarks. Therefore, the image gets assigned to the most probable trained

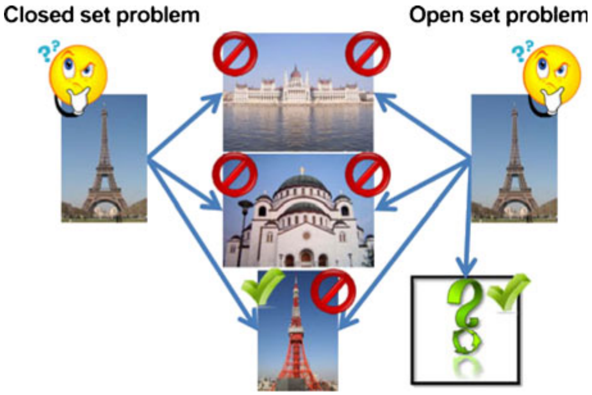


**Fig. 2** The closed and the open set problems. In the closed set problem, each test picture is assumed to correspond to one of the known (trained) landmarks. However, in the open set problem the test picture may also correspond to an unknown landmark

landmark and the corresponding tag is propagated to the test image. This is comparable to an identification task in biometrics. However, in the *open set problem* the test picture may correspond to an unknown landmark. This problem is comparable to a watchlist task in biometrics where the goal is to distinguish between known and unknown persons (landmarks) and to propagate the tags only for the known ones. For example, in Fig. 2 we assume that the system is trained with only three known landmarks: Budapest (Parliament), Belgrade (Church St. Sava) and Tokyo (Tower). Given the input test image of Paris (Eiffel Tower), the system gives different results for the closed and open set problems. In the closed set problem, our system finds that Tokyo (Tower) is the most suitable model for the test image. If we consider the open set problem, the system does not retrieve any of the trained models since the matching scores between the object models and the test image do not exceed a predefined threshold. The open and closed set problems are separately evaluated in Section 4 as detection and recognition tasks, respectively.

In a more detailed way, the object duplicate detection module provides a matching score matrix $S_{i,j}$. It represents the pair-wise comparison of the trained images (landmarks) $i$, $i \in [1, M]$, and the test images $j$, $j \in [1, N]$, where $M$ and $N$ are number of training and test images, respectively.

In the closed set problem, we find the maximum score for each test image $j$ and propagate the geotag of the corresponding training image $i$. The assignment matrix $C_{i,j}$, $i \in [1, M]$ and $j \in [1, N]$, is formed in the following way:

$$C_{i,j} = \begin{cases} 1, \text{ if } S_{i,j} = \max\limits_{i \in [1,M]} \{S_{i,j}\}; \\ 0, \text{ otherwise.} \end{cases} \tag{1}$$

In this case, each test image gets assigned with exactly one tag from the training photo dataset.

In the open set problem the tag propagation is only done if the corresponding score exceeds a predefined threshold $\hat{S}$. The assignment matrix $O_{i,j}$, $i \in [1, M]$ and $j \in [1, N]$, is defined as:

$$O_{i,j} = \begin{cases} 1, \text{ if } S_{i,j} = \max\limits_{i \in [1,M]} \{S_{i,j}\} \wedge S_{i,j} \geq \hat{S}; \\ 0, \text{ otherwise.} \end{cases} \tag{2}$$

In this case, each test image can get assigned zero or one tag from the training set depending on the value of the threshold $\hat{S}$.

Based on the assignment matrix $C_{i,j}$ or $O_{i,j}$ the tags are propagated. If the corresponding value is 1, the tag associated with training image $i$ is propagated to the test image $j$. If the corresponding value is 0, no tag is propagated.

### 3.3 User trust modeling

As described before, we introduce users in our system in order to simulate a real social network. The methodology used in this paper is to extract a sub network from a large social network, in a way that every user in this sub network annotates every landmark in the subset of the dataset. Upon this sub network, we build up an automatic propagation system in order to decrease the annotation time and increase the accuracy of the system. In this case, our system relies on user-provided tags which may sometimes be spam annotations given on purpose or wrong tags given by mistake. The users are evaluated and only tags from users whose trust model exceeds a predefined threshold are propagated to other images of the database.

First, we evaluate the trust or reliability of users by making use of their past behavior in tagging. We want to distinguish between users who provide reliable geotags, and those who do not. After user evaluation and trust model creation, tags will be propagated to other photos in the database only if the user is trusted. Assuming that there are $L$ users who tag $M$ training images, a matrix $U_{i,l}$, $i \in [1, M]$ and $l \in [1, L]$, is defined as:

$$U_{i,l} = \begin{cases} 1, & \text{if user } l \text{ tags image } i \text{ correctly;} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The process of comparing the predicted tags to ground truth tags can be done automatically or manually. As an example, automatic tag checking can be done by making use of WordNet [8] and external resources with images and text (e.g. Wikipedia). Given the predicted geotag, WordNet returns a set of the closest words (tags) to that geotag, and this set of tags is used to acquire ground truth images from Wikipedia. Then, the object duplicate detection is performed on retrieved images to see if the tags from WordNet correspond to the given image. Nevertheless, we considered only manually defined ground truth for our experiments.

Here we introduce a simple user trust ratio $T_l$ for user $l$. It is computed as the percentage of the correctly tagged images among all images tagged by user $l$:

$$T_l = \frac{\sum_{i=1}^{M} U_{i,l}}{M}. \tag{4}$$

Only tags from users who are trusted are propagated to other photos in the dataset. In other words, if the user trust ratio $T_l$, $l \in [1, L]$, exceeds a predefined threshold $\hat{T}$, then all his/her tags are propagated. Otherwise, none of his/her tags are propagated.

In this paper, ground truth data are used for the estimation of the user trust ratio. However, for a real photo sharing system, such as Panoramio, it is not necessary to collect ground truth data since user feedback can replace them. The main idea is that users evaluate tagged images by assigning a true or a false flag to the tag associated with an image. If the user assigns a false flag, then he/she needs to suggest a correct tag for the image. The more misplacements a user has, the more untrusted he/she is. By applying this method, spammers and unreliable users can be efficiently detected and eliminated. Therefore, the user trust ratio is calculated as the ratio between the number of true flags and all associated flags over all images tagged by that user. The number of misplacements in Panoramio is analogous to the number of wrongly tagged images in our approach.

In case that a spammer attacks the system, other users can collaboratively eliminate the spammer. First, the spammer wants to make other users untrusted, so he/she assigns many false flags to the tags given by those other users and sets new, wrong, tags to these images. In this way, the spammer becomes trusted. Then, other users correct the tags given by the spammer, so that the spammer becomes untrusted and all of his/her feedbacks in the form of flags are not considered in the whole system. Finally, previously trusted users, who were untrusted due to spammer attack, recover their status. Following this scenario, the user trust ratio can be constructed by making use of the feedbacks from other users who agree or disagree with the tagged location. However, due to the lack of a suitable dataset which provides user feedback, the evaluation of the user trust scenario is based on the simulation of the social network environment.

## 4 Experiments

In this section the performance of the proposed geotag propagation method is evaluated and analyzed in two application scenarios. The considered dataset is described in Section 4.1, followed by the explanation of the two scenarios in Section 4.2. In Section 4.3 the evaluation is presented and finally the results are discussed for the two scenarios in Sections 4.4 and 4.5, respectively.

### 4.1 Dataset

A new dataset was created in order to evaluate the proposed geotag propagation method. We are interested in images that depict geographically unique landmarks. For instance, pictures taken by tourists are ideal because they often focus on the unique and interesting landmarks of a place. The dataset is obtained from Google Image Search, Flickr and Wikipedia by querying the associated tags for famous landmarks.

The resulting dataset consists of 1,320 images: 22 cities (such as Amsterdam, Barcelona, London, Moscow, Paris) and 3 landmarks for each of them (objects or areas in those cities, such as Bird's Nest Stadium, Sagrada Familia, Reichstag, Golden Gate Bridge, Eiffel Tower). Each landmark has 20 image samples. Figure 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). As can be seen from these samples, images with a large variety of view points and distances are considered for each landmark. Figure 5 provides all cities and landmarks from our dataset.

The dataset is split into a training and a test set. Training images are chosen carefully so that they provide a wide angle view of those landmarks without other dominating objects. One image from each landmark is chosen as a training image. All other images from the dataset are test images.

In order to make our approach more computationally feasible, all images are downsized to a maximum size of $500 \times 500$ pixels and JPEG compressed before further processing.

### 4.2 Scenarios

In the *tag propagation scenario* we evaluate our automatic geotag propagation algorithm without including users and their mistakes in the annotation process. First, training images are selected for each landmark. Moreover, for each training image, negative and positive test pictures are selected. For each landmark there are 19 positive images in the test set. Negative images are all images which do not contain the ground truth landmark, namely all images which depict one of the other 65 landmarks. This leads to $19 \times 65 = 1{,}235$ negative images in the test set. In the evaluation of this method in Section 4.4, we consider both the open and closed set problems.

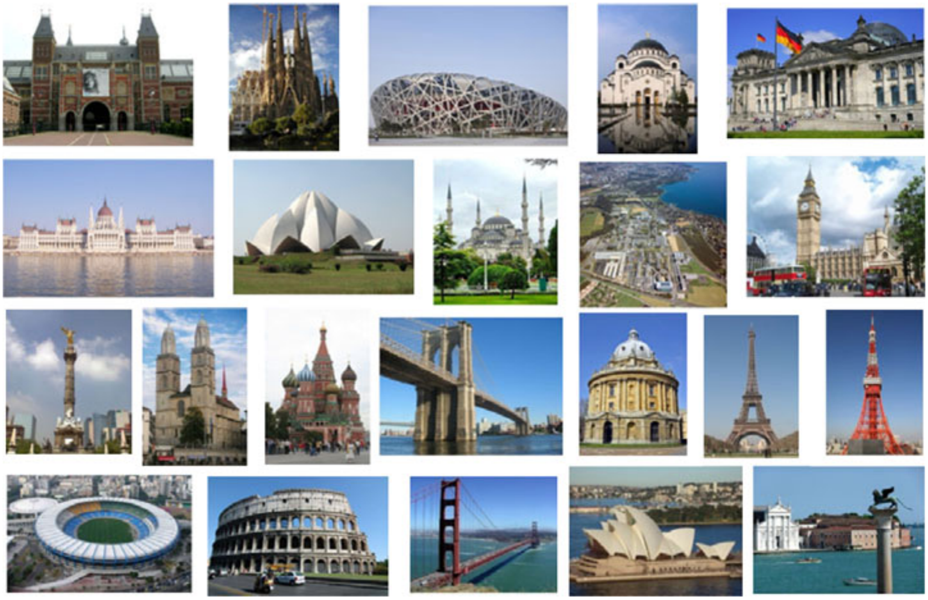


**Fig. 3** Sample landmarks for each of the 22 cities within the dataset. The dataset covers a large variety of landmarks including buildings, bridges, monuments, etc.

In the *user trust scenario* we simulate a social network environment. As explained in Section 3.3, due to the lack of a suitable dataset which provides user feedback from Panoramio, the evaluation of the user trust scenario in this paper is based on the simulation of the social network environment. We performed experiments in which $L = 44$ users were asked to tag $M = 66$ photos from the dataset, putting the

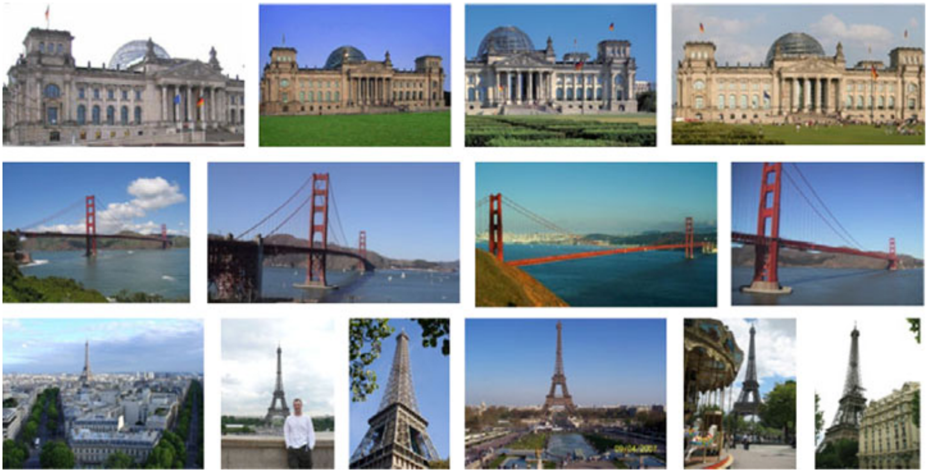
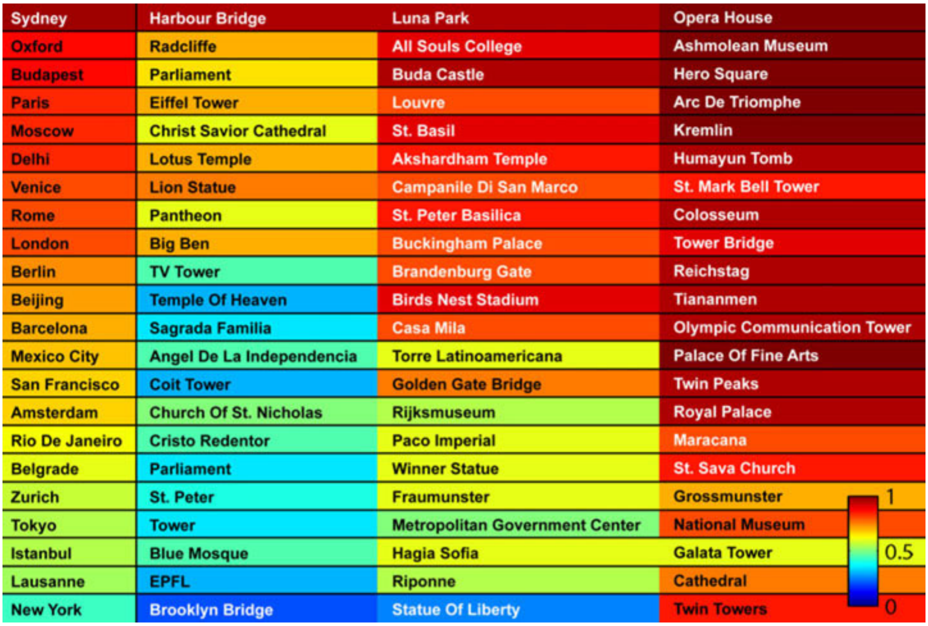


**Fig. 4** Images for 3 selected landmarks: Berlin (Reichstag), San Francisco (Golden Gate Bridge) and Paris (Eiffel Tower). The images contain a large variety of views, distances, and partial occlusions



| | | | |
|---|---|---|---|
| Sydney | Harbour Bridge | Luna Park | Opera House |
| Oxford | Radcliffe | All Souls College | Ashmolean Museum |
| Budapest | Parliament | Buda Castle | Hero Square |
| Paris | Eiffel Tower | Louvre | Arc De Triomphe |
| Moscow | Christ Savior Cathedral | St. Basil | Kremlin |
| Delhi | Lotus Temple | Akshardham Temple | Humayun Tomb |
| Venice | Lion Statue | Campanile Di San Marco | St. Mark Bell Tower |
| Rome | Pantheon | St. Peter Basilica | Colosseum |
| London | Big Ben | Buckingham Palace | Tower Bridge |
| Berlin | TV Tower | Brandenburg Gate | Reichstag |
| Beijing | Temple Of Heaven | Birds Nest Stadium | Tiananmen |
| Barcelona | Sagrada Familia | Casa Mila | Olympic Communication Tower |
| Mexico City | Angel De La Independencia | Torre Latinoamericana | Palace Of Fine Arts |
| San Francisco | Coit Tower | Golden Gate Bridge | Twin Peaks |
| Amsterdam | Church Of St. Nicholas | Rijksmuseum | Royal Palace |
| Rio De Janeiro | Cristo Redentor | Paco Imperial | Maracana |
| Belgrade | Parliament | Winner Statue | St. Sava Church |
| Zurich | St. Peter | Fraumunster | Grossmunster |
| Tokyo | Tower | Metropolitan Government Center | National Museum |
| Istanbul | Blue Mosque | Hagia Sofia | Galata Tower |
| Lausanne | EPFL | Riponne | Cathedral |
| New York | Brooklyn Bridge | Statue Of Liberty | Twin Towers |

**Fig. 5** The recognition rate for all landmarks. Each row represents one city from our dataset and the right three columns represent three landmarks in each of the cities. The first column shows the sorted average recognition rates for each city

name of the landmark depicted in the image. For creating the user trust model, if these geotags are correlated with the landmark in the image, we assume that the image is correctly tagged. After having created the user trust ratio $T_l$, $l \in [1, L]$, we perform tag propagation based on those annotated images. By counting the number of correctly tagged images in this scenario we can mimic the user behavior in Panoramio, where the number of misplacements is considered.

## 4.3 Evaluation

In this section the evaluation methods for the user driven tag propagation system are described. While the tag propagation scenario is evaluated as a closed and an open set problem, the user trust scenario is only evaluated as a closed set problem.

An *open set problem* can be evaluated as a typical detection task, where an image has to be classified as known or unknown. Given the ground truth and the predicted labels, the numbers of true positives ($TP$), false positives ($FP$) and false negatives ($FN$) are computed. Precision recall (PR) curves can be also derived, which plot the recall ($R$) versus the precision ($P$) with:

$$P = \frac{TP}{TP + FP}, \tag{5}$$

$$R = \frac{TP}{TP + FN}. \tag{6}$$

The F-measure is calculated to determine the optimum threshold ($\hat{S}$ in (2)) for the object duplicate detection. It can be computed as the harmonic mean of the $P$ and $R$ values:

$$F = \frac{2 \cdot P \cdot R}{P + R}. \tag{7}$$

Thus it considers precision and recall equally weighted.

For the evaluation of the specific scenarios a ground truth matrix is defined as:

$$GT_{i,j} = \begin{cases} 1, \text{ if } Landmark(i) = Landmark(j); \\ 0, \text{ otherwise.} \end{cases} \tag{8}$$

where $i \in [1, M]$, $j \in [1, N]$, $M$ is the number of training images and $N$ is the number of test images.

Given the assignment matrix $O_{i,j}$ for the open set problem, $TP$, $FP$ and $FN$ can be calculated as

$$TP = \sum_{i,j} GT_{i,j} \cdot O_{i,j}, \tag{9}$$

$$FP = \sum_{i,j} \left(1 - GT_{i,j}\right) \cdot O_{i,j}, \tag{10}$$

$$FN = \sum_{i,j} GT_{i,j} \cdot \left(1 - O_{i,j}\right). \tag{11}$$

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

$$RR = \frac{T}{A}. \tag{12}$$

For tag propagation, $T$ and $A$ can be calculated as

$$T = \sum_{i,j} GT_{i,j} \cdot C_{i,j}, \tag{13}$$

$$A = \sum_{i,j} GT_{i,j}. \tag{14}$$

For user trust modeling, $T$ and $A$ are defined as

$$T = \sum_{i,j,l:T_l \geq \hat{T}} GT_{i,j} \cdot C_{i,j} \cdot U_{i,l}, \tag{15}$$

$$A = \left(\sum_{i,j} GT_{i,j}\right) \cdot \left(\sum_{l:T_l \geq \hat{T}} 1\right). \tag{16}$$

where $\hat{T}$ is the threshold for the user trust ratio, index $l \in [1, L]$ and $L$ is the number of users. In other words, $A$ is the number of propagated tags for each trusted user and $T$ is the truly propagated tags among those. Therefore, a propagated tag is only considered as correct if the annotated tag was correct and propagated correctly.

### 4.4 Results of the tag propagation scenario

In this section the results of the tag propagation scenario are discussed.

First, the *closed set problem* is evaluated as a recognition task. The recognition rate for all landmarks is shown in Fig. 5. The average recognition rate for each city is presented in the first column of this figure and the values are sorted. In our dataset, there are three landmarks in each of the cities which are represented in the right columns of the figure.

The performance of the tag propagation varies considerably between different cities, but also across the individual landmarks within a city. According to common visual properties, all landmarks are split into different groups, such as castles, churches, bridges, towers/statues, stadiums and ground structure, and the further evaluation of the tag propagation scenario is based on these groups. Interestingly, by taking a closer look at the results in Fig. 5, one can perceive that the performances of landmarks within the same group show small variations.

Figure 6 provides the recognition rates averaged over the different groups, which shows variation between the groups. Our approach performs the best for the group of castles, while stadiums show the worst results. The average $RR$ across all landmarks is 71%. The recognition errors are solely caused by the object duplicate detection. Since the user trust scenario includes also tagging errors, the $RR$ of the second scenario is expected to be lower.

Secondly, the *open set problem* is evaluated as a detection task through the PR curves shown in Fig. 7. The PR curves show significant difference between the different groups of landmarks. The proposed tag propagation scenario performs well with castles or other buildings which have more salient regions. In case of towers, it performs worse because the landmark does not have enough discriminative features. However, in case of stadiums, the performance is low due to the large variety of viewpoints.

The F-measures for the different detection thresholds $\hat{S}$ are calculated to determine the optimal threshold value. Figure 8 shows the F-measures across the different groups. The F-measures for the different thresholds are calculated and the optimal

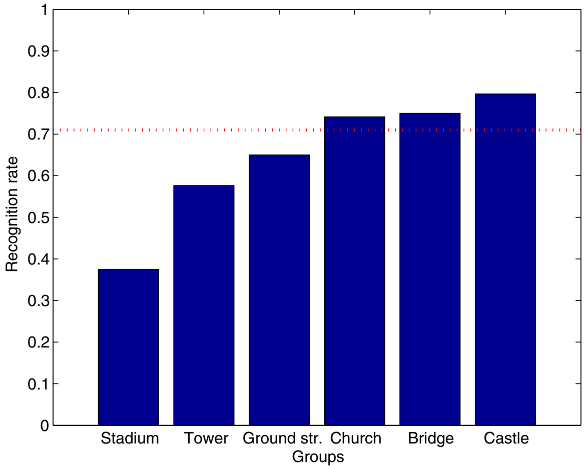
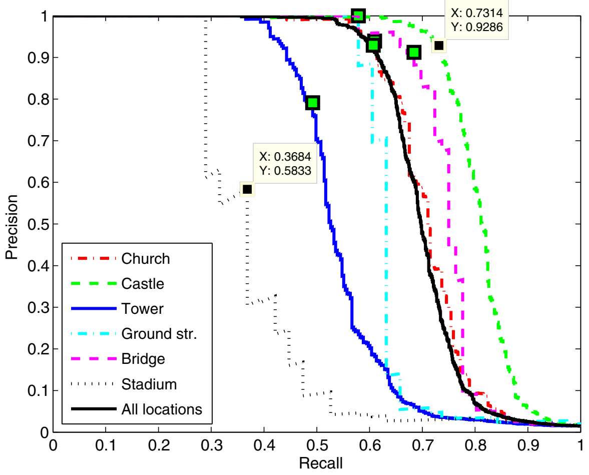


**Fig. 6** The recognition rate across the different landmark groups in the closed set problem (*bars*) and the recognition rate of all landmarks (*dashed line*). Landmarks have been grouped according to their visual characteristics

**Fig. 7** Precision versus recall curves for the open set problem across the different landmark groups. Markers show the optimal precision and recall, considering the F-measure

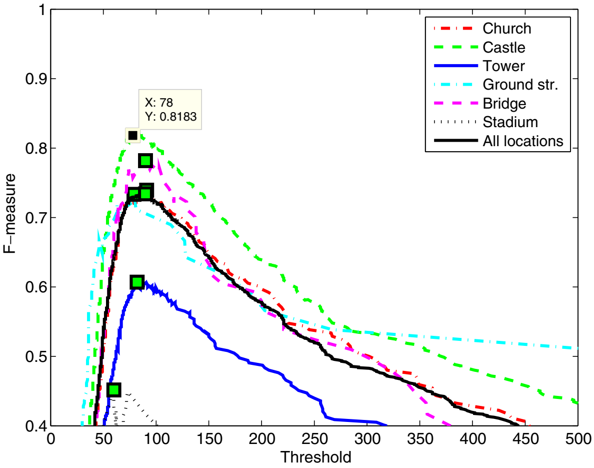


threshold is chosen for the maximum F-measure and shown by green markers. The optimal threshold value does not vary much depending on landmarks (standard deviation of 13%). The final F-score for the open set problem averaged over the whole dataset is 73%.

### 4.5 Results of the user trust scenario

Figure 9 shows the user trust ratio calculated by (4). This user trust ratio can be dynamically updated by making use of several resources with ground truth data as already described in Section 3.3. For this reason two curves are plotted in Fig. 9. The first curve, marked with "x", shows the user trust ratio calculated for a subset of 20 images among 66 training images, which simulates a preliminary result within the dynamic system. The second curve, marked with "o", shows the results after taking into account all 66 images. According to this figure, one can conclude that already 20

**Fig. 8** F-measure versus detection threshold $\hat{S}$ across different landmark groups. *Green markers* show the optimal thresholds



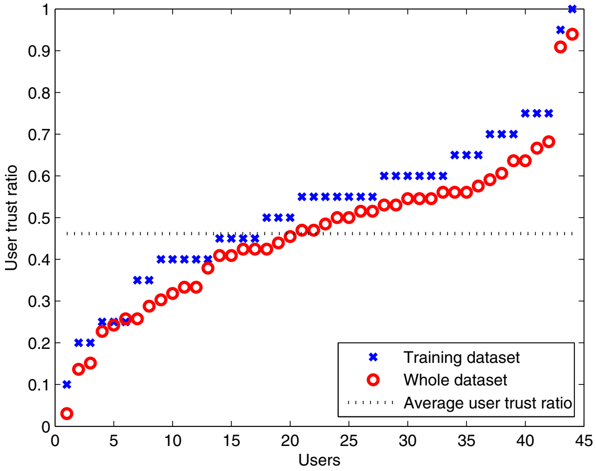


**Fig. 9** Sorted trust ratios for the different users for a reduced and the complete set of training images. The results show wide variety of knowledge among the users

images show good prediction of the user trust model in comparison with the whole dataset. In the further analysis the predicted trust model which uses 66 images is exploited. The results in Fig. 9 show that the user trust ratios vary in a wide range from 0 to 100%. Another interesting observation is that for 20 images the user trust ratios are higher than that for the entire dataset of 66 images. The reason for this can be in human nature that annotation of 66 images without break makes subjects (users) tired and less precise after a while, even if the whole experiment takes no longer than 20 mins. The figure also shows the average user trust ratio for the whole dataset, which is equivalent to the accuracy of the user tagging.

Figure 10 shows the accuracy of the system and the percentage of the number of propagated tags versus the threshold set for the user trust model. The optimal accuracy using object duplicate detection for geotag propagation is 71%, as shown in Section 4.4. However, in this scenario the error of the user tagging step leads to a decrease of the performance. This error is caused by wrong tags given by the users.

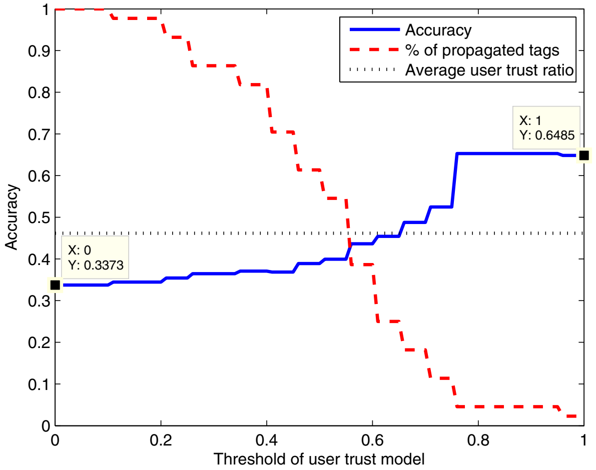


**Fig. 10** The recognition rate of the geotag propagation system and the percentage of the propagated tags versus the threshold $\hat{T}$ for the user trust ratio

The optimal results can be reached if we set the threshold $\hat{T}$ to a high value, but then the number of propagated tags becomes very low. On the other hand, when the threshold is low, more tags are propagated. These curves could be used to determine an appropriate threshold for the proposed user trust model. The higher the threshold for the user trust model is, the more reliable the geotag propagation system is. At a threshold of 0, the accuracy of the system is equal to that without a user trust model, since all the user tags are propagated. In this case the accuracy of the system is 34%. The figure also shows the average user trust ratio of 46%, which is the same as the accuracy when the users tag all the images in the dataset (1,320 images) and not only 66 images. Therefore, if we consider a large social network system where landmarks and users are selected in a way that each landmark is annotated by each user, our system shows that the best performance is achieved by choosing the most trusted user and propagating his/her annotations through the whole database of images. More precisely, in our dataset, the user annotates 1,320/66 = 20 times less images and the performance of the system ($RR$) increases from value of 46–65%. As a conclusion, by using the proposed model less manual tagging is needed, while the performance of the system increases significantly.

Figure 11 illustrates the relationship between the accuracy of the tag propagation system and the number of propagated tags by plotting them against each other. The maximum number of propagated tags can be much higher than the number of images, since several tags can be assigned to an image by different users. The black marker indicates the average tagging accuracy of the system without the user trust model and tag propagation presented in this paper. In this case, if users tag 44 × 66 = 2,904 photos (44 users in our experiments and each of them tags 66 images), the average accuracy of 46% can be achieved. This is equivalent to what we currently have in Flickr or Panoramio, where users simply tag photos independently and these tags are not being propagated. However, by introducing a user trust model and tag propagation into the system, we can improve the accuracy of the system and propagate more correct tags to untagged images in the dataset. This is depicted with the left part of the blue curve, which is above the average user trust ratio, we can still

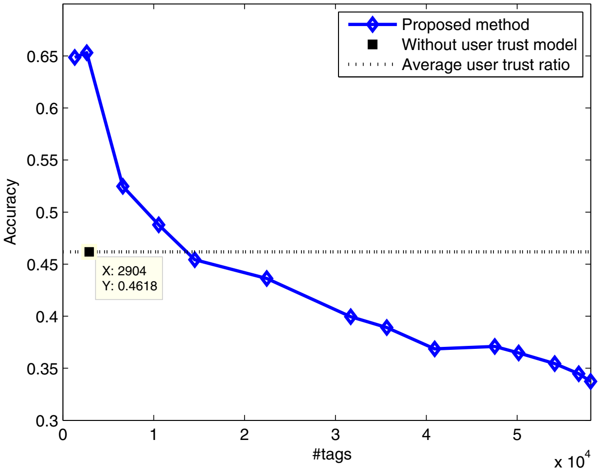


**Fig. 11** The recognition rate of the geotag propagation system versus the number of the propagated tags

propagate more than 10,000 tags from trusted users, while keeping accuracy higher than 46%.

## 5 Conclusion

Social networks are gaining popularity for sharing interests and information. Especially photo sharing and tagging is becoming more and more popular. Among other tags, geotags in form of geographical locations provide efficient information for grouping or retrieving images. Since manual annotation of these tags is time consuming, automatic tag propagation based on visual similarity offers a very interestingly good solution.

In this work, we have developed an efficient system for automatic geotag propagation by associating locations with distinctive landmarks and using object duplicate detection for tag propagation. The adopted graph based approach reliably establishes the correspondence between a small set of tagged images and a large set of untagged images. Based on this correspondences and a user trust ratio derived for each user, only reliable geotags are propagated. This leads to an increased accuracy of the tag propagation and a decrease of tagging efforts.

For assessing the quality of the tag propagation system we have considered two scenarios. First, we have analyzed the performance of the tag propagation alone which leads to a promising average accuracy of 71% over all the landmarks. Furthermore, we have shown that the performance varies considerably among different landmark types depending on their visual characteristics. Second, we have analyzed the influence of errors within the tag annotation which causes wrong tags to be propagated in the database. By considering a simple user trust model the accuracy of the system could be considerably improved. In this way, the proposed user trust model can be generalized to photo sharing platforms such as Panoramio or Flickr.

Currently the user trust model relies on predefined ground truth to estimate the user trust ratios. Future work will focus on automatic ground truth generation using WordNet and Wikipedia to obtain tagged image samples. In addition, we will work on developing new approaches for creating user trust models by considering user trust values assigned by particular user to other users and employing some ranking algorithm such as PageRank.

**Acknowledgements** This work was supported by the Swiss National Foundation for Scientific Research in the framework of NCCR Interactive Multimodal Information Management (IM2), the Swiss National Science Foundation Grant “Multimedia Security” (number 200020-113709), and partially supported by the European Network of Excellence PetaMedia (FP7/2007-2011).

## References

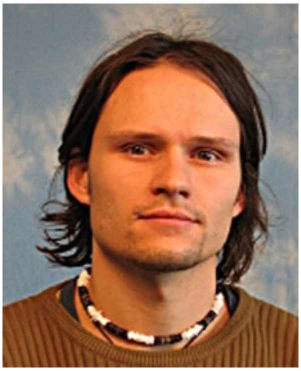
1. Ballard DH (1981) Generalizing the hough transform to detect arbitrary shapes. Pattern Recogn 13(2):111–122
2. Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. Comput Netw ISDN Syst 30(1–7):107–117
3. Cao L, Luo J, Huang T (2008) Annotating photo collections by label propagation according to multiple similarity cues. In: Proceeding of the 16th ACM international conference on multimedia (ACM MM 2008), pp 121–130

4. Cao L, Yu J, Luo J, Huang T (2009) Enhancing semantic and geographic annotation of web images via logistic canonical correlation regression. In: Proceedings of the 17th ACM international conference on multimedia (ACM MM 2009), pp 125–134
5. Carboni D, Sanna S, Zanarini P (2006) GeoPix: image retrieval on the geo web, from camera click to mouse click. In: Proceedings of the 8th ACM international conference on human-computer interaction with mobile devices and services (Mobile HCI 2006), pp 169–172
6. Crandall D, Backstrom L, Huttenlocher D, Kleinberg J (2009) Mapping the world's photos. In: Proceedings of the 18th international conference on World Wide Web (WWW 2009), pp 761–770
7. Facebook Statistics. Available at: http://www.facebook.com/press/info.php?statistics
8. Fellbaum C (ed) (1998) WordNet an electronic lexical database. MIT Press, Cambridge, London
9. Gammeter S, Bossard L, Quack T, Van Gool L (2009) 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)
10. Hays J, Efros AA (2008) IM2GPS: estimating geographic information from a single image. In: Proceedings of the IEEE international conference on computer vision and pattern recognition (CVPR 2008), pp 1–8
11. Hollenstein L, Purves R (2010) Exploring place through user-generated content: using Flickr to describe city cores. Journal of Spatial Information Science (JOSIS) 1:1–29
12. International Press Telecommunications Council (2009) IPTC photo metadata standard, IPTC Core 1.1 and IPTC Extension 1.1. Tech. rep.
13. Jøsang A, Ismail R, Boyd C (2007) A survey of trust and reputation systems for online service provision. Decis Support Syst 43(2):618–644
14. Kennedy LS, Naaman M (2008) Generating diverse and representative image search results for landmarks. In: Proceedings of the 17th international conference on World Wide Web (WWW 2008), pp 297–306
15. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
16. Marti S, Garcia-Molina H (2006) Taxonomy of trust: categorizing P2P reputation systems. Comput Netw 50(4):472–484
17. Massa P, Avesani P (2005) Controversial users demand local trust metrics: an experimental study on Epinions.com community. In: Proceedings of the international conference on artificial intelligence (IJCAI 2005), pp 121–126
18. Mikolajczyk K, Schmid C (2002) An affine invariant interest point detector. In: Proceedings of the 7th European conference on computer vision (ECCV 2002), pp 128–142
19. Nister D, Stewenius H (2006) Robust scalable recognition with a vocabulary tree. In: Proceedings of the IEEE international conference on computer vision and pattern recognition (CVPR 2006), pp 2161–2168
20. Quack T, Leibe B, Van Gool L (2008) 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), pp 47–56
21. Sahami M, Dumais S, Heckerman D, Horvitz E (1998) A Bayesian approach to filtering junk e-mail. Tech. Rep. WS-98-05, AAAI-98 workshop on learning for text categorization
22. Sigurbjörnsson B, van Zwol R (2008) Flickr tag recommendation based on collective knowledge. In: Proceeding of the 17th international conference on World Wide Web (WWW 2008), pp 327–336
23. Technical Standardization Committee on AV & IT Storage Systems and Equipment (2002) Exchangeable image file format for digital still cameras: exif Version 2.2. Tech. Rep. JEITA CP-3451
24. Vajda P, Dufaux F, Minh TH, Ebrahimi T (2009) Graph-based approach for 3D object duplicate detection. In: Proceedings of the international workshop on image analysis for multimedia interactive services (WIAMIS 2009), pp 254–257
25. Vajda P, Goldmann L, Ebrahimi T (2009) Analysis of the limits of graph-based object duplicate detection. In: Proceedings of the international symposium on multimedia, pp 600–605
26. Wikipedia—Flickr. Available at: http://en.wikipedia.org/wiki/Flickr
27. Wu L, Yang L, Yu N, Hua X (2009) Learning to tag. In: Proceedings of the 18th international conference on World Wide Web (WWW 2009), pp 361–370
28. Zheng Y, Zhao M, Song Y, Adam H, Buddemeier U, Bissacco A, Brucher F, Chua T, Neven H (2009) 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), pp 1085–1092

**Ivan Ivanov** is a research assistant and PhD student in Multimedia Signal Processing Group at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. His research interests include multimedia content annotation and tag propagation by merging textual tags and visually similar objects in social network collections of images and videos. He received Dipl.-Ing. (M.Sc.) degree in Electrical Engineering from the University of Belgrade, Serbia, in 2006. He worked as a Hardware Design Engineer for Texas Instruments, France, where he participated in the development of low-power VLSI multimedia applications for portable devices. After that, he worked as a Radio Access Network Conceptual Planning Expert in Vip mobile, Serbia, focusing on the definition of national 2G & 3G radio access network parameters and the implementation of new radio access technologies. He is a student member of IEEE.

**Peter Vajda** received his M.Sc. degree in Computer Science from the Vrije Universiteit, Amsterdam, Netherlands, in July 2006 and in Program Designer Mathematician from Eötvös Loránd University, Budapest, Hungary, in July 2007. He performed his diploma work on selection mechanisms in evolution computing and on using prediction algorithms in human-computer interaction. Since September 2007, he is research assistant and PhD student in Professor Ebrahimi's group at Ecole Polytechnique Fédéral de Lausanne (EPFL), Lausanne, Switzerland. He is involved in several European projects such as, Visnet II Network of Excellence, K-Space and PetaMedia. His research interests include mobile visual search and multimedia content analysis in still image and video. He is a student member of IEEE.

**Jong-Seok Lee** is a research scientist in the Multimedia Signal Processing Group (MMSPG), EPFL, Lausanne, Switzerland. He received the Ph.D. degree in Electrical Engineering and Computer Science in 2006, from KAIST, Daejeon, Korea. He was an adjunct professor at the School of Electrical Engineering and Computer Science, KAIST, in 2007. He is the author or co-author of more than 30 research publications. His research interests include multimedia signal processing, multimedia quality assessment, human-computer interaction, and machine learning.



**Lutz Goldmann** is a research scientist in the Multimedia Signal Processing Group (MMSPG), EPFL, Lausanne, Switzerland. He received his Dipl.-Ing. (M.Sc.) degree in Electrical Engineering from the Technical University of Dresden (TUD), Germany in 2002 and the Dr.-Ing. (Ph.D) degree in Electrical Engineering from the Technical University of Berlin (TUB), Germany in 2009. In 2002 he joined Siemens CT IC2, Munich, Germany as a research student where he developed image enhancement techniques for video coding artifact removal. Between 2003 and 2008 he worked as a research assistant at the Technical University of Berlin (TUB), Germany on the detection and recognition of humans within images and videos. He was actively involved in several national and European projects, such as GraVis, VISNET, 3DTV, K-Space and VISNET II. He is the author or co-author of more than 20 research publications. His research interests include 2D and 3D image and video analysis, multimedia quality assessment, and machine learning.

**Touradj Ebrahimi** is Professor at EPFL heading its Multimedia Signal Processing Group. He is also adjunct Professor with the Center of Quantifiable Quality of Service at Norwegian University of Science and Technology (NTNU). Prof. Ebrahimi has been the recipient of various distinctions and awards, such as the IEEE and Swiss national ASE award, the SNF-PROFILE grant for advanced researchers, Four ISO-Certificates for key contributions to MPEG-4 and JPEG 2000, and the best paper award of IEEE Trans. on Consumer Electronics. He became a Fellow of the international society for optical engineering (SPIE) in 2003. His research interests include still, moving, and 3D image processing and coding, visual information security (rights protection, watermarking, authentication, data integrity, steganography), new media, and human computer interfaces (smart vision, brain computer interface). He is the author or the co-author of more than 200 research publications, and holds 14 patents.