

# Comparative Study of Trust Modeling for Automatic Landmark Tagging

Ivan Ivanov, Peter Vajda, Pavel Korshunov, and Touradj Ebrahimi

**Abstract**—Many images uploaded to social networks are related to travel, since people consider traveling to be an important event in their life. However, a significant amount of travel images on the Internet lack proper geographical annotations or tags. In many cases, the images are tagged manually. One way to make this time-consuming manual tagging process more efficient is to propagate tags from a small set of tagged images to the larger set of untagged images automatically. In this paper, we present a system for automatic geotag propagation in images based on the similarity between image content (famous landmarks) and its context (associated geotags). In such a scenario, however, an incorrect or a spam tag can damage the integrity and reliability of the automated propagation system. Therefore, for reliable geotags propagation, we suggest adopting a user trust model based on social feedback from the users of the photo-sharing system. We compare this socially-driven approach with other user trust models via experiments and subjective testing on an image database of various famous landmarks. Results demonstrate that relying on user feedback is more efficient, since the number of propagated tags more than doubles without loss of accuracy compared to using other models or propagating without trust modeling.

**Index Terms**—Geotags, object duplicate detection, social networks, social tagging, spam combatting, tag propagation, trust modeling.

## I. INTRODUCTION

SOCIAL networks and photo sharing websites have become increasingly popular in recent years, since people use them to interact with each other by sharing their own interests or activities and exploring shared content (e.g., photo, video, text, and audio) of others. This sharing trend has resulted in a continuously growing volume of publicly available photos on Flickr<sup>1</sup>, Picasa<sup>2</sup>, and Facebook<sup>3</sup>. For instance, 219 billion photos have been uploaded on Facebook since 2005 [1]. Tagging is one

of the popular mechanisms that helps managing large volume of photos. Tags, when combined with search technologies, are essential in resolving user queries targeting shared photos. 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 decreases the precision of search, because photos without proper annotations are typically much harder to retrieve than correctly annotated photos.

A significant subset of shared photos in social networks or photo sharing websites is travel related. Travel is an important type of event for which people like to share, annotate, and search pictures. Usually, the most salient region in a travel related image corresponds to a specific landmark or object. When users annotate such images, they link a geotag to the object depicted in the image. For the majority of travel images on the Internet, however, proper geographical annotations are not available. In order to speed up the time-consuming manual tagging process, geotags can be propagated based on the similarity between image content (usually famous landmarks) and context (associated geotags).

One important challenge in tagging is to identify most appropriate tags for given content, and at the same time, to eliminate noisy or spam tags. Shared photos can be associated with inappropriate tags for several reasons. First of all, users are human beings and make mistakes. Moreover, it is possible to provide wrong tags on purpose for advertisement, self-promotion, or to increase the rank of a particular tag in automatic search engines. Finally, wrong machine tags, such as longitude and latitude, can be automatically assigned to images captured with cameras equipped with GPS devices due to bad or noisy communication channels with GPS satellites or wireless access points. Kennedy *et al.* [2] analyzed the Flickr website and revealed that the tags provided by users are often imprecise and only around 50% of tags are correctly assigned to images. Fig. 1 shows examples of imprecise or spam tags and incorrect geotags found on Flickr website. Beside the tag-photo association, spam objects can take other forms, i.e., possibly manifesting as spam photos or spam users (spammers). Therefore, for the practical tag propagation system, it is important to consider user trust information derived from users' tagging behavior. Trust provides a natural security policy stipulating that users or photos with low trust values should be investigated or eliminated.

In this paper, we consider traveling as one of the human activities highly influenced by social media in recent years [3]. As majority of travelers read blogs and reviews, and watch photos to select destinations, it is important to increase photos "findability" through geotagging of famous landmarks in images.

Manuscript received June 12, 2012; revised October 26, 2012; accepted January 06, 2013. Date of publication January 25, 2013; date of current version May 16, 2013. This work was supported by the Swiss National Foundation for Scientific Research in the framework of NCCR Interactive Multimodal Information Management (IM2), and by the Swiss National Science Foundation Grant "Multimedia Security" (200020-113709). The associate editor coordinating the review of this manuscript and approving it for publication was Yan Sun.

I. Ivanov, P. Korshunov, and T. Ebrahimi are with the École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (e-mail: ivan.ivanov@epfl.ch; pavel.korshunov@epfl.ch; touradj.ebrahimi@epfl.ch).

P. Vajda is with Stanford University, Stanford, CA 94305-9510 USA (e-mail: pvajda@stanford.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2013.2242889

<sup>1</sup><http://www.flickr.com>

<sup>2</sup><http://picasa.google.com>

<sup>3</sup><http://www.facebook.com>

TABLE I  
SUMMARY OF REPRESENTATIVE RECENT TECHNIQUES THAT COMBINE GEOGRAPHICAL CONTEXT AND VISUAL CONTENT FOR AUTOMATIC GEOTAGGING OF IMAGES

Reference	Descriptor	Method	Application
Hays and Efros [10]	visual features	the probability distribution for the location of an unknown image is found on the globe using a purely data-driven scene matching	non-landmark (scene) location recognition
Kennedy and Naaman [11]	visual and textual features	for a given location diverse and representative images are generated based on geotagged community images	visual summary of landmarks
Zheng <i>et al.</i> [12]	visual features & GPS coordinates	travel blogs and geotagged images are analyzed and a list of tourist landmarks is established based on the information from nearest neighbors	landmark recognition
Quack <i>et al.</i> [13]	visual and textual features & GPS coordinates	objects and events are retrieved from a large-scale collection of geotagged images using pair-wise similarity	event/scene understanding

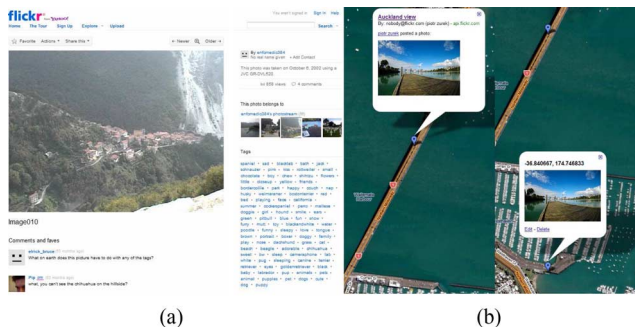


Fig. 1. Examples of imprecise or spam tags and incorrect geotags in Flickr: (a) wrong tags—only a few tags in the list are related to the image, while the rest is irrelevant (e.g., yellow, love, doggy), (b) incorrect geotags—the left side of the picture shows the point placed at incorrect location acquired from GPS-enabled camera, while the right side shows the manually created Google Maps point with correct data and the incorrect point slightly above on the bridge.

For accurate and reliable automatic geotagging, we previously proposed to combine the object (e.g., famous landmark) duplicate detection with the user trust modeling [4]. The user trust modeling reduces the risk of propagating wrong tags caused by spamming or faulty annotation. In a real-life scenario, an image with unknown landmark will be automatically tagged with either one geotag or none, depending on the level of similarity with the known (trained) landmarks. This scenario is denoted as open set problem in our previous work [4]. The less complex scenario is closed set problem, where each test image depicting unknown landmark is assumed to correspond to exactly one of the known landmarks. Therefore, the test image gets assigned to the most probable trained landmark and the corresponding geotag is propagated to the test image. Although we use the proposed automatic geotagging system, in this paper we do not focus on its evaluation of computational efficiency, nor on comparison of different geotagging methods.

The focus of this paper is on comparison of several techniques for trust modeling in social tagging systems, namely Koutrika *et al.* [5], Liu *et al.* [6], Xu *et al.* [7], Krestel and Chen [8], and Ivanov *et al.* [4]. The last trust model is our approach, which was proposed together with the system architecture for automatic geotagging applied in this paper. In this paper, we compare our socially-driven user trust model with four other models. The effectiveness of these models is demonstrated through a set of experiments on an image database with different landmarks. To create an environment for trust modeling where object duplicate detection performs well, we consider only closed

set problem. This constrain allows us to thoroughly analyze and compare trust models, while having less errors introduced in the system by the object duplicate detection module. Using robust trust model and robust object duplicate detection, we can create more accurate system. Spam in social systems can not be controlled, however, the performance of object duplicate detection can be improved, which will lead to more accurate system for automatic geotagging with trust modeling.

The remaining sections of this paper are organized as follows. We introduce related work of geotagging in Section II. Selected trust models are presented in Section III. Section IV summarizes our approach for geotag propagation between images. Evaluation methodology and results for comparison of trust models are presented in Sections V and VI, respectively. Finally, Section VII concludes the paper with a summary and some perspectives for the future work.

## II. GEOTAGGING IN SOCIAL NETWORKS AND SHARING WEBSITES

The proposed system is related to different research fields including visual analysis, geographic information systems, and social networking and tagging systems. Therefore, the goal of this section is to review the most relevant work in the fields of joint analysis of visual content and geographical context, and manual tagging, while the next section provides insight into trust modeling in social tagging systems.

In the last several years, an important trend in multimedia understanding is modeling and extracting value from geographical context, such as GPS coordinates, and visual content, such as photo description. Different research problems and significant approaches in this field are summarized by Luo *et al.* [9]. In this section, we focus on some of the representative image retrieval approaches that rely on a variety of image or landmark descriptors combined with geographic information. These approaches are summarized in Table I.

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, the algorithm finds the visual nearest neighbors in the database and estimates a geolocation of the image from the GPS coordinates of the 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.

Kennedy and Naaman [11] 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. The visual features (global color and texture features, and Scale Invariant Feature Transform (SIFT)) are analyzed to cluster landmark images into visually similar groups. The method has been proven to be effective for extraction of the representative image sets for a given landmark. But since it cannot be applied to untagged images, its applicability is limited. The recent work of Zheng *et al.* [12] automatically finds frequently photographed landmarks from a large collection of geotagged photos. The authors 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<sup>4</sup>, and visually cluster photos gathered by querying Google Images<sup>5</sup>. However, the test set they use is quite limited—728 images in total for a 124-category problem, or less than 6 test images per landmark.

Another application that combines textual and visual techniques has been proposed by Quack *et al.* [13]. The authors developed a system that crawls photos on the internet and identifies clusters of images referring to a common object (physical items at 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>6</sup> articles are attached to the images and the validity of these associations is checked. Gammeter *et al.* [14] 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 [13] and [14], GPS coordinates are used to precluster objects which may not be always available.

A commercial application by Google, called Google Goggles<sup>7</sup>, is created for landmark search on mobile phones. It also detects logos, book and DVD covers, artworks, and products.

Most of the photo sharing websites (e.g., Flickr, Picasa, Panoramio<sup>8</sup>, Zoomr<sup>9</sup>), 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) [15], International Press Telecommunications Council (IPTC) [16]), or manually annotated via geocoding.

The main disadvantages of the above systems is that they rely on GPS coordinates to derive geographical annotation, which is

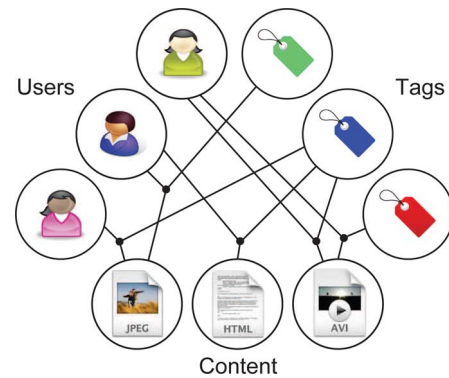


Fig. 2. General model of a social tagging system is represented as a tripartite graph structure which includes three kinds of nodes (objects): *users*, *content*, and *tags*. An edge linking a user, a tag, and a content represents a tag assignment.

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 [17] indicated that geotagging should follow the way people actually describe locations, i.e., it is more convenient to use: Church of Saint Sava in Belgrade, rather than: latitude 44.798083 and longitude 20.46855. Therefore, there is a growing interest in the research community to derive geographic locations of the scenes in photos based on visual and text features.

### III. TRUST MODELING IN GEOTAGGING APPLICATIONS

When information is exchanged on the Internet, malicious individuals are everywhere trying to take advantage of the information exchange structure for their own benefit, while bothering and spamming others. In this section, we present and discuss several techniques for combatting noise and spam through trust modeling in social tagging systems. First, we introduce the model of a social tagging system. Then, we present in details the five recent techniques for user trust modeling that are suitable for a specific application of geotagging and can be used in geotagging the shared content and efficient propagation of such tags throughout the untagged content.

The entities (or objects) that make up the model of a social tagging system [18] are shown in Fig. 2. The model consists of *users* who interact with the system, *content* (resources or documents), which is a piece of information such as photo, video, textual document, or web pages; and *tags*, the descriptions assigned to the piece of the content by users. The action of associating a tag to a content by a user is usually referred to as *tag assignment* [6]. Depending on the system under consideration, a user can assign one or several tags to each content.

Trust modeling methods can be categorized into two classes according to the target of trust, i.e., content and user trust modeling, as we previously reported in [19]. *Content trust modeling*

<sup>4</sup><http://www.wikitravel.com>

<sup>5</sup><http://images.google.com>

<sup>6</sup><http://www.wikipedia.org>

<sup>7</sup><http://www.google.com/mobile/goggles>

<sup>8</sup><http://www.panoramio.com>

<sup>9</sup><http://www.zoomr.com>

TABLE II  
SUMMARY OF FIVE TRUST MODELING TECHNIQUES USED FOR COMBATTING NOISE AND SPAM IN SOCIAL TAGGING SYSTEMS

Reference	Content	Method	Dataset
Koutrika <i>et al.</i> [5]	bookmarks	a coincidence-based model for query-by-tag search, which estimates the level of agreement among different users in the system for a given tag	Delicious, real & simulated
Liu <i>et al.</i> [6]	bookmarks	an iterative approach to identify spam content by its information value extracted from the collaborative knowledge	Delicious, real
Xu <i>et al.</i> [7]	bookmarks	an iterative approach to compute the goodness of each tag with respect to a content and the authority scores of the users	MyWeb 2.0, real
Krestel and Chen [8]	bookmarks	a TrustRank-based approach using features which model tag co-occurrence, content co-occurrence and co-occurrence of tag-content	BibSonomy, real
Ivanov <i>et al.</i> [4]	images	an approach based on the feedback from other users who agree or disagree with a tag associated with an image	Panoramio, real

TABLE III  
NOTATIONS USED TO DESCRIBE TRUST MODELS CONSIDERED IN THIS PAPER

Notation	Explanation
$U$	Set of users
$D$	Set of documents (content)
$T$	Set of tags
$P \in U \times D \times T$	Set of tag assignments
$u \in U$	A user
$d \in D$	A document (content)
$t \in T$	A tag
$p \in P$	A tag assignment
$trust^{model}(u)$	Trust value of a user $u$ calculated for particular model
$L$	Total number of users
$M$	Number of training documents (photos)
$N$	Number of test photos
$ x : condition $	Number of $x \in \{u, d, t, p\}$ which satisfy condition

is to classify content as spam or legitimate. In this case, the target of trust is a content, and thus a trust value is given to each content. In *user trust modeling*, trust is given to each user based on the information extracted from a user's account, his/her interaction with other participants within the social network, and/or the relationship between the content and tags that the user contributed to the social network. Given a user trust value, the user might be flagged as a legitimate user or spammer. Table II summarizes five user trust models, which we then describe in more details (in the same order as they are presented in the table). And Table III summarizes the notations used for their detailed description. These methods are different in the targeted media content, for which the geotagging is used, the applications they are intended for, and the required level of participation from the users of the geotagging system.

#### A. A Coincidence-Based Model

Koutrika *et al.* [5] were the first to explicitly discuss methods of tackling spamming activities in social tagging systems. The authors studied the impact of spamming through a framework for modeling social tagging systems and user tagging behavior. They proposed a method for ranking content matching a tag based on taggers' reliability in social bookmarking service Delicious. Their coincidence-based model for query-by-tag search estimates the level of agreement among different users in the system for a given tag. A bookmark is ranked high if it is

tagged correctly by many reliable users. A user is more reliable if his/her tags more often coincide with other users' tags.

In more formal way, the following calculations are performed:

$$c(u) = \sum_{d,t:\exists P(u,d,t)} \sum_{u_i \in U: u_i \neq u} |p : \exists P(u_i, d, t)| \quad (1)$$

$$score(d, t) = \frac{\sum_{u:\exists P(u,d,t)} c(u)}{\sum_{u \in U} c(u)} \quad (2)$$

$$trust^{Koutrika}(u) = \sum_{d,t:\exists P(u,d,t)} score(d, t) \quad (3)$$

where  $c(u)$ , coincidence factor of the user  $u$ , is the number of other users  $u_i$  who assigned the same tag  $t$  to the same document  $d$  as the user  $u$  did. Score of the document  $d$  with respect to the tag  $t$ , denoted as  $score(d, t)$ , is calculated as a normalized value of  $c$  over all users who assigned  $t$  to  $d$ . Finally, a trust value of the user  $u$ ,  $trust^{Koutrika}(u)$ , is the sum of  $score(d, t)$  over all tag assignments by  $u$ .

Koutrika *et al.* performed a variety of evaluations of their trust model on controlled (simulated) dataset by populating a tagging system with different user tagging behavior models, including a good user, bad user, targeted attack model and several other models. Using controlled data, interesting scenarios that are not covered by real-world data could be explored. It was shown that spam in tag search results using the coincidence-based model is ranked lower than in results generated by, e.g., a traditional occurrence-based model, where content is ranked based on the number of posts that associate the content to the query tag.

#### B. A Wisdom of Crowds Model

Liu *et al.* [6] proposed a simple but effective approach for detecting spam content in Delicious, by harvesting the wisdom of crowds. An information value of a bookmark is defined as the average number of times that each tag of the content is assigned by different users. A low information value of a bookmark indicates a divergence from crowds, which can be considered as a spam content. Furthermore, this method was extended to user trust modeling by aggregating the information values for each user.

All measures are defined as follows:

$$it(d, t) = \frac{|u : \exists P(u, d, t)|}{\sum_{t' \in T} |u : \exists P(u, d, t')|} \quad (4)$$

$$ic(u, d) = \frac{\sum_{t: \exists P(u, d, t)} it(d, t)}{|t: \exists P(u, d, t)|} \quad (5)$$

$$I(d) = \frac{|u: \exists P(u, d, \cdot)|}{\sum_{d' \in D} |u: \exists P(u, d', \cdot)|} \quad (6)$$

$$trust^{Liu}(u) = \sum_{d: \exists P(u, d, \cdot)} I(d) \cdot ic(u, d), \quad (7)$$

where  $it(d, t)$  represents the tag's  $t$  tagging information value with respect to document  $d$ , and  $ic(u, d)$  is the information value of the content (document)  $d$  with respect to user  $u$ . The importance of the document  $d$  is defined by  $I(d)$ . Finally, a trust value of the user  $u$ ,  $trust^{Liu}(u)$ , is calculated as the weighted average of the information value of the content tagged by user  $u$ , with the importance of the document as weight.

An interesting point is that, for the time being, Liu *et al.* collected the largest dataset for trust modeling by crawling Delicious [19]. This dataset had around 82 thousand users, 1.1 million tags, 9.3 million bookmarks and 17.4 million tag-bookmark associations.

### C. An "Authority" Model Based On Goodness of Tags

Xu *et al.* [7] introduced the concept of "authority" in social bookmarking systems, where they measured the goodness of each tag with respect to a content by the sum of the authority scores of the users who have assigned the tag to the content. Authority scores and goodness are iteratively updated by using Hyperlink-Induced Topic Search (HITS) algorithm, which was initially used to rank web pages based on their linkage on the web [20].

Following measures are defined and iteratively calculated:

$$s_{i+1}(d, t) = \sum_{u: \exists P(u, d, t)} trust_i^{Xu}(u) \quad (8)$$

$$trust_i^{Xu}(u) = \frac{\sum_{d, t: \exists P(u, d, t)} s_i(d, t)}{|t: \exists P(u, \cdot, t)|}, \quad (9)$$

where  $i \in [1 \dots Q]$ ,  $s_i(d, t)$  is the goodness of each tag  $t$  with respect to a content  $d$ , and  $trust_i^{Xu}(u)$  represents a trust value (authority score) of the user  $u$ . Initial settings in this iterative approach are:  $s_0(d, t) = 0, \forall t, d$  and  $trust_0(u) = 1, \forall u$ . The number of iterations is set to  $Q = 100$ .

### D. A Cooccurrence Model

In contrast to the approach of Xu *et al.* [7], Krestel and Chen [8] iteratively updated values for users only. The authors proposed to use a spam value propagation technique to propagate trust values through a social graph in BibSonomy, where edges between nodes (in this case, users) indicate the number of common tags supplied by users, common content annotated by users and/or common tag-content pairs used by users. Starting from a manually assessed set of nodes labeled as spammers or legitimate users with the initial spam values, a TrustRank metric is used to calculate and iteratively update spam values for all users. TrustRank metric is previously introduced in [21]

to semiautomatically separate reputable from spam web pages. This metric relies on an important empirical observation called approximate isolation of the good set: good pages seldom point to bad ones.

All measures are calculated as follows:

$$W(u_1, u_2) = |t: \exists P(u_1, \cdot, t), P(u_2, \cdot, t)| + |d: \exists P(u_1, d, \cdot), P(u_2, d, \cdot)| + |d, t: \exists P(u_1, d, t), P(u_2, d, t)| \quad (10)$$

$$Tr(u_1, u_2) = \frac{W(u_1, u_2)}{\sum_{v \in U} W(u_1, v)} \quad (11)$$

$$trust_i^{Krestel}(u) = \alpha \cdot \sum_{v \in U} Tr(u, v) \cdot trust_{i-1}^{Krestel}(v) - (1 - \alpha)d(u), \quad (12)$$

where  $i \in [1 \dots Q]$ ,  $W(u_1, u_2)$  is the weight of the edge between users  $u_1$  and  $u_2$  in the social graph and  $Tr(u_1, u_2)$  is the corresponding transition matrix. A trust value of the user  $u$ ,  $trust_i^{Krestel}(u)$ , is iteratively calculated. Initial setting in this iterative approach is:  $trust_0(u) = d(u), \forall u$ , where  $d(u)$  represents the trust values of the seed users. The number of iterations is set to  $Q = 100$ .

The approach of Krestel and Chen is more sophisticated than the approach of Xu *et al.* [7] in that multiple relationships, such as tag cooccurrence, content cooccurrence and tag-content cooccurrence, can be taken into account, rather than considering only the tag-content pairs shared by users.

### E. User Reliability Based Model

In this section, we describe our own approach for user trust modeling in image tagging, which was proposed in Ivanov *et al.* [4]. 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  $R(i, u)$ ,  $i \in [1 \dots M]$  and  $u \in [1 \dots L]$ , is defined as:

$$R(i, u) = \begin{cases} 1, & \text{if user } u \text{ tags image } i \text{ correctly;} \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

The process of comparing the propagated tags to ground truth tags can be done automatically using tag similarity measures, for example WordNet [22] or Google distance [23] measures. Nevertheless, we considered only manually defined ground truth for our experiments.

A trust value for user  $u$ ,  $trust^{Ivanov}(u)$ , is computed as the percentage of the correctly tagged images among all images tagged by user  $u$ :

$$trust^{Ivanov}(u) = \frac{\sum_{i=1}^M R(i, u)}{M}. \quad (14)$$

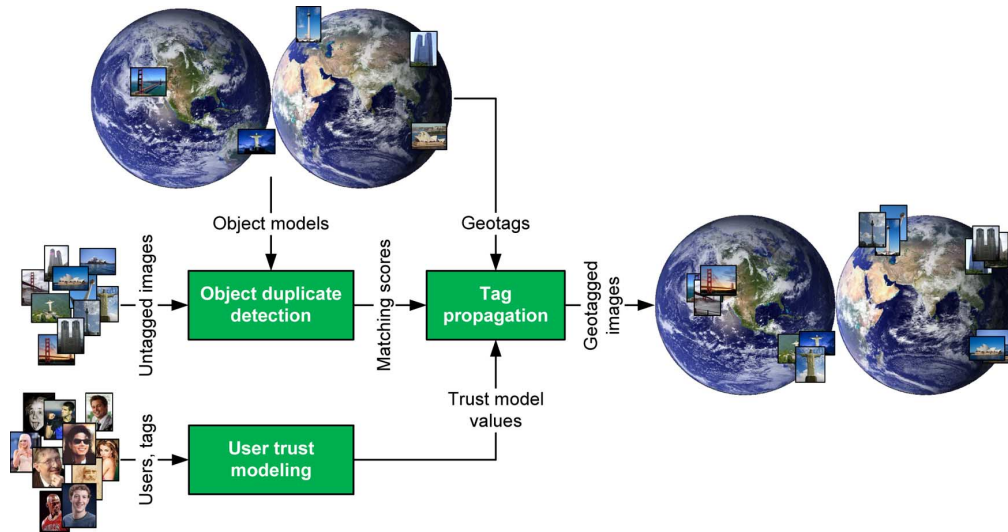


Fig. 3. Overview of the system for geotag propagation in images. The object duplicate detection is trained with a small set of images with associated geotags. The created object (landmark) models are matched against untagged images. The resulting matching scores serve as an input to the tag propagation module, which propagates the corresponding tags to the untagged images. Given a user trust model, only the tags from reliable users are propagated.

In this approach, ground truth data are used for the estimation of the user trust value. However, for a practical 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 value 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 the trusted 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 value 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 as described in details in [4].

#### IV. AN AUTOMATED LANDMARK TAGGING

Based on the user reliability trust modeling described in Section III-E, we built the solution for geotag propagation between images. The main innovation of such system is the

combination of object duplicate detection and user trust modeling for accurate and reliable geotag propagation. The system architecture has been proposed previously in [4] and is illustrated here in Fig. 3. It contains three functional modules, each of which has a specific task: object duplicate detection, tag propagation, and user trust modeling. As the focus of this paper is on trust modeling, the object duplicate detection [24] and tag propagation [4] modules are only summarized briefly below. The user trust modeling module is powered by alternating the five trust models described in Section III, and then the performance of the trust module and the whole system is measured, and results are reported and analyzed in Section VI.

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

##### A. 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. This is especially true for images related to travel, where tourists tend to take a lot of photos from different distances and view-points around a famous landmark. The basic idea of applying object duplicate detection for geotag propagation is that travel images 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 [25] and each of these regions is described using SIFT features [26]. These features are robust to arbitrary changes in viewpoints. Then, hierarchical k-means clustering [27] 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 in the test phase. Finally, a spatial graph model is constructed to improve the accuracy of the feature matching with a test image. The graph model 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. Feature matching is performed using a one-to-one nearest neighbor matching, where the hierarchical clustering is used to efficiently resolve the nearest neighbor search. 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 [28] 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], [29].

### B. Tag Propagation

The goal of the tag propagation module is to propagate the geotags from the tagged to the untagged 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), for example, Paris (Eiffel Tower) and Budapest (Parliament).

It was shown in [4] that tag propagation module supports two application scenarios: closed and open set problem. In this paper, we will consider only closed set problem to create an

environment for trust modeling where object duplicate detection performs well, as already explained in Section I. In the closed set problem, each test image is assumed to correspond to exactly one of the known (trained) landmarks. Therefore, the test image gets assigned to the most probable trained landmark, based on the matching scores provided by the object duplicate detection module, and the corresponding tag is propagated to the test image. The closed set problem is evaluated in Section V as recognition task, while the evaluation of the open set problem was presented as detection task in [4].

In a more detailed way, the tag propagation can be explained as follows. The object duplicate detection module provides a matching score matrix  $S_{i,j}$ . It represents the pair-wise comparison of the trained images (landmarks)  $i$ ,  $i \in [1 \dots M]$ , and the test images  $j$ ,  $j \in [1 \dots N]$ , where  $M$  and  $N$  are number of training and test images, respectively. Then we find the maximum score for each test image  $j$  and propagate the geotag of the corresponding training image  $i$ . The assignment matrix  $C_{i,j}$ ,  $i \in [1 \dots M]$  and  $j \in [1 \dots N]$ , is formed in the following way:

$$C_{i,j} = \begin{cases} 1, & \text{if } S_{i,j} = \max_{i \in [1 \dots M]} \{S_{i,j}\}; \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

In this case, each test image gets assigned with exactly one tag from the training photo dataset. Based on the assignment matrix  $C_{i,j}$  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.

### C. User Trust Modeling

The goal of the user trust modeling module is to measure trust or reliability of the users in geotagging.

Different users are introduced into the system. They perform annotation of a small set of images through geotagging. The trust module collects all tags from the users or their feedback on tags, and aggregate them to produce trust values. For a given trust model described in Section III, the user trust values  $trust(u)$ ,  $u \in [1 \dots L]$ , are calculated. Only tags from users who are trusted are propagated to other photos in the dataset. In other words, if the user trust value  $trust(u)$ , exceeds a predefined threshold  $\hat{T}$ , then all his/her tags are propagated. Otherwise, none of his/her tags are propagated. Also, based on this user trust value, the user might be flagged as a legitimate user, if  $trust(u) \geq \hat{T}$  or spammer, if  $trust(u) < \hat{T}$ .

As a final outcome of the whole system, only geotags coming from trusted users are propagated to the same landmark found in the nontagged set of images.

## V. EXPERIMENTS

The effectiveness of the proposed system for automatic geotagging based on user trust modeling is demonstrated through a set of experiments on an image database containing various landmarks. Furthermore, several techniques for trust modeling, described in Section III, are compared and contrasted to each other in detail.

### A. Dataset

A dataset of images was created by us as part of the reported work in [4] to evaluate the proposed geotag propagation method.



Fig. 4. Sample landmarks for each of the 22 cities within the dataset. The dataset covers a large variety of landmarks including buildings, bridges, monuments, etc.

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. Recognition rate for all landmarks. Each row represents one city from our dataset and the right three columns represent three landmarks in each city. The values in the first column are the average recognition rates for each city, sorted from highest to lowest.

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 1320 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, and Eiffel Tower). Each landmark has 20 image samples. Fig. 4 shows a single image for a single landmark from each of the 22 considered cities. Images with a large variety of view points and distances are considered for each landmark. Fig. 5 summarizes all cities and landmarks contained in the 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 used as test images.

We manually formed a ground truth data by assigning several tags describing landmarks depicted in images.

## B. Scenarios

The performance of the described geotag propagation method is evaluated and analyzed in two application scenarios: tag propagation and user trust scenario. 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 that do not contain the ground truth landmark, namely all images which depict one of the other 65 landmarks. This leads to  $19 \times 65 = 1235$  negative images in the test set. In the evaluation of this scenario in Section VI-A, we consider the closed set problem, while evaluation of the open set problem is described in more details in [4].

In the *user trust scenario*, we simulate a social network environment. As explained in Section III-E, due to the lack of a suitable dataset, which provides user geotags and feedback from photo sharing website Panoramio, the evaluation of the user trust scenario in this paper is based on the simulation of the social network environment. We recruited  $L = 47$  participants, among whom 66% were males and 34% were females, aged 16–63 (average age was 29), with different backgrounds (architects, researchers, engineers, doctors, high school students) and cultural differences (from 8 different countries located mostly in Europe). Participants were asked to tag  $M = 66$  photos from the dataset, putting the name of the landmark depicted in the image. We collected 3295 tags (658 of them were unique tags) and they were used to create different user trust models as per Section III. For the model of Krestel and Chen [8], 12 users were manually selected as reliable in geotagging. They were researchers and architects who are assumed to have reliable knowledge of landmarks due to frequent travels and educational background. After having created the user trust values  $trust(u)$ ,  $u \in [1 \dots L]$ , for all trust models described in Section III, we perform tag propagation based on those annotated images. Selected trust models are then compared.

## C. Evaluation

In this section, the evaluation methods for user driven tag propagation system are described. Both scenarios, tag propagation scenario and user trust scenario, are evaluated as a closed set problem.

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

$$RR = \frac{Tc}{A}. \quad (16)$$

First, ground truth matrix ( $GT$ ) is created for each test image  $j$  and of the corresponding training image  $i$ :

$$GT_{i,j} = \begin{cases} 1, & \text{if } Landmark(i) = Landmark(j); \\ 0, & \text{otherwise} \end{cases} \quad (17)$$



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

Second, for tag propagation scenario using object duplicate detection (ODD) method [24],  $Tc$  and  $A$  are calculated as:

$$Tc^{ODD} = \sum_{i,j} GT_{i,j} \cdot C_{i,j} \quad (18)$$

$$A = \sum_{i,j} GT_{i,j}, \quad (19)$$

where  $C_{i,j}$  assignment matrix was defined in Section IV-B.

Finally, for the user trust scenario, which combines the tag propagation method and user trust modeling,  $Tc$  and  $A$  are defined as:

$$Tc^{Trust} = \sum_{i,j,u:trust(u) \geq \hat{T}} GT_{i,j} \cdot C_{i,j} \cdot U_{i,u} \quad (20)$$

$$A = \left( \sum_{i,j} GT_{i,j} \right) \cdot \left( \sum_{u:trust^{model}(u) \geq \hat{T}} 1 \right), \quad (21)$$

where  $\hat{T}$  is the threshold for the user trust value, index  $u \in [1 \dots L]$  and  $L$  is the number of users. In other words,  $A$  is the number of tags (one for each image) for each trusted user and  $Tc^{Trust}$  is the correctly propagated tags among those. A propagated tag is considered correct only if the annotated tag was the same as the ground truth tag.

## VI. RESULTS

### A. Results of the Tag Propagation Scenario

Considering only object duplicate detection [24] in the tag propagation scenario without trust modeling, we compute the recognition rate for all landmarks in the dataset as shown in Fig. 5. Each field in the figure is highlighted with the color ranging in the spectrum from red to blue, with red corresponding to the recognition rate closer to 1 and blue to the rate closer to 0. First column of the figure shows the average recognition rate for each city, sorted from highest to lowest values. In the dataset, we have three landmarks for each city as reflected in the three right columns of the figure.

The performance of object duplicate detection varies considerably for different cities, but also across the individual landmarks within a city. To investigate these variations, all landmarks, according to common visual properties, were divided into different groups such as castles, churches, bridges, towers/statues, stadiums, and ground structure. Fig. 6 shows that average recognition rates vary considerably for each of these categories. With the average rate of 71% across all the landmarks, the object duplicate detection demonstrates the highest recognition rate for the castles' category and the lowest for the stadiums'. These results demonstrate that if we rely only on the object duplicate detection for tags propagation, on average, we can expect from the accuracy of propagation to be no more than 71%. This expectation, however, is based on the assumption that all propagated tags are correct and reliable, with no spam tags or tags mistakenly assigned to images. Therefore, in the extended more practical scenario, where we assume some images to be

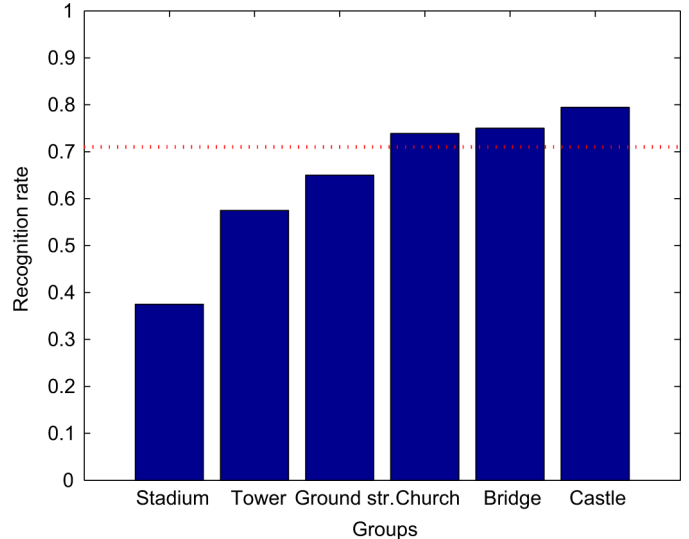


Fig. 6. Recognition rate across the different landmark categories in the closed set problem (bars) and the recognition rate of all landmarks (dashed line). Landmarks have been grouped according to their visual characteristics.

tagged wrongly, the accuracy of the propagation may decrease dramatically. To minimize the impact from the spam or mistaken tags on tag propagation, in the next section, we consider different user trust models and compare their efficiency in filtering out the wrong tags.

### B. Results of the User Trust Scenario

To compare different user trust models, we first analyze the distribution of their trust values given the manually assigned tags by the human participants (see Section V-B for more details). The values for each trust model were computed as described in Section III. Obtained user trust values were normalized to 1 for each trust model. Then, the trust values were split into five equally distributed histogram bins with the following ranges: 0 – 0.2, 0.2 – 0.4, 0.4 – 0.6, 0.6 – 0.8, and 0.8 – 1. Fig. 7 shows the distribution of the total number of users with trust values in different bins for each of the trust model. From the results, it can be noted that the distributions for most of the user trust models are not uniform. However, the tags to our dataset assigned by the human participants can be regarded following a uniform distribution, assuming, participants unbiasedly tagged the depicted generally well-known landmarks. Therefore, useful, adequate, and practical user trust model should also reflect this uniformity in the gathered tags from participants. From Fig. 7, we can notice that only two out of five compared user trust models, Koutrika *et al.* [5] and Ivanov *et al.* [4], demonstrate the uniformity in their assignment of the trust values to the participated users, while the rest of the models mark majority of the users as untrusted.

To understand the reasons for such bias in some of the user trust models, we plotted user trust values given by each approach to all users in Fig. 8. The figure shows that trust values vary considerably between different users, but also across different models. For example, trust values of the users enumerated with 2, 16 and 47 span almost the entire range of the normalized trust value, namely, from 0 to 1, for all selected models.

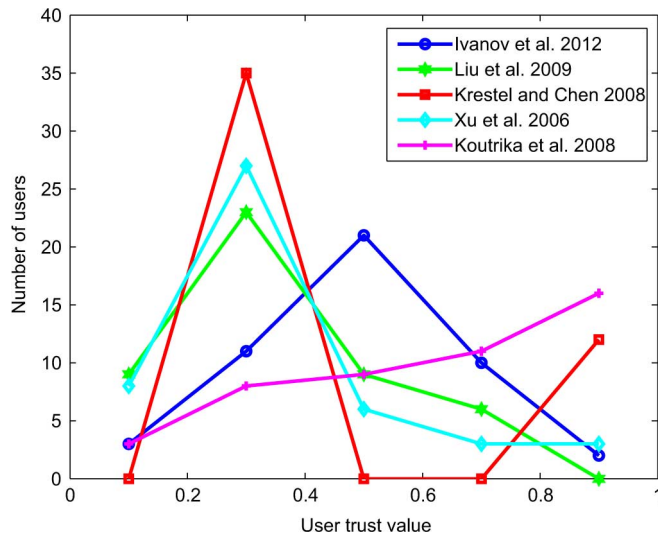


Fig. 7. Distribution of the normalized trust values for different user trust models. Different user trust models are depicted with different line colors and different markers. The results show a wide variety of distributions, mainly not uniform, which leads to the conclusion that users possess different knowledge in landmark recognition and thus people are more or less reliable in geotagging.

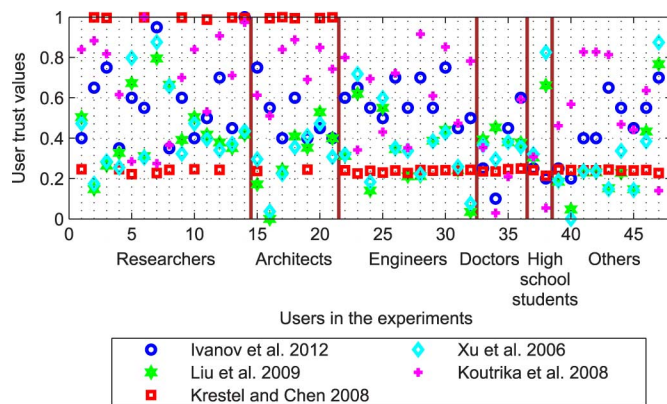


Fig. 8. Distribution of the normalized trust values for all users by different trust models. Each user has five trust values given by each of the trust models.

Since it is difficult to compare selected trust models for each user separately, we also grouped the participants according to their background. In our experiments, users are split into 6 different categories: researchers (13 users), architects (7 users), engineers (12 users), doctors (4 users), high school students (2 users), and others (9 users who did not indicate their background). By looking at the average trust levels for each group of users, as shown in Fig. 9, we can observe that the trust values from Ivanov *et al.*, Koutrika *et al.* are higher for researchers and architects than for engineers, doctors, and high school students. Researchers and architects had larger number of geotags, since they travel often and, therefore, have a good knowledge of landmarks and their locations. High school students had considerably fewer geotags. Therefore, the trust values of researchers and architects produced by Ivanov *et al.* and Koutrika *et al.* are high, since these models give higher trust values to users who have more common tags with other users. Doctors and high school students might travel less, and thus their trust values are

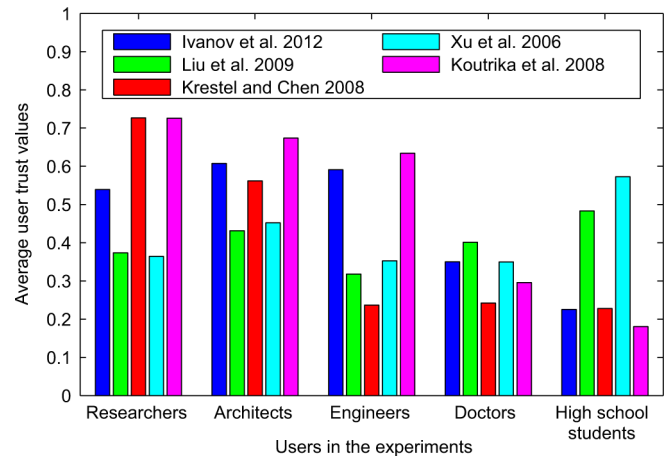


Fig. 9. Average normalized trust values for different groups of users. The user category “others” is not shown, since the background of these users was not known.

lower. On the other hand, models by Liu *et al.* [6] and Xu *et al.* [7] give larger trust values to the high school students. The reason is that these models assign higher trust weight to users who provide less tags in total but more of which are reliable, i.e., common with tags of the other users. One model that stands out of the rest is by Krestel and Chen [8]. Fig. 8 demonstrates that one group of users have the highest trust value 1, while the rest of users are given value about 0.2. The users with trust value 1 are the 12 users that were manually chosen to be the trusted users to ‘seed’ the algorithm by Krestel and Chen. The reason for such significant disparity in trust values between the trusted users and others is that Krestel and Chen’s model was designed for scenarios when most of the data is spam [8]. Their user trust model is very sensitive to the tags, which deviate from the tags by the original set of the trusted users. Therefore, this model demonstrates such significant bias when used with our set of geotags from the typical nonmalicious participants.

To analyze the performance of the studied user trust models in tags propagation scenario, we plot the accuracy of propagation versus the number of tags in Fig. 10. We vary the number of propagated tags by adjusting the threshold on what is the acceptable user trust value of each model. The accuracy of the tags propagation is computed as the ratio of the correct (based on the ground truth) tags to the total number of tags assigned (propagated) to images. 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. Each tag is propagated to different images. Therefore the curves in the Fig. 10 show a trade-off between propagating tags to more images but less accurately and propagating tags to less images but more accurately. The black marker indicates the average tagging accuracy of the system when neither the user trust model nor automated tag propagation is used. In our experiments, it corresponds to users assigning  $47 \times 66 = 3102$  tags to images (47 users in our experiments with each of them tagging 66 images at least once). The resulted average tagging accuracy is 52%. This accuracy is equivalent to what currently Flickr or Panoramio have, where users simply tag photos independently with no propagation used.

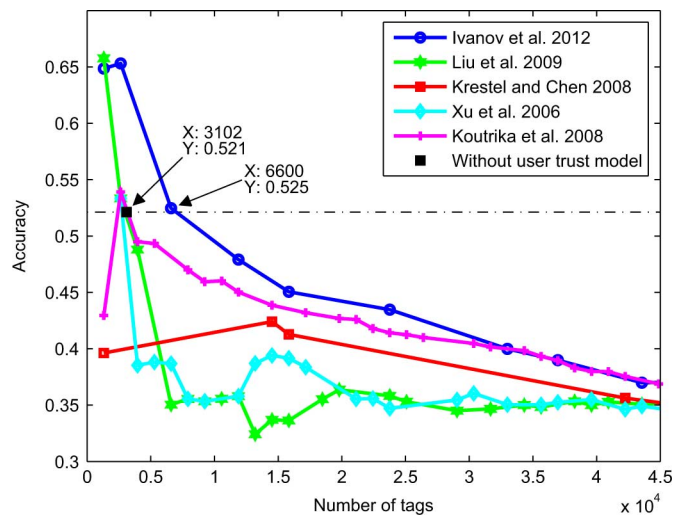


Fig. 10. Recognition rate of the geotag propagation system versus the number of the propagated tags. Different user trust models are depicted with different line colors and markers.

However, by using automated tag propagation that relies on the trust model based on user feedback, we can improve the accuracy of the tagging system and propagate more tags to the untagged images in the dataset. This improvement is illustrated by the left part of the blue curve (our method), which is above the average user trust value of 52%. It means that more than 6600 tags (see Fig. 10) can be propagated, twice more than without a trust model, from the trusted users, while keeping accuracy higher above 52%. Other trust-based methods, such as by Koutrika *et al.*, also perform well, though, they show less impressive results than the tag propagation based on our user trust model. However, the advantage of the algorithm by Koutrika *et al.* is that it is simple and does not need any ground truth or seed data. Methods of Liu *et al.*, Xu *et al.*, and Krestel and Chen are not able to perform well in the tag propagation scenario. Our method showed good performance in this simulated social network environment, since the algorithm includes users' tagging behavior through feedback from other people as an important factor in calculating trust value, rather than simply relying on the user contributed tags.

To further justify the usage of trust modeling in the automatic landmark tagging system, we measure the accuracy of propagation and the percentage of the number of propagated tags versus the threshold set for the user trust modeling. The results are shown in Fig. 11 for the socially-driven approach by Ivanov *et al.* The optimal accuracy using object duplicate detection for geotag propagation is 71%. 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. 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 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

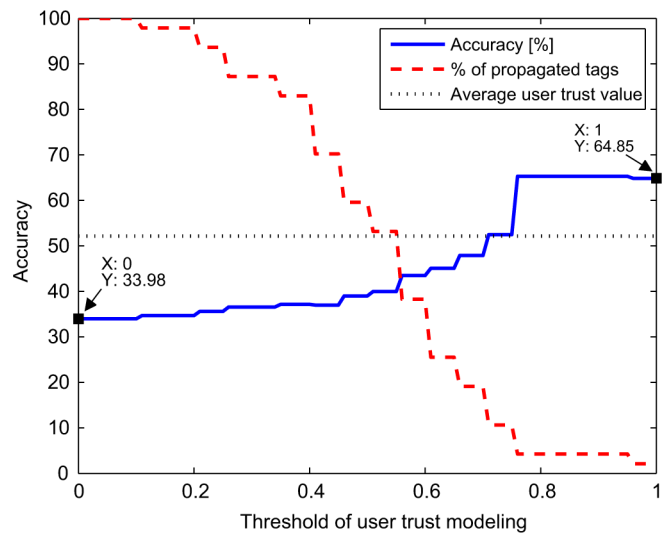


Fig. 11. Recognition rate of the geotag propagation system and the percentage of the propagated tags versus the threshold  $T$  for the user trust modeling in the socially-driven approach by Ivanov [4].

user tags are propagated. In this case the accuracy of the system is 34%. This additionally confirms that the accuracy of propagation can be significantly improved by including trust modeling in the automatic landmark tagging system.

## VII. CONCLUSION

In this paper, we have presented different techniques for user trust modeling that are suitable for geotagging and can be used in geotag propagation systems. The problem of having trustworthy geotags of the content is important in social networks, because of their increasing popularity as means of 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.

The particular focus of this paper is on the 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 these correspondences and a trust value of the model derived for each user, only reliable geotags are propagated, which leads to a decrease of tagging efforts. We have analyzed the performance of the tag propagation alone which leads to a promising average accuracy of 71% over all the landmarks. We have also shown that the performance varies considerably among different landmark types depending on their visual characteristics. We have analyzed the influence of wrongly annotated tags, which causes even more wrongly propagated tags in the database. Several trust models were evaluated and compared. The results show that by propagating tags based on the trust modeling relying on users' tagging behavior, the larger number of tags (more than twice) can be propagated with the same accuracy compared to using other trust models

that simply rely on the user contributed tags or if using no trust modeling at all.

Since this type of comparative study is a pioneering work, a future study may consider a more careful selection of participants, for example, equal distribution of participated users in terms of group sizes and background. In this paper, we compared trust modelling for automatic tagging considering closed set problem, as we could precisely measure number of tags in the system. However, we expect that the open set problem would also work fairly good, granted that we have a “good” thresholding method for object duplicate detection step. Most of the current techniques for noise and spam reduction focus only on textual tag processing and user profile analysis, while visual features of multimedia content can also provide useful information about the relevance of the content and content-tag relationship. In the future, a promising research direction would be to combine multimedia content analysis with conventional tag processing and user profile analysis.

## REFERENCES

- [1] C. Taylor, The Most Important Facebook Number: 140.3 Billion, Oct. 2012 [Online]. Available: <http://mashable.com/2012/10/05/the-most-important-facebook-number-140-billion>
- [2] L. S. Kennedy, S.-F. Chang, and I. V. Kozintsev, “To search or to label?: Predicting the performance of search-based automatic image classifiers,” in *Proc. ACM MIR*, Oct. 2006, pp. 249–258.
- [3] The Impact of the Internet and Social Media on Travel, Funsherpa.com, Jul. 2012 [Online]. Available: <http://blog.funsherpa.com/2012/07/travel-infographic>
- [4] I. Ivanov, P. Vajda, J.-S. Lee, L. Goldmann, and T. Ebrahimi, “Geotag propagation in social networks based on user trust model,” *Multimedia Tools and Applicat.*, vol. 56, no. 1, pp. 155–177, Jan. 2012.
- [5] G. Koutrika, F. A. Effendi, Z. Gyöngyi, and P. H. Garcia-Molina, “Combating spam in tagging systems: An evaluation,” *ACM Trans. Web*, vol. 2, no. 4, pp. 22:1–22:34, Oct. 2008.
- [6] K. Liu, B. Fang, and Y. Zhang, “Detecting tag spam in social tagging systems with collaborative knowledge,” in *Proc. IEEE FSKD*, Aug. 2009, pp. 427–431.
- [7] Z. Xu, Y. Fu, J. Mao, and D. Su, “Towards the semantic web: Collaborative tag suggestions,” in *Proc. ACM WWW*, May 2006, pp. 1–8.
- [8] R. Krestel and L. Chen, “Using co-occurrence of tags and resources to identify spammers,” in *Proc. ECML PKDD*, Sep. 2008, pp. 38–46.
- [9] J. Luo, D. Joshi, J. Yu, and A. Gallagher, “Geotagging in multimedia and computer vision—A survey,” *Multimedia Tools and Applicat.*, vol. 51, no. 1, pp. 187–211, 2011.
- [10] J. Hays and A. A. Efros, “IM2GPS: Estimating geographic information from a single image,” in *Proc. IEEE CVPR*, 2008, pp. 1–8.
- [11] L. S. Kennedy and M. Naaman, “Generating diverse and representative image search results for landmarks,” in *Proc. WWW*, 2008, pp. 297–306.
- [12] Y. T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T. Chua, and H. Neven, “Tour the world: Building a web-scale landmark recognition engine,” in *Proc. IEEE CVPR*, 2009, pp. 1085–1092.
- [13] T. Quack, B. Leibe, and L. V. Gool, “World-scale mining of objects and events from community photo collections,” in *Proc. IEEE CIVR*, 2008, pp. 47–56.
- [14] S. Gammeter, L. Bossard, T. Quack, and L. van Gool, “I know what you did last summer: Object level auto-annotation of holiday snaps,” in *Proc. ICCV*, Oct. 2009, pp. 614–621.
- [15] Technical Standardization Committee on AV & IT Storage Systems and Equipment, Exchangeable Image File Format for Digital Still Cameras: Exif Version 2.2, Electronics and Information Technology Industries Association, Japan, Tech. Rep. JEITA CP-3451, Apr. 2002.

- [16] IPTC Standards Committee, IPTC photo metadata standard, IPTC Core 1.1 and IPTC Extension 1.1, International Press Telecommunications Council, Jul. 2010.
- [17] L. Hollenstein and R. Purves, “Exploring place through user-generated content: Using flickr to describe city cores,” *J. Spatial Inf. Sci.*, vol. 1, no. 1, pp. 1–29, 2010.
- [18] C. Marlow, M. Naaman, D. Boyd, and M. Davis, “Ht06, tagging paper, taxonomy, flickr, academic article, to read,” in *Proc. ACM HT*, Aug. 2006, pp. 31–40.
- [19] I. Ivanov, P. Vajda, J. S. Lee, and T. Ebrahimi, “In tags we trust: Trust modeling in social tagging of multimedia content,” *IEEE Signal Proc. Mag.*, vol. 29, no. 2, pp. 98–107, Mar. 2012.
- [20] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *J. ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999.
- [21] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, “Combating web spam with TrustRank,” in *Proc. VLDB*, Aug. 2004, pp. 576–587.
- [22] A. Budanitsky and G. Hirst, “Evaluating wordnet-based measures of lexical semantic relatedness,” *Comput. Linguist.*, vol. 32, no. 1, pp. 13–47, Mar. 2006.
- [23] R. L. Cilibrasi and P. M. B. Vitanyi, “The google similarity distance,” *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 370–383, Mar. 2007.
- [24] P. Vajda, I. Ivanov, L. Goldmann, J.-S. Lee, and T. Ebrahimi, “Robust duplicate detection of 2D and 3D objects,” *Int. J. Multimedia Data Eng. Manage.*, vol. 1, no. 3, pp. 19–40, Jul.–Sep. 2010.
- [25] K. Mikolajczyk and C. Schmid, “An affine invariant interest point detector,” in *Proc. ECCV*, 2002, pp. 128–142.
- [26] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Computer Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [27] D. Nister and H. Stewenius, “Robust scalable recognition with a vocabulary tree,” in *Proc. IEEE CVPR*, 2006, pp. 2161–2168.
- [28] D. H. Ballard, “Generalizing the Hough transform to detect arbitrary shapes,” *Pattern Recogn.*, vol. 13, no. 2, pp. 111–122, 1981.
- [29] P. Vajda, L. Goldmann, and T. Ebrahimi, “Analysis of the limits of graph-based object duplicate detection,” in *Proc. Symp. Multimedia*, 2009, pp. 600–605.



**Ivan Ivanov** is a research assistant and Ph.D. student in the Multimedia Signal Processing Group, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He received the Dipl. Ing. (M.Sc.) degree in electrical engineering from the University of Belgrade, Serbia, in 2006. In 2006 and 2007, he worked as a hardware design engineer for Texas Instruments, France, where he participated in the development of low-power very large-scale integration multimedia applications for portable devices. He also worked as a radio access network conceptual planning expert in Vip mobile, Serbia, focusing on the implementation of second- and third-generation radio access technologies. His current research interests include social media processing and analysis, visual search, and computer vision.



**Peter Vajda** is a visiting assistant professor in the Department of Electrical Engineering at Stanford University. He received the doctorate degree from EPFL in Switzerland. He was involved in several European projects such as Visnet II Network of Excellence (Networked Audiovisual Media Technologies), K-Space (The Knowledge Space of Technology to Bridge the Semantic Gap), and PetaMedia (Peer-to-peer Tagged Media). Within these European projects, he was working in the area of object detection for video surveillance applications, image segmentation evaluation, and metadata propagation based on object duplicate detection. His current research interests include large-scale image retrieval system for mobile visual search applications and personalized real-time video news.



**Pavel Korshunov** received the Dipl. Sci. (M.Sc.) degree in mathematics and computer science from Saint-Petersburg State University, Russia, in 2002. In 2010, he received the Ph.D. degree in computer science from National University of Singapore under the supervision of Wei Tsang Ooi. In 2007, he worked as a Ph.D. intern at FX Palo Alto Laboratory (FXPAL) under the supervision of Larry Row. He joined Touradj Ebrahimi's Multimedia Signal Processing Group at EPFL as postdoctoral researcher in 2011. His research interests include computer vision and video analysis, video streaming, video and image quality assessment, video quality for video analysis, crowdsourcing, video and image encoding, high dynamic range imaging, and privacy issues in video surveillance systems.



**Touradj Ebrahimi** is currently a professor at EPFL heading its Multimedia Signal Processing Group. He 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, seven ISO certificates for key contributions to MPEG-4, JPEG 2000, JPSearch, and JPEG XR standards, and the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS Best Paper Award. He is the head of the Swiss delegation to MPEG, JPEG, and SC29 and chair of the Advisory Group on Management in SC29. His research interests include still, moving, and three-dimensional image processing and coding, visual information security, new media, and human-computer interaction.