Exploring Temporal Proximity and Spatial Distribution of Terms in Web-based Search of Event-Related Images

Massimiliano Ruocco
Dept. of Computer and Information Science
Norwegian University of Science and Technology
NO-7491 Norway
ruocco@idi.ntnu.no

Heri Ramampiaro
Dept. of Computer and Information Science
Norwegian University of Science and Technology
NO-7491 Norway
heri@idi.ntnu.no

ABSTRACT

Pictures in media sharing applications are increasingly accompanied with geotags. For this reason, we stress the importance of exploring the possibility of applying spatial, as well as the temporal dimensions in searching event-related pictures. Specifically, we propose extended query expansion models that exploit the information about the temporal neighbourhoods among pictures in a collection and leverage on the spatio-temporal distribution of the candidate expansion terms to re-weight and expand the initial query. To evaluate our approach, we conduct extensive experiments on a large dataset consisting of 88 million pictures from Flickr. The results from these experiments demonstrate the viability and effectiveness of our method with respect to retrieval performance, considering both a large dataset and query pictures with restricted size of terms.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and Retrieval—*Information Search and Retrieval*

Keywords

Event Retrieval, Image Clustering, Relevance Feedback

1. INTRODUCTION

The explosion of photos shared on the web has not only opened many possibilities but also resulted in new needs, and hence new challenges. Although recent developments and technological advances have helped the user to access public photos on the web – e.g., through media sharing applications, the amount of available information makes the access to these photos still a less straightforward task. To partly address this challenge, the development of event-related image retrieval systems has been proposed [12]. An event-related image retrieval system is a retrieval system optimized to retrieve all pictures related to a specific event. Here, an event has a specific semantic meaning. Focusing on media-sharing applications, an *event* can be *"something happening in a certain place at a certain time and tagged with a certain term"* [16]. So in an event-retrieval system, the intent of a user might be to retrieve resources related to a particular event, or to use a given tagged photo representing an event to retrieve other photos related to any similar events from a large image collection. Here, we mainly focus on the latter. Due to their characteristics, pictures in photo sharing applications such as *Flickr*[1] and *Panoramio*[2] are particularly interesting. Pictures in such applications are accompanied by contextual metadata, containing heterogeneous fields, such as camera-specific data, *Title*, *Tags*, *Description*, temporal information – i.e., the picture capture and upload times, and geolocation. In this work we study how we can exploit the above metadata in retrieving event-related pictures.

1.1 Goals and Contributions

Due to the characteristics of tags, there are several challenges that we must address. First, *Tags* are unstructured, subjective and full of noise, which would, in turn, affect the retrieval performance. Second, we can assume that many of the queries are short – i.e., containing only few tags, which is in itself a challenge. Third, a complete collection of images from photo-sharing applications is inherently large, which must be dealt with.

In this view, the main goal of this work is to tackle the above challenges focusing on situations where a user searches for pictures related to a specific event, represented by an image with a small number of tags. To our best of knowledge, this area is still new and only few approaches are available – e.g., [19, 12]. Moreover, within information retrieval, existing work has mainly focused on applying temporal information in the retrieval models [8]. At the same time, the most related approach, such as [19], is promising with respect to retrieval performance but mainly applying visual features, thus putting higher requirements on the underlying processing power to achieve web-scalability.

In this paper, we show that by mining and extracting the geo-profiles of the terms in the tags, we can gain an improved retrieval performance, especially when the query images does not contain geo-tags at all. To the best of our knowledge, this is the first work to explore this in depth. Previous work has mainly focused on point-of-interests (POI) extraction [17] and trajectory mining [20]. With the constantly increasing number of geotagged pictures[3] in – e.g., Flickr, exploring this dimension is important.

To this end, the main contributions of our paper are as follows. First, we conduct a study comparing the effectiveness of different retrieval models when using only the textual metadata in event-

24th ACM Conference on Hypertext and Social Media
1–3 May 2013, Paris, France

[1]See `http://www.flickr.com/`
[2]See `http://www.panoramio.com/`
[3]Already in 2009, more than 3.3% – i.e., around 100 Million pictures are geotagged. See also `http://code.flickr.com/2009/02/04/100000000-geotagged-photos-plus/`

related image retrieval. As part of this, we do a thorough analysis of how the different combinations of textual fields affect the retrieval effectiveness depending on the adopted retrieval model. Second, we propose a new weighting model for a query expansion step based temporal proximity in combination with existing term weighting and similarity models. Third, we develop a new extended model that also includes the mined spatial profile for tag terms. Our extensive evaluation shows that using both of our new models yields better retrieval performance than the baseline models, especially with short queries – i.e., pictures with only 1 to 3 tags.

1.2 Related Work

Extraction of pictures related to real-life events is an active research field [16, 6], and in the past decades, detection of events from textual document streams and databases has been extensively treated in the literature [1, 4]. However, despite being an active research field, retrieving and matching event-related images as a field is still less mature. Most existing related approaches have been aimed at *extracting* events from different kinds of datasets. To our best knowledge, only few works have addressed the problems of *retrieval* of events in connection to media sharing. Most of these approaches were presented in the Social Event Detection (SED) task at MediaEval 2011[4] [12], where the main objective was to propose event retrieval systems for Flickr pictures. Most related to our approach is the work by Trad et al. [19]. Similar to our approach, the authors proposed a methods to match a given (query) picture representing an event to pictures representing the same events in a picture collection. The query image is provided with both temporal and spatial information, and the matching algorithm is based first on visual similarity, followed by a reranking step based on geo-temporal coherence. To handle scalability, they use MapReduce in the content analysis and indexing process. The main difference to our work is that rather than applying visual features, our method uses the textual data only. This also allows us to work on a much larger data set.

2. EXTENDED QUERY EXPANSION MODELS FOR EVENT RETRIEVAL

A query expansion approach is a two step approach consisting of (1) choosing the terms to be used in the expansion, and (2) assigning the weight to the chosen terms. With respect to step (1), there are several approaches that have been suggested. Among these, we specifically considered an existing methods for QE that has been proven to be very effective: the Kullback-Liebler (KL) divergence-based approach [5]. With the *KL divergence* approach, the idea is to analyse the term distributions, and maximize the divergence between the distribution of terms from the top-k retrieved documents and the distribution of terms over the entire collection [5]. The terms chosen for the query expansion are those contributing to the highest divergence – i.e., the highest KL-score [5]. This means that expansion terms with low probability in the entire collection and high probability on the retrieved top-k documents will be given more weights than other terms. To calculate the KL-score for a given term t in the feedback (top-k) documents, we use the following equation [5]: $KL = P_{Rel}(t) \log{[P_{Rel}(t)/P_{Coll}(t)]}$, where $P_{Rel}(t)$ and $P_{Coll}(t)$ are the probability that t appears in the top-k documents and the collection, respectively. Here, $P_{Rel}(t)$ can be estimated by the normalized term frequency of t in the top-k documents, whereas $P_{Coll}(t)$ can be computed as the normalized frequency of t in the entire collection.

[4]See `http://www.multimediaeval.org/mediaeval2011/`

After the expansion terms have been selected using one of the approaches above, we can proceed to step (2) – i.e, re-weighting the terms in the query. One of the classical approach to re-weight query terms is the Rocchio's algorithm [15]. More specifically, we use the Rocchio's Beta equation [13] as follows: $\hat{w}(t_q) = tf_{q_{t_q}}/\max tf_q + \beta w(t_q)/\max w$. Here, $\hat{w}(t_q)$ is the new weight of a term t_q of the query; $w(t_q)$ is the weight from the expansion model – i.e., $KL_{Div}(t_q)$; $\max w$ is the maximum weight from the expanded weight model; and $\max tf_q$ is the maximum term frequency in the query and $tf_{q_{t_q}}$ is the frequency of the term in the query.

There are few approaches that have proposed the integration of the temporal information in a pseudo relevance feedback framework. For example, in [9] the authors proposed incorporating time information into the relevance model [11]. The main difference with our approach is the way we use the characteristics of an event and combine this with the temporal proximity of the term distribution as a feature in the term selection process.

We assume that *all* pictures in our collection contains a temporal annotation identifying when the picture was taken – i.e., a timestamp. Further we hypothesize that pictures related to the same event have some temporal proximity or temporal closeness. This means that the more temporally close to the query and the retrieved pictures are, the more likely that they are related to the same event.

Temporal-Proximity Aware KL Divergence. As first improvement, we actively use the assumption about temporal proximity mentioned before. In both of the presented baseline query expansion models, the core premise is that a query expansion word should be more common in the feedback documents and less common in the whole collection.

We extend this assumption as follow. Let a *good* expansion term be a term that is added to the user query or got an increased weight during the retrieval process, and that it improves the retrieval effectiveness – e.g., increasing the mean average precision (MAP). In an event-related retrieval system, we hypothesize that *a term related to the same event as of the user query* is a good expansion term. Hence, we can formulate the following intuition: the distribution of a *good* candidate term should commonly co-occur as much as possible in the documents that is temporally close to the query picture, and less common in the whole collection. This is the same as having a high divergence between the distribution of the co-occurrence of the candidate expansion terms and the query terms in the set of temporal neighbours pictures, and the distribution in the whole collection.

The idea is that in addition to the original KL-divergence computation, our weighting process also considers the divergence of the term distributions within a time slice $\mathcal{L}$, centered in the timestamp of the query image, and the co-occurrence of the term with the query terms within the same time slice. Now, let $\theta^{\mathcal{L}}_{[t,t_i]}$ be the distribution of the co-occurrence between the term t and the query terms $t_i \in Q$ within the set of temporal neighbours, and $\theta^{Coll}_{[t,t_i]}$ denote the distribution of the co-occurrence terms in the whole collection. Then, our temporal-aware KL score can be computed as $sc_Q(t) = \sum_{t_i \in Q} KL(\theta^{\mathcal{L}}_{[t,t_i]} \| \theta^{Coll}_{[t,t_i]})$ – i.e.,

$$sc_Q(t) = \sum_{t_i \in Q} P_{\mathcal{L}}([t,t_i]) \log\left[\frac{P_{\mathcal{L}}(t,t_i)}{P_{Coll}(t|t_i)}\right]. \quad (1)$$

In this re-weighting process, the new weight of a candidate expansion term t is the sum of the divergence between $\theta^{\mathcal{L}}_{[t,t_i]}$ and $\theta^{Coll}_{[t,t_i]}$, for all the $t_i \in Q$. In other words, a candidate expansion term gets a higher weight if the divergence between these two distributions $\theta^{\mathcal{L}}_{[t,t_i]}$ and $\theta^{Coll}_{[t,t_i]}$ is high.

Further, $P_{\mathcal{L}}(t|t_i)$ is the co-occurrence probability of the terms t and t_i within a time interval $\mathcal{L}$, and $P_{Coll}(t|t_i)$ is the co-occurrence probability of the terms t and t_i within the whole collection. We evaluate the co-occurrence probability as proposed in [18] by adding a normalization factor, such that

$$P_{\mathcal{L}}(t|t_i) = \frac{\left[\frac{n_d^{\mathcal{L}}(t,t_i)}{n_d^{\mathcal{L}}(t)+n_d^{\mathcal{L}}(t_i)}\right]}{|\mathcal{D}_{\mathcal{L}}|}, \quad P_{Coll}(t|t_i) = \frac{\left[\frac{n_d^{Coll}(t,t_i)}{n_d^{Coll}(t)+n_d^{Coll}(t_i)}\right]}{|\mathcal{D}|},$$

where $\mathcal{D}$ is the whole dataset and $D_{\mathcal{L}} \subset D$ is a part of a set of the collection documents D that have timestamps within the time interval $\mathcal{L}$. $n_d^{\mathcal{L}}(t,t_i)$ and $n_d^{Coll}(t,t_i)$ are the number of documents in the set $D_{\mathcal{L}}$ and D, respectively, in which the terms t and t_i co-occur. Similarly, $n_d^{\mathcal{L}}(t)$ and $n_d^{Coll}(t)$ are the number of documents tagged with the term t in the set $D_{\mathcal{L}}$ and D, respectively.

To include the influences of both scores in the calculation of the final expansion weight $W(t)$, the last two models can be mixed in a linear combination, given by $KLT(t) = \gamma KL(t) + (1-\gamma)KL^{\mathcal{L}}(t)$, where γ is factor used to determine the amount of influence each score has on the final weight.

Exploring Term Spatial Distribution. As explained in our hypothesis related to the meaning of *event*, pictures related to the same event tend to appear in a limited geographical area. In this work, we consider query pictures that are not geotagged since this would mostly reflect the reality of media sharing application and social media, in general. For example, considering Flickr database, only 3.3% of the pictures are geotagged, i.e.– around 100 million of pictures. So the probability of having a geotagged picture is low. Still, having this amount of pictures can be useful in extracting the spatial profiles of the tag terms.

We propose a method to find a *good* expansion term t, given a set of query terms $Q = \{t_i\}_i$. By including the spatial dimension (profiles), a good expansion term is a term related to the same event as the query picture. This means that a good expansion term are those commonly co-occurring in documents that are temporally close to the query picture and in a geographic delimited area, and less common in in the whole temporal timeline in the same delimited area. To mimic a real-world problems, we only consider query pictures without geotagged.

In this work, we define picture locations by discretizing the world map. Inspired by [22], we first divide the world map into M *tiles* $\Theta = \{\mathcal{T}_k\}_{k=1..M}$ of size 1 degree. Since the size corresponding to 1 degree varies depending the latitude values, the discretization process does not produce equal tiles. Nevertheless, this approximation is still useful since most of the areas with high population density are mainly close to the latitude near the equator.

To include the spatial dimension to score a candidate expansion term, we start from a similar hypothesis to the one proposed above but including the geographical dimension. This means that a good expansion term t is the one that yield a high divergence between the distribution of the pictures tagged with the query terms, the expansion terms in a temporal time slice $\mathcal{L}$ and a tile $\mathcal{T}_k$; and the distribution of terms still in the geographical tile $\mathcal{T}_k$ but occur in the whole timeline. Formally, this divergence is computed based on the KL-divergence as $sc_Q^3(t,\mathcal{T}_k) = \sum_{t_i \in Q} KL(\theta^{\mathcal{L}}_{[t,t_i,\mathcal{T}_k]} \| \theta^{Coll}_{[t,t_i,\mathcal{T}_k]})$ – i.e.,

$$sc_Q^3(t,\mathcal{T}_k) = \sum_{t_i \in Q} P_{\mathcal{L}}(t|t_i,\mathcal{T}_k) \log \left[\frac{P_{\mathcal{L}}(t|t_i,\mathcal{T}_k)}{P_{Coll}(t|t_i,\mathcal{T}_k)}\right]. \quad (2)$$

Here, $P_{\mathcal{L}}(t|t_i,\mathcal{T}_k)$ is the co-occurrence probability of the query term t_i and expansion term t, within a time interval $\mathcal{L}$ and a geographical tile $\mathcal{T}_k$. Similarly, $P_{Coll}(t|t_i,\mathcal{T}_k)$, is the same probability without the temporal restriction. We approximate this probability as follows:

$$P^G_{Coll}(t|t_i,\mathcal{T}_k) = \frac{\left[\frac{n_d^{Coll}(t,t_i|\mathcal{T}_k)}{n_d^{Coll}(t|\mathcal{T}_k)+n_d^{Coll}(t_i|\mathcal{T}_k)}\right]}{|\mathcal{T}_k|} \quad (3)$$

$$P^G_{\mathcal{L}}(t|t_i,\mathcal{T}^{\mathcal{L}}_k) = \frac{\left[\frac{n_d^{\mathcal{L}}(t,t_i|\mathcal{T}_k)}{n_d^{\mathcal{L}}(t|\mathcal{T}_k)+n_d^{\mathcal{L}}(t_i|\mathcal{T}_k)}\right]}{|\mathcal{T}^{\mathcal{L}}_k|} \quad (4)$$

We compute the pair of probabilities $P^G_{Coll}(t|t_i,\mathcal{T}_k)$ and $P^G_{\mathcal{L}}(t|t_i,\mathcal{T}^{\mathcal{L}}_k)$ for each tile $\mathcal{T}_k \in \Theta$. Then, we calculate the divergence between the two distribution values, tile by tile, and chose the maximum divergence value as the final score. In order to include the influence of KLT, we mix the models in a linear combination as follows: $KLST(t) = \sigma KLT(t) + (1-\sigma)\max\{KL^{\mathcal{L}}_{\mathcal{T}_k}(t)\}_{\mathcal{T}_k}$, where $KL^{\mathcal{L}}_{\mathcal{T}_k}(t)$ is the KL divergence between the distributions in Equation 3 and Equation 4.

3. EXPERIMENTAL RESULTS

3.1 Dataset and Evaluation Metrics

To evaluate our method, we use the *Upcoming dataset* [3] as the ground-truth for our experiments. This dataset consists of 270.425 pictures from Flickr, taken between 1st of January 2006 and 31st December 2008, each of which belongs to a specific event from the Upcoming event database[5], with 9.515 unique events. Each event is composed by a variable number of images, varying from 1 to 2.398 pictures. Further, to make our experiment more realistic with respect to scalability, we decided to build an additional dataset by merging the Upcoming dataset with other pictures gathered from Flickr[6] covering a time period from 01.01.2006 to 31.12.2010 and without spatial restrictions. Our final dataset now contains 88 million pictures, of which approx. 19 millions are without any tags, and around 23.5% with 1 to 3 tags.

To perform our experiments, we first indexed all image tags using Terrier[7]. As part of the dataset preparation we perform a preprocessing step consisting of tokenization based on whitespace and punctuation marks, and English stopword removal. Then, we randomly selected a set of pictures from each event cluster in the Upcoming dataset and use these as queries.

To assess the effectiveness of our approach, we compare our models with existing models. Such models serve as baselines for our evaluation, including the *Vector Space Model* (**TFIDF**) [2], *Okapi BM25* (**BM25**) [14], *Hiemstra Language Modelling* (**LM**) weighting model [7] and *KL divergence retrieval model* (**KLDM**) [10]. For both BM25 and LM, we use the default parameter values – i.e., for BM25 we set $k_1 = 1.2$, $k_3 = 8$ and $b = 0.75$, and for LM is $c = 0.15$.

To evaluate the retrieval performance, we use standard in information retrieval evaluation metrics, including the Mean Average Precision (MAP) and R-Precision (RP) [2]. Further, to make sure that any improvements are statistically significant, we perform paired two-sample one-tailed t-tests at $p < 0.05$ or 95 % confidence interval. Therefore, any stated improvements in this paper are all statistically significant, unless otherwise specified.

3.2 Results

[5]See `http://upcoming.yahoo.com/`

[6]We used Flickr API to do this. See also `http://www.flickr.com/services/api/`

[7]Terrier is a well-known java-based open source programming library tool for retrieval. See `http://www.terrier.org/`

3.2.1 Field Effectiveness

Our first experiment aims at exploring the effectiveness of using Flickr images as queries. To assess this effectiveness and analyse the role of the fields in the metadata, we use different combinations of the textual metadata as queries and document representation. Specifically, we evaluate how *Title*, *Tag* and their combination affect the retrieval effectiveness. To do this, we first represent the documents by *Title* only, then by *Tag* only, and finally by *Description* metadata only. Thereafter, we test different combinations of these fields as follows: *Title* and *Tag*; and *Title*, *Tag* and *Description*. The set of queries is formed by randomly selecting 1 picture from each event cluster in the Upcoming Dataset. Only event clusters containing more than 500 pictures from a total of 50 clusters are considered. Hence, the total number of queries is 50 for each sample. This random sampling is repeated 5 times in order to obtain five set of 50 queries, giving us 250 queries in total.

	TFIDF		**BM25**		**LM**		**KLDM**	
Comb	MAP	RP	MAP	RP	MAP	RP	MAP	RP
	Query: *Tag* field							
TAG_{TAG}	.685	.695	**.687**^{234}	**.697**^{234}	.691	.704	**.693**^{23}	**.704**^{23}
TAG_{TIT}	.064	.082	.085	.105	.067	.083	.064	.081
TAG_{DES}	.281	.290	.281	.287	.434	.448	.281	.288
TAG_{TT}	**.695**^{123}	**.707**^{123}	.530	.540	**.696**^{123}	**.708**^{123}	**.691**^{23}	**.704**^{23}
	Query: *Title* field							
TIT_{TAG}	.498	.502	**.500**^{234}	**.506**^{234}	**.484**^{234}	**.492**^{234}	**.503**^{234}	**.510**^{234}
TIT_{TIT}	.350	.358	.324	.332	.357	.364	.353	.360
TIT_{DES}	**.550**^{124}	**.559**^{124}	.459	.467	.460	.468	.460	.468
TIT_{TT}	.113	.129	.106	.124	.127	.140	.130	.147
	Query: *Tag* and *Title* field							
TT_{TAG}	.663	.680	**.669**^{23}	**.686**^{23}	.468	.484	**.667**^{23}	**.682**^{23}
TT_{TIT}	.117	.139	.108	.129	.120	.144	.129	.154
TT_{DES}	.369	.376	.287	.295	.288	.297	.289	.297
TT_{TT}	**.673**^{123}	**.690**^{123}	**.665**^{23}	**.683**^{23}	**.693**^{123}	**.705**^{123}	**.670**^{23}	**.686**^{23}

Table 1: MAP and R-precision (RP) by querying the dataset using the *Tag* field only, *Title* field only, and both fields. The numbers 1, 2, 3 and 4 in superscript indicate the statistical significance improvements on the dataset indexed with TAG field, TIT field, DES field and TT fields, respectively.

Table 1 summarizes the results from our experiments using different fields of the pictures as queries. Here TAG_{TAG} means we use the tag field in both the indexing and the query, whereas TAG_{TIT} means we apply tag (*TAG*) in the indexing but title (*TIT*) in the query, and so on. *TT* stands for tags and title combination, while *DES* is the description field.

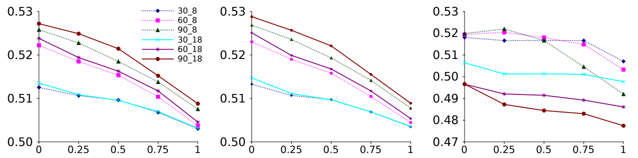
With the results in the tables above, we can make the following observations. First as shown in Table 1, querying using the title resulted in the lowest MAP and R-precision (RP) values compared to querying with the title and the tags. Further, with the best results in each section of the table, for each retrieval model, the most representative field for each picture was the *Tag* field, with which the MAP and the RP values were the highest. Finally, when querying using Title, either in combination with the Tag field or alone, we can see that in all the cases, the highest MAP and RP values were obtained when the same fields were used both to represent the documents/images and to generate the set of queries. In summary, since these results are conclusive, we can safely base our experiments to test our query expansion (QE) step using the combination TAG_{TAG}.

3.2.2 Evaluating the Extended QE Models

In this experiment, we evaluated the approaches proposed in Section 2. Similar to the previous experiment, we first randomly selected 100 queries from the event clusters, containing more than 100 pictures. Then we selected pictures with less than 3 tags.

To perform a complete set of experiments, we considered different values of the following parameters. First, as query expansion parameters, we varied the value of K such that $K \in \{30, 60, 90, 120\}$ and the values of n such that $n \in \{8, 18\}$. Second, as a parameter for **KLT**, we varied the time slice $\mathcal{L}$ in the following set: {1 day, 3 days, 7 days}[8].

(1) The Impact of γ on Mixed KL. With this set of experiments, we tested the impact of the parameter γ in Section 2 used to linearly combine the **KLT** and the standard **KL** divergences. We varied its values from 0 to 1 such that $\gamma \in \{0, 0.25, 0.5, 0.75, 1\}$, 0 means that we only have the contribution of **KLT** and to 1 we only have the contribution of **KL**. We repeated the experiment for the six combinations of the number of query expansion terms n, and the number of top-K documents considered in the query expansion process.



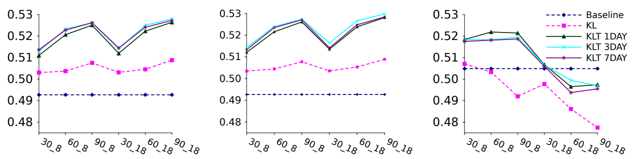
(a) TFIDF (b) BM25 (c) LM

Figure 1: MAP values as function of K and n (expressed as $\{K\}\_\{n\}$), for the 3 different retrieval and query expansion (QE) models, with different values of γ.

Figure 1 shows the MAP values as function of the different values of γ. As can be seen in this figure, for all the six combinations, the MAP values decreased when the γ values decreased. This means that the mix of both of the contributions was not effective, but the most important contribution came from our **KLT** divergence.

(2) KL vs KLT. To further assess the performance of our **KLT** approach, we compared it with the baseline approach, using the linear combination above, with $\gamma = 0$.

First, we compared **KL** and **KLT**. The result from this experiment is summarized in Figure 2. As can be observed, by using BM25 and TFIDF retrieval models in the initial retrieval step, our **KLT** outperforms **KL**, with all combinations of K and n. With LM and $n = 30$, the **KLT** also outperforms the baseline model. With $n = 60$, the **KLT** still outperforms **KL** but in this case the query expansion process is not so effective. All the improvements from **KL** and the baseline methods are statistically significant at 95 % confidence interval.



(a) TFIDF (b) BM25 (c) LM

Figure 2: The MAP values as function of K and n (expressed as $\{K\}\_\{n\}$), with the 3 different retrieval and QE models.

(3) KLT vs KLST. In this subsection we compare the temporal-aware query expansion model with the spatiotemporal-aware query

[8]In addition to the above models, we also implemented the *Mixture Model* [21] and the Relevance Model [11]. However, the results were comparable to the KL query expansion models.

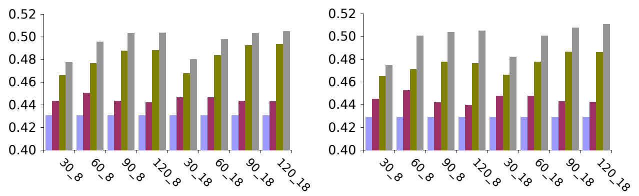
expansion model **KLST**. We use the values of $\gamma = 0$ for the linear combination between **KL** and **KLT**, which has been shown to yield the best result. Further, we set $\delta = 0.5$ to compute $KLST(t)$ as discussed in Section 2. Due to the space limitation, we do not present any tuning process for the δ value. Table 2 shows the percentage of improvement of MAP and RP values between the **KLST** query re-weighting model and the other models. The numbers 1, 2 and 3 in superscript in the table indicate the statistical significance improvements on the baseline, **KL** and **KLT** re-weighting models, respectively.

	Δ**MAP**(%)			Δ**RP**(%)		
	KL	**KLT**	**KLST**	**KL**	**KLT**	**KLST**
30_8	2.97	8.20	**10.84**^{12}	3.74	8.29	**10.59**^{12}
60_8	4.57	10.66	**15.06**^{123}	5.46	9.79	**16.66**^{123}
90_8	2.91	13.19	**16.77**^{123}	3.02	11.30	**17.42**^{123}
120_8	2.61	13.34	**16.94**^{123}	2.47	11.01	**17.63**^{123}
30_18	3.62	8.58	**11.44**^{123}	4.35	8.61	**12.30**^{123}
60_18	3.62	12.25	**15.55**^{123}	4.35	11.30	**16.63**^{123}
90_18	2.95	14.33	**16.85**^{123}	3.19	13.38	**18.30**^{123}
120_18	2.84	14.52	**17.25**^{123}	3.14	13.29	**19.05**^{123}

Table 2: Percentage of improvement of MAP and RP using different re-weighting model on BM25.

In all the retrieval models, we observe that **KLST** outperforms both the baseline KL reweighting model, **KL**, and the temporal-proximity-aware model, **KLT**. Without loss of generality Table 2 shows the results using the **BM25** retrieval model, considering $3DAY$ as temporal window. As can be observed from this table, **KLST** is from 10.6% to 19% better than the baseline method. Moreover, with the best MAP and RP values, **KLST** is six times better than the baseline **KL** and around 50% better than the **KLT** model.

Figure 3 further illustrates our comparison with respect to MAP and RP values. This figure confirms our observation about the effectiveness of our **KLST** model.



(a) MAP (b) RP

Figure 3: Comparison of the MAP and RP values for KLT and KLST against the baseline query expansion model, as function of the values of K and n (expressed as {K}_{n}), using BM25.

4. CONCLUSIONS

Photosharing applications, such as Flickr, contain many pictures related to real life events, many of which are annotated with time and location information. The main goal of this work has been to improve existing retrieval models by exploiting this information in search of event-related images. To achieve this goal, we have proposed an extended query expansion model that exploits the temporal information of the pictures and the spatial distribution of the terms. We thoroughly evaluated our approach by first analysing the retrieval effectiveness with respect to different combinations of metadata fields, and using different standard retrieval models. Thereafter, we conducted several experiments to assess the effectiveness of our two proposed query expansion models; one based on temporal proximity of tag terms, and an other based on spatial distribution of tag terms. We compared both methods with existing baseline approaches. The results of these experiments have shown that our approach outperforms the state-of-the-art query expansion models, and that the improvements were statistically significant at a $p < 0.05\%$ level. In particular, we demonstrated that our method is effective even when the amount of information surrounding a picture is small. Finally, by testing our approach on a large dataset, and still getting good results, we can conclude that our approach is also scalable.

Acknowledgments

This work is supported by the Research Council of Norway, grant number 176858 under the VERDIKT program.

References

[1] J. Allan, R. Papka, and V. Lavrenko. On-line new event detection and tracking. In *Proc. of ACM SIGIR 1998*, pages 37–45. ACM, 1998.

[2] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology behind Search*. Addison-Wesley, New York, 2011.

[3] H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in social media. In *Proc. of WSDM 2010*, pages 291–300, 2010.

[4] T. Brants, F. Chen, and A. Farahat. A system for new event detection. In *Proc. of ACM SIGIR 2003*, pages 330–337. ACM, 2003.

[5] C. Carpineto, R. de Mori, G. Romano, and B. Bigi. An information-theoretic approach to automatic query expansion. *ACM Transaction on Information System*, 19:1–27, 2001.

[6] N. Gkalelis, V. Mezaris, and I. Kompatsiaris. A joint content-event model for event-centric multimedia indexing. In *Proc. of the 2010 IEEE Fourth International Conference on Semantic Computing (ICSC '10)*, pages 79–84. IEEE Computer Society, 2010.

[7] D. Hiemstra. *Using language models for information retrieval*. Univ. Twente, 2001. ISBN 978-90-75296-05-1.

[8] R. Jones and F. Diaz. Temporal profiles of queries. *ACM Trans. Inf. Syst.*, 25(3), July 2007. ISSN 1046-8188.

[9] M. Keikha, S. Gerani, and F. Crestani. Time-based relevance models. In *Proc. of ACM SIGIR 2011*, pages 1087–1088. ACM, 2011.

[10] J. Lafferty and C. Zhai. Document language models, query models, and risk minimization for information retrieval. In *Proc. of ACM SIGIR 2001*, pages 111–119. ACM, 2001.

[11] V. Lavrenko and W. B. Croft. Relevance based language models. In *Proc. of ACM SIGIR 2001*, pages 120–127. ACM, 2001.

[12] S. Papadopoulos, R. Troncy, V. Mezaris, B. Huet, and I. Kompatsiaris. Social Event Detection at MediaEval 2011: Challenges, Dataset and Evaluation. In *MediaEval 2011 Workshop*, Pisa, Italy, September 1-2 2011.

[13] J. Pérez-Agüera and L. Araujo. Comparing and combining methods for automatic query expansion. *Advances in Natural Language Processing and Applications Research in Computing Science*, 33:177–188, 2008.

[14] S. E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proc. of ACM SIGIR 1994*, pages 232–241, 1994.

[15] J. Rocchio. *Relevance Feedback in Information Retrieval*, pages 313–323. 1971.

[16] M. Ruocco and H. Ramampiaro. A scalable algorithm for extraction and clustering of event-related pictures. *Multimedia Tools and Applications Journal (MTAP)*, pages 1–34, 2012. ISSN 1380-7501.

[17] M. Ruocco and H. Ramampiaro. Heterogeneous tag-point patterns for ranking and extracting hot-spot related tags. In *Proc. of the 4th ACM SIGSPATIAL Workshop on Location-Based Social Networks (LSBN 2012)*. ACM, 2012.

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

[19] M. R. Trad, A. Joly, and N. Boujemaa. Large scale visual-based event matching. In *Proc. of ACM ICMR 2011*, pages 53:1–53:7. ACM, 2011.

[20] Z. Yin, L. Cao, J. Han, J. Luo, and T. S. Huang. Diversified trajectory pattern ranking in geo-tagged social media. In *SDM*, pages 980–991, 2011.

[21] C. Zhai and J. Lafferty. Model-based feedback in the language modeling approach to information retrieval. In *Proc. of CIKM 2001*, pages 403–410. ACM, 2001.

[22] H. Zhang, M. Korayem, E. You, and D. J. Crandall. Beyond co-occurrence: discovering and visualizing tag relationships from geo-spatial and temporal similarities. In *Proc. of ACM WSDM 2012*, pages 33–42. ACM, 2012.