GeoTime-Based Tag Ranking Model for Automatic Image Annotation

Mădălina Mitran, Guillaume Cabanac, Mohand Boughanem

GeoTime-Based Tag Ranking Model for Automatic Image Annotation

Mădălina Mitran
IRIT UMR 5505 CNRS, France
University of Toulouse
Madalina.Mitran@irit.fr

Guillaume Cabanac
IRIT UMR 5505 CNRS, France
University of Toulouse
Guillaume.Cabanac@irit.fr

Mohand Boughanem
IRIT UMR 5505 CNRS, France
University of Toulouse
Mohand.Boughanem@irit.fr

ABSTRACT

In this paper, we propose a novel GeoTime-based tag ranking model to address the automatic image annotation problem. Our aim is to assign relevant descriptors to a query image using textual, spatial, and temporal clues from nearby images. The assumption behind our model is that tags associated with images that are closer in time and space with a query image are more likely to be relevant to it.

Given a query image we retrieve the images (available in community image databases, such as `flickr.com` and `panoramio.com`) located in its close geographical area using its GPS coordinates (i.e., latitude and longitude). Once these images retrieved, we take advantage of their metadata (e.g., users' social contributions and information stored in the EXIF descriptor) in order to suggest relevant tags. While most of state-of-the-art approaches used visual and textual factors to suggest and rank tags, our model uses also temporal and spatial proximity factors. To capture these proximity factors we exploited similarity methods and kernel functions. Finally, the top-ranked tags are used to annotate a query image.

We conducted a series of experiments on a dataset consisting of over 201,000 Flickr images from the Paris geographic area. The experimental results showed that our GeoTime-based tag ranking model yields significant improvement over two state-of-the-art baselines.

Keywords

Image Annotations, Social Multimedia, Tag Ranking Model, Spatial and Temporal Proximities, Flickr, Kernel Functions

1. INTRODUCTION

The automatic image annotation task faces the problem of finding terms (tags) that best describe a given query image provided by a user as input to the annotation system. The evolution of new media (e.g., image sharing platforms) and the development of digital technologies (e.g., digital cameras, smartphones) yields an exponential increase of the number of images available on the web. This evolution makes the image annotation problem increasingly challenging.

In our view, automatic image annotation faces two main challenges. The first challenge is to find the terms that best describe a query image (i.e., its candidate tag set). Previous work used content-based techniques represented by visual features [14], text-based approaches represented by the text surrounding the image in the document from where it appeared [4], and more recently, community-based approaches represented by the metadata available in image sharing platforms (e.g., title of an image, comments, tags, and so on) [13, 2]. To make use of the most recent approaches, in our work we employed tag metadata that are generally assigned by users who uploaded images in online repositories, such as `flickr.com` and `panoramio.com`.

The second challenge is to rank the tags recommended for a query image. Therefore, the problem that arises is to identify the factors and the methods to combine them in order to obtain the best scores for the meaningful tags. To tackle this problem, some authors used supervised and unsupervised ranking methods to combine tag frequency, spatial proximity, image content similarity, and users activities factors [13]. Other authors used linear [13] and statistical models [10] to combine high level (e.g., user tags) and low level information (e.g., global color, edge features, SIFT, and SURF operators). Our work furthers prior works by introducing the temporal factor represented by the time when the images were taken. To the best of our knowledge this factor was not used so far as a measure to rank tags. Besides the temporal factor, the tag ranking model that we propose takes into account textual and spatial factors too. Thus, another aspect of our work consists in the methods employed to rank tags according to each of these three factors.

In this paper, we address these two challenges and measure how much improvement is yielded using the proposed GeoTime-based tag ranking model in the automatic image annotation task. We evaluate our model on a dataset consisting of over 201,000 images of the Paris geographic area obtained using the Flickr API. Moreover, we compare its performance with two state-of-the-art baselines from [13, 10]. The results obtained show significant improvements of our model over the two baselines and its effectiveness for the use of the temporal and spatial proximity factors in the automatic image annotation task. The proposed model is thus able to assign relevant descriptors for a given query image.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes our GeoTime-based tag rank-

ing model. Section 4 details our experimental setup, including the test collection, the evaluation metrics, the two baselines, and the model parameter selection. The results are discussed in section 5. Finally, section 6 concludes the paper and outlines further work.

2. RELATED WORK

In this paper we address the automatic image annotation problem. In the following, we will review image annotation techniques aiming to describe a query image with meaningful tags.

Sevil et al. [11] proposed an automatic image tag expansion approach, using visual and textual factors from other related images. The initial tags added by the users are used to retrieve related images together with their tags. Furthermore, tags associated with the retrieved images are weighted according to the visual similarity between the retrieved images and the query image. Finally, the tags with the highest weights are presented to the users. On the other hand, Sigurbjörnsson and van Zwol [12] present a tag recommendation system using the collective knowledge from the image sharing platform `flickr.com`. The authors used tag co-occurrence statistics in order to recommend annotations for partially tagged images.

Other authors used the title of a query image together with its visual content to retrieve related images together with their tags [2]. Furthermore, tags associated with the retrieved images are weighted according to their popularity and finally added to the recommended tag list. Other works [16] proposed an automatic approach that exploits the semantic correlation between image content and tags using Kernel Canonical Correlation Analysis. Eventually, they used this correlation and the input-independent tag popularity to recommend tags.

Our work differs from the previous studies in two aspects. First, instead of using visual and textual factors to retrieve images for a given query image, we propose to use a recent technique based on location information. In this way, we ensure the scalability of the model, and also we avoid any irrelevant initial user tags that can lead to poor results. Second, assuming that tags associated with images that are closer in time and space with a query image have a higher probability to be relevant, we propose to capture temporal and spatial proximities between a query image and a retrieved image in order to weight these tags.

More similar to our work are those of Silva and Martins [13], Moxley et al. [7], and Sergieh et al. [10]. They used the geographical coordinates of a query image to retrieve related images together with their metadata and several features in order to assign scores for tags. Silva and Martins [13] used a set of estimators based on: tag frequency, spatial proximity of the image, image content similarity, and the number of different users employing the tag. Furthermore, they proposed to combine the multiple estimators through supervised learning to rank methods and unsupervised rank aggregation methods. In [7] tags are ranked according to their local frequency in comparison to their global frequency. Sergieh et al. [10] proposed a probabilistic model that combines two kinds of information: high level information represented by user tags and low level information represented by the visual similarity between the query image and the set of similar images.

The review of related work shows that a particular factor, namely the temporal factor was overlooked. None of these works considered it in the tag ranking process, which can lead to poor results when we have to suggest and rank tags for event query images.

Finally, while these approaches presented only statistical methods and probabilistic models to combine different factors, we propose a GeoTime-based tag ranking model similar to the positional language models for information retrieval [6, 3]. The difference consists in the type of data for which they are used, in information retrieval for documents and, in our case, for images. The model proposed in this paper exploits spatial, temporal, and textual factors to design similarity methods and kernel functions in order to capture the temporal and spatial proximities between a query image and a nearby image.

3. TAG RANKING APPROACH

In this section, we first present an overview of our proposed model and then we detail it.

3.1 Model Overview

As shown in Figure 1, our approach consists in two steps. First, given the latitude and the longitude of a query image, we retrieve several other images together with their user social contributions (i.e., tags). These images available in online community databases are retrieved within a fixed radius, r, of the location of the query image. Once these images found, we exploit their tags to create a candidate tag set for the given query image. Second, we exploit spatial, temporal, and textual factors to assign scores for each tag from the candidate tag set. Finally, the tags with the highest scores are used to annotate a query image. In the following, we introduce our GeoTime-based tag ranking model consisting of three functions: `textual-based function`, `spatial-based function`, and `temporal-based function`.

3.2 GeoTime-Based Tag Ranking Model

In order to present our model more formally, let us introduce the following notations: i_q is the query image, i_r is a retrieved image, I is the collection of all images from our dataset, and I_R (produced as described in section 3.1) is the set of images retrieved for each i_q ($I_R \subseteq I$). Each retrieved image, $i_r \in I_R$, is represented by the tuple ($annot_{i_r}$, $spatial_{i_r}$, $temporal_{i_r}$), where:

- $annot_{i_r}$ is the tag set associated with the i_r image ($annot_{i_r} = \{tg_1, tg_2, \ldots, tg_n\}$). $annot_{i_r} \subseteq Annot_{I_R}$, where $Annot_{I_R}$ represents the candidate set of unique tags obtained from the images of I_R,
- $spatial_{i_r}$ is the location of the retrieved image defined by the pair ($longitude_{i_r}, latitude_{i_r}$),
- and $temporal_{i_r}$ is the image capture timestamp defined by the temporal information.

Spatial and temporal metadata are available in the EXIF format [1] as part of image files (when taken with devices embedded with GPS sensors). Otherwise, these metadata are given by the users.

To assign scores to each tag tg_n from the query image candidate tag set $Annot_{I_R}$, we estimate the conditional probability $P(tg_n|i_q)$, i.e., the probability of tg_n given the query image i_q. Assuming $P(i_q)$ to be a uniform prior probability

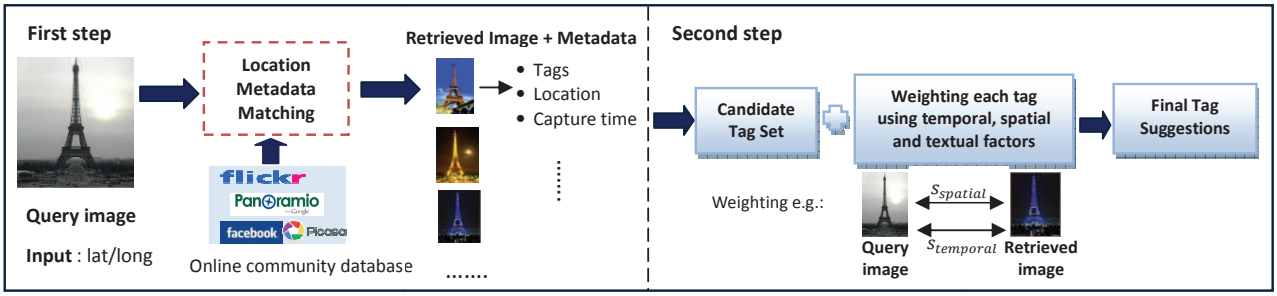

Figure 1: Model overview. The location information is first used to retrieve images for a query image. Then spatial, temporal, and textual factors are used to weight tags.

and the conditional independence between tg_n and i_q, we obtain the following equation:

$$P(tg_n|i_q) \propto \sum_{i_r \in I} P(i_r)P(tg_n|i_r)P(i_q|i_r) \quad (1)$$

where $P(i_r)$ is the probability of i_r to occur in the close area of the query image: $d(i_q, i_r) < r$, where $d(i_q, i_r)$ is the spatial distance between two images and r represents a fixed radius (in meters). Assuming $P(i_r)$ to be a uniform probability, represented by a constant value, the equation 1 can be simplified as:

$$P(tg_n|i_q) \propto \sum_{i_r \in I_R} P(tg_n|i_r)P(i_q|i_r) \quad (2)$$

where the conditional probabilities $P(tg_n|i_r)$ and $P(i_q|i_r)$ are the weight of the candidate tag tg_n given by the retrieved image i_r, and the proximity between i_r and i_q, respectively. In this paper, we estimate the $P(i_q|i_r)$ probability by considering temporal and spatial proximities. Therefore, the proposed tag ranking model predicts tag scores according to three functions: textual, spatial, and temporal. Furthermore, we rewrite the equation 2 as:

$$P(tg_n|i_q) \propto score(tg_n, i_q) = \sum_{i_r \in I_R} s_{textual}(tg_n, i_r)s_{spatial}(i_q, i_r)s_{temporal}(i_q, i_r) \quad (3)$$

The following sections present each of the three functions in more details.

3.2.1 Textual-based function

Given the candidate tag set $Annot_{I_R}$ associated with a i_q, one way to estimate the textual score of each tag $tg_n \in Annot_{I_R}$ is to use a frequency-based approach (i.e., similar with the tf measure from Information Retrieval) as follows:

$$s_{textual}(tg_n, i_r) = P(tg_n|i_r) = \frac{tf(tg_n, i_r)}{|i_r|} \quad (4)$$

where $tf(tg_n, i_r)$ measures the occurrence of the tag tg_n in the annotations of an image i_r, and $|i_r|$ represents the total number of tags that annotate i_r.

$$tf(tg_n, i_r) = \begin{cases} 1 & \text{if } tg_n \in annot_{i_r}, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

A similar method is employed by Sergieh et al. [10] where the value $\sum_{i_r \in I_R} s_{textual}(tg_n, i_r)$ represents the weighted vote of the tag tg_n. In our work we used this measure in two ways. First, as a textual-based function in our GeoTime-based tag ranking model, and second as a baseline for our experiments presented in section 4.

3.2.2 Spatial-based function

We assume that a tag is more important for a query image if it occurs in images located close to the location of the query image (i.e., the spatial distance between a query image and a retrieved image is small). Thus, the spatial score of each tag $tg_n \in Annot_{I_R}$ is computed as:

$$s_{spatial}(i_q, i_r) = P_s(i_q|i_r) \quad (6)$$

where $P_s(i_q|i_r)$ is the probability that captures the spatial proximity between two images. To compute this probability we represented the spatial distance between two geographical points in terms of latitude and longitude, $d(i_q, i_r)$, by the great circle method[1], also used in [13]. In the following, we present two methods to compute the $P_s(i_q|i_r)$ probability:

a) The first one, similar to the one presented in [13], transforms the spatial distance, $d(i_r, i_q)$, into a spatial similarity according to the following equation:

$$P_s(i_q|i_r) = \frac{1}{1 + d(i_q, i_r)} \quad (7)$$

To avoid the division by zero when the query image and the retrieved image are taken in the same location we add 1 to the spatial distance.

b) For the second method we propose to use a proximity-based density function. By doing so, the $P_s(i_q|i_r)$ probability is calculated using the kernel functions $K_\sigma(i_q, i_r)$ that decrease as the spatial distance between i_q and i_r increases. We describe later the kernel functions that we used in this paper, as they also appear in the next section.

3.2.3 Temporal-based function

We assume that the description of a query image is time-dependent. For example, at Wembley stadium there are concerts and also football games at different time points. Therefore, images taken from the same location at different time points are related to different events. So, we assume that a tag is more important for a query image if it occurs

[1]http://en.wikipedia.org/wiki/Great-circle_distance

in images captured close in time to it (i.e., the temporal distance between a query image and a retrieved image is small). Thus, the temporal score of each tag $tg_n \in Annot_{I_R}$ is computed as:

$$s_{temporal}(i_q, i_r) = P_t(i_q|i_r) \quad (8)$$

where $P_t(i_q|i_r)$ is the probability that captures the temporal proximity between two images. To compute $P_t(i_q|i_r)$ we represented the temporal information of images (i.e., date/time) by the Unix Time (UxT) format as the number of seconds elapsed since `January 1, 1970`[2]. Thus, this probability can be expressed by the temporal distance between the temporal representation of the query image (UxT_{i_q}) and the temporal representation of the retrieved image (UxT_{i_r}). In the following, we present two methods to compute this probability:

a) The first one converts the temporal distance into temporal similarity according to the following equation:

$$P_t(i_q|i_r) = \frac{1}{1 + |UxT_{i_q} - UxT_{i_r}|} \quad (9)$$

We added 1 to the temporal distance to avoid the division by zero when the query image and the retrieved image are captured in the same moment of time.

b) The second method is similar to the second method presented for the spatial-based function. The main difference is that the kernel functions $K_\sigma(i_q, i_r)$ decrease as the temporal distance between i_q and i_r increases.

Previous works used kernel functions to capture the proximity of the words in a document [6, 3]. However, here we aim to capture the temporal and spatial proximity of tags that are associated with retrieved images. These proximities have not been considered in the automatic image annotation process to the best of our knowledge. In this paper, we investigate the Gaussian and Laplace kernel functions (equations 10 and 11) that proved to perform well in the literature [6, 3].

- **Gaussian Kernel**

$$K_\sigma(i_q, i_r) = exp\left[\frac{-(i_q - i_r)^2}{2\sigma^2}\right] \quad (10)$$

- **Laplace Kernel**

$$K_\sigma(i_q, i_r) = exp\left[\frac{-|i_q - i_r|}{\sigma}\right] \quad (11)$$

In order to obtain the final score for each candidate tag tg_n we normalized each score x (i.e., $score(tg_n, i_q)$) using the distribution of scores X in $Annot_{I_R}$.

$$Normalized(x, X) = \frac{x - min(X)}{max(X) - min(X)} \quad (12)$$

Finally, the candidate tags with the highest scores will be used in the automatic image annotation process.

4. EXPERIMENTAL SETUP

In this section we present our experimental setup: the dataset, the evaluation metrics, the two baselines, and the model parameter selection that we conducted to evaluate the effectiveness of our tag ranking model.

[2]http://en.wikipedia.org/wiki/Unix_time

4.1 Dataset

The dataset used in this research consisted in over 201,000 Flickr geo-tagged images from the Paris geographic area (available under request). These images were crawled through the Flickr API[3], with the mention that we crawled only the images for which the time, location, and tag information were available. The number of unique tags for this dataset reaches 81,000 and the average number of tags per image is 9.18. These images come with various metadata, such as: image id, image title, image tags, latitude, longitude, date/time, and the image URL. As a test set, we randomly selected 200 images from the dataset and as gold standard we considered the original tags of each image.

4.2 Evaluation Metrics

In order to calculate the performance of our model we compared the original tag list of a query image with the suggested annotations. Moreover, we used well-known Information Retrieval metrics such as: Precision of the 5 recommended tags (P@5), R-precision (Rprec), and Mean Average Precision (MAP), used also to evaluate state-of-the-art approaches [13, 12]. Considering that a system could recommend relevant tags that are not among the original query image tags, these metrics should be used only as relative performance indicators, as mentioned in the work of [8].

4.3 Baselines

We compared the effectiveness of our GeoTime-based tag ranking model to two baselines.

The first baseline, B_1, (similar to the tf measure from Information Retrieval) is represented by the number of nearby retrieved images of a query image that contain the tag. This is one of the most common state-of-the-art approaches for the automatic image annotation task. The equation employed for this baseline is presented in section 3.2.1 and is similar to the ones used in the work of Sergieh et al. [10] and Hsiao et al. [5].

As a second baseline, B_2, we considered the best performing aggregation method presented in the work of Silva and Martins [13]. This method is based on the CombMNZ score of the following four factors: number of nearby images using the tag, number of different `flickr.com` users employing the tag, number of web visits made for the nearby images using the tag, and the minimum distance between the GPS coordinates of the nearby images that use the tag and the GPS coordinates of the query image. This baseline is given by the equation 13.

$$B_2 = CombMNZ(tg_n, i_q) = p * \sum_{j=1}^{k} score_j(tg_n, i_q) \quad (13)$$

where $score_j(tg_n, i_q)$ is the score received by the tag for each of the individual factor and p represents the number of non-zero similarities.

4.4 Model Parameter Selection

Our model has two parameters: σ which is the interpolation coefficient used by the kernel functions, and r which is the geographic radius used to retrieve images and implicitly to form the candidate tag set for a query image. In order to find the best parameters, we tested different values

[3]http://www.flickr.com/services/api/

Table 1: **Empirical best values for σ coefficient for both kernel functions (Gaussian and Laplace) and both temporal (t) and spatial (s) factors.**

	kernel	σ	P@5	P@10	MAP	R-prec
s	Gaussian	4	0.4237	0.2492	0.4559	0.4434
	Laplace	2	**0.4305**	**0.2497**	**0.4684**	**0.4624**
t	Gaussian	2	0.4667	0.2898	0.5281	0.5195
	Laplace	2	**0.4767**	**0.2927**	**0.5343**	**0.5225**

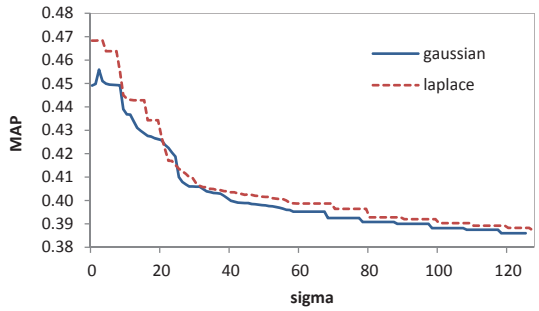

Figure 2: Sensitivity to the σ parameter of kernel functions for the spatial factor.

for σ in the range of [2, 128] and for r we tried the following values: 10m, 100m, 300m, 500m, and 1km. Values of r greater than 1km do not conduct to better results as demonstrated also in [7]. A large radius provides a greater number of images making the computational time cost to increase. On the other hand, a small radius cannot ensure the necessary number of images in order to form the candidate tag set for a query image. Therefore, taking into account these two observations we decided to use the value of 300m in our experiments for the spatial radius.

The best performance for the interpolation coefficient σ for both kernel functions and for both temporal and spatial factors are reported in Table 1. As we can notice in the table the Laplace kernel outperforms the Gaussian kernel in all the measures that we considered. Thus, we use it to rank tags for both temporal and spatial factors in our model.

Figures 2 and 3 relate the sensitivity in terms of MAP to the σ parameter in the range of [2, 128] for both kernel functions. We notice once again that the Laplace kernel provides better results than the Gaussian kernel over different parameter settings.

5. RESULTS AND DISCUSSION

In this section, we present the evaluation results of our GeoTime-based tag ranking model and its effectiveness compared to the two baselines described in section 4.

Since our tag ranking model can be configured according to two spatial-based functions (i.e., spatial similarity and kernel spatial functions) and two temporal-based functions (i.e., temporal similarity and kernel temporal functions) we first investigated the impact of each individual function on the performance. The results are presented in Table 2 where the top section shows the results for the spatial-based functions and the bottom section the results for the temporal-based functions. Table 2 shows that the Laplace kernel function yields a significant improvement for both spatial and temporal functions for all metrics. We can thus argue that the Laplace kernel function performs well and would be useful to use it in the GeoTime-based tag ranking model.

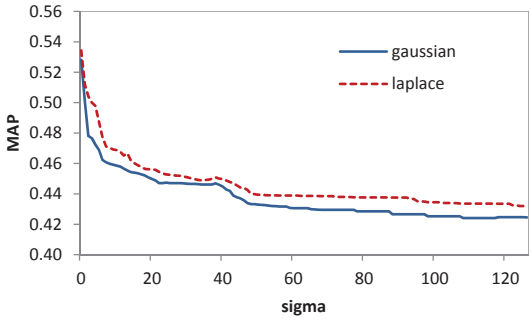

Figure 3: Sensitivity to the σ parameter of kernel functions for the temporal factor.

If we compare the performance between the Laplace temporal and Laplace spatial functions, we observe that for all metrics the Laplace temporal function outperforms the Laplace spatial function. The intuition behind this behavior is that events may occur in the location where the query image was taken, making the event-specific tags have a good ranking in the candidate tag set of a query image. For example, for a Eiffel Tower query image taken on the Bastille Day, the images retrieved in its nearby geographical area (not necessarily taken on the Bastille Day) are mostly annotated with tags like, `Eiffel Tower`, `Paris`, `France`, and so on, and only a few are annotated with tags like, `14 July`, `national holiday`, and `Bastille Day` (especially those taken on the Bastille Day). Therefore, considering only textual and spatial factors these tags would not get high scores. On the contrary, the temporal factor can identify these tags and boost their scores in the candidate tag set. The significant improvement obtained by using the overlooked temporal factor across the spatial one, 14% in terms of MAP and 8% in terms of Precision at rank 5, show us that it plays an important role in the GeoTime-based tag ranking model.

Finally we compared our model with the two baselines presented in section 4 and report the results in Table 3. The top section of Table 3 shows the evaluation results in terms of Precision at rank 5 (P@5), MAP, and Rprec. The best values in each column are in bold face. As can be seen from the table, none of the baselines improved the GeoTime-based tag ranking model for all metrics.

As the precision at rank 5 of our model goes up to 49%, it means that on average 2.45 of the top 5 tags are good descriptors for a query image. In addition, our model has the highest MAP, which suggests its effectiveness and stability across the two baselines. We also notice (Table 3) that the baseline obtained from the work of Silva and Martins [13] outperforms the classical tf measure from Information Retrieval used in several automatic image annotation works.

The bottom section of Table 3 shows the improvement that our tag ranking model yields over the first and second baselines. The improvements achieved are statistically significant tested using the Student's bilateral and paired t-test [15] ($p < 0.01$), found reliable for this type of experiments [9]. As can be noticed from the table, the improvements (for all metrics) between the first baseline and the

Table 2: Results for the two spatial and for the two temporal functions using our Flickr collection.

	P@5	MAP	Rprec
Spatial functions			
spatial similarity	0.416	0.448	0.441
Laplace kernel	**0.431**	**0.468**	**0.462**
Temporal functions			
temporal similarity	0.463	0.515	0.500
Laplace kernel	**0.465**	**0.534**	**0.522**

Table 3: Results for our GeoTime-based tag ranking model (GT-TRM) over the two baselines (bline) presented in section 4. A dag (†) and a star (*) indicate statistically significant improvement over the first and second baseline respectively.

	P@5	MAP	Rprec
bline 1	0.261	0.345	0.301
bline 2	0.361	0.414	0.391
GT-TRM	**0.493 †***	**0.566 †***	**0.543 †***
Improvement			
bline 1 vs bline 2	38.5%	19.8%	30.1%
bline 1 vs GT-TRM	89.2%	63.8%	80.4%
bline 2 vs GT-TRM	36.5%	36.6%	38.6%

proposed model are greater than the improvements between the second baseline and the proposed model. This was expected because of the noticeable performance of the second baseline across the first one. With the improvements that we obtained we can conclude that our GeoTime-based tag ranking model improves the automatic image annotation effectiveness compared to state-of-the-art research.

6. CONCLUSION

In this paper we addressed the automatic image annotation problem to find and rank tags according their relevance to a query image. In our work we used data from image sharing platforms and we focused on textual information provided by these ones and on temporal and spatial information stored in the image EXIF descriptor. We integrated this information into a tag ranking model following the intuition that textual, spatial, and temporal factors together plays an important role in the tag ranking process. Furthermore, we proposed to use kernel functions to compute the spatial and temporal proximity factors and we showed that we obtain better results when employing them.

We have evaluated the proposed GeoTime-based tag ranking model on a dataset obtained using the Flickr API. We showed that the overlooked temporal factor, not used so far in the tag ranking process, plays an important role in this process. Moreover, we concluded that the model showed to be effective across two state-of-the-art baselines which proves an improvement in the automatic image annotation process.

Further work will include experiments on other datasets using larger test and training sets. Another direction of research will be to investigate the impact of other kernel functions (e.g. triangular kernel, cosine kernel, rectangular kernel) and also to look forward for other available resources that could be used to improve the effectiveness of our ranking model, such as a synonym database in order to expand the tags for a query image.

7. REFERENCES

[1] Exchangeable image file format for digital still cameras: Exif version 2.2. Technical report, Japan Electronics and Information Technology Industries Association, 2002.

[2] S. Barai and A. F. Cardenas. Image annotation system using visual and textual features. In *DMS '10*, pages 289–296. Knowledge Systems Institute, 2010.

[3] S. Gerani, M. J. Carman, and F. Crestani. Proximity-based opinion retrieval. In *SIGIR '10*, pages 403–410. ACM, 2010.

[4] Z. Gong, L. H. U., and C. W. Cheang. Web image indexing by using associated texts. *Knowledge and Information Systems*, 10(2):243–264, 2006.

[5] J.-H. Hsiao, C.-S. Chen, and M.-S. Chen. A novel language-model-based approach for image object mining and re-ranking. In *ICDM '08*, pages 243–252. IEEE Computer Society, 2008.

[6] Y. Lv and C. Zhai. Positional language models for information retrieval. In *SIGIR '09*, pages 299–306. ACM, 2009.

[7] E. Moxley, J. Kleban, and B. S. Manjunath. Spirittagger: a geo-aware tag suggestion tool mined from flickr. In *MIR '08*, pages 24–30. ACM, 2008.

[8] A. Rae, B. Sigurbjörnsson, and R. van Zwol. Improving tag recommendation using social networks. In *RIAO '10*, pages 92–99. CID - Le Centre de Hautes Etudes Internationales D'Informatique Documentaire, 2010.

[9] M. Sanderson and J. Zobel. Information retrieval system evaluation: effort, sensitivity, and reliability. In *SIGIR '05*, pages 162–169. ACM, 2005.

[10] H. M. Sergieh, G. Gianini, M. Döller, H. Kosch, E. Egyed-Zsigmond, and J.-M. Pinon. Geo-based automatic image annotation. In *ICMR '12*, pages 46:1–46:8. ACM, 2012.

[11] S. G. Sevil, O. Kucuktunc, P. Duygulu, and F. Can. Automatic tag expansion using visual similarity for photo sharing websites. *Multimedia Tools Appl.*, 49(1):81–99, 2010.

[12] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW '08*, pages 327–336. ACM, 2008.

[13] A. Silva and B. Martins. Tag recommendation for georeferenced photos. In *GIS-LBSN '11*, pages 57–64. ACM, 2011.

[14] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1349–1380, 2000.

[15] Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

[16] Z. Wang and B. Li. Learning to recommend tags for on-line photos. In *2nd International Workshop on Social Computing, Behavior Modeling, and Prediction*, pages 1–9. Springer, 2009.