Collaborative event annotation in tagged photo collections

Article *in* Multimedia Tools and Applications · May 2012

DOI: 10.1007/s11042-012-1154-5


Christos Zigkolis
Aristotle University of Thessaloniki


Symeon Papadopoulos
The Centre for Research and Technology, Hellas


Ioannis (Yiannis) Kompatsiaris
The Centre for Research and Technology, Hellas


Athena Vakali
Aristotle University of Thessaloniki

Noname manuscript No.
(will be inserted by the editor)

Collaborative Event Annotation in Tagged Photo Collections

Christos Zigkolis · Symeon Papadopoulos · George Filippou · Yiannis Kompatsiaris · Athena Vakali

Received: date / Accepted: date

Abstract Events constitute a significant means of multimedia content organization and sharing. Despite the recent interest in detecting events and annotating media content in an event-centric way, there is currently insufficient support for managing events in large-scale content collections and limited understanding of the event annotation process. To this end, this paper presents CrEve, a collaborative event annotation framework which uses content found in social media sites with the prime objective to facilitate the annotation of large media corpora with event information. The proposed annotation framework could significantly benefit social media research due to the proliferation of event-related user-contributed content. We demonstrate that, compared to a standard "browse-and-annotate" interface, CrEve leads to a 19% increase in the coverage of the generated ground truth in a large-scale annotation experiment. Furthermore, the paper discusses the results of a user study that quantifies the performance of CrEve and the contribution of different event dimensions in the event annotation process. The study confirms the prevalence of spatio-temporal queries as the prime option of discovering event-related content in a large collection. In addition, textual queries and social cues (content contributor) were also found to be significant as event search dimensions. Finally, it demonstrates the potential of employing automatic photo clustering methods with the goal of facilitating event annotation.

Keywords event authoring · multimedia annotation · ground truth generation

1 Introduction

The digital presence and online activities of users in social media sites is largely a reflection of their everyday life experiences. Social network users share such content on a daily basis leading to a rapid increase in the size of the media content available

Christos Zigkolis · Symeon Papadopoulos · Yiannis Kompatsiaris
Informatics and Telematics Institute, CERTH
E-mail: {chzigkol,papadop,ikom}@iti.gr

Christos Zigkolis · Symeon Papadopoulos · George Filippou · Athena Vakali
Department of Informatics, AUTH
E-mail: {gfilip, avakali}@csd.auth.gr

online. Due to its "social" nature, a large fraction of user-contributed content pertains to the participation and coverage of real-life events, such as concerts, sports games and celebrations. Naturally, such content abundance creates the need for efficient content organization and indexing in order for such content to be efficiently searchable.

The requirement for event-centric content organization also stems from social media research, where vast amounts of content need to be annotated with reliable event information, thus enabling large-scale studies of real-world social activities and comprehensive benchmarks of the performance of event mining methods. To date, multimedia annotation has largely relied on manual effort, which is time-consuming and error-prone. For instance, the annotation of the NUS-WIDE photo collection, comprising 269,648 images that are annotated with 81 concepts was estimated to have cost about 3000 man-hours [7]. Still, this large time effort was possible, because the ground truth creators employed heuristics to reduce the effective dataset size that the annotators would go through, at the expense of a small annotation error. Similar ground truth annotation efforts, such as MIR-Flickr [12], also employ annotation heuristics in order to speed up the annotation process, while others, such as TRECVid [33] rely on result pooling, thus leading to the creation of non-exhaustive annotation. It becomes clear that new sophisticated tools are necessary for the support of annotating large media collections.

Event-centric annotation has recently attracted interest [22] facing even harder annotation challenges. Events constitute highly complex concepts and are defined at multiple levels of granularity. For instance, an event can last from few hours (e.g. small gig at a bar) to many days (e.g. large music festival). Similarly, in terms of space, events may be both focused (e.g. a speech taking place in a specific room) and highly scattered (e.g. a parade spanning a large part of a city centre). Other times, the spatio-temporal features of content items are not available or are simply insufficient to associate them with a target event. Existing annotation tools (cf. Section 2.1) are mostly based on a single type of information (e.g. temporal) to group content which is restrictive given that there are additional dimensions associating content with an event (location, participants, etc.).

Taking the aforementioned limitations into consideration, our primary objective in this study is to provide support to the process of event-centric annotation by proposing the **CrEve** framework[1] (CrEve is an abbreviation of the phrase Create Event). This framework offers users a set of query facilities covering different aspects of data (i.e. location, time, textual, ownership etc) for exploring the collection and associating photo content with real-life events. The proposed framework addresses the following issues:

– **Magnitude of photo collections:** Discovering only the part of content related to events is a non-trivial task given the great number of available photos in large-scale event annotation problems. CrEve offers query facilities based on textual, spatial, temporal and provenance filters with the goal of facilitating the photo collection exploration process along different dimensions of an event.
– **Inconsistency of photo metadata:** The quality of photo metadata is highly variable in large collections of user-generated content. There are numerous instances where tags or geographic information are missing. In such cases, inferred relations among photos, as a result of the visual similarity search and event-based clustering capabilities offered by our framework are useful for discovering photos relevant to a target event and thus increasing the recall of the resulting ground truth.

[1] A demo version of CrEve can be found here, http://www.clusttour.gr/creve

- **Annotation bias and quality:** Users capture events from different perspectives. In addition, real-life events often involve more than one scenes. Annotating complex real-life events is thus particularly prone to subjective bias and the quality of annotation greatly depends on the prior knowledge of the annotators with respect to the event. CrEve supports the collaborative aspect of annotation by making annotators aware of other users' annotations with the goal of establishing a shared view among the annotators with respect to the event of interest.

Apart from the proposed event annotation framework, we also conducted a study with the goal to evaluate both the utility and effectiveness of CrEve and the contribution of different event dimensions in the annotation process. Therefore, the main contributions of this paper can be summarized as follows:

- **Creating event-based annotations for photo collections:** The proposed framework for annotating event-related content provides a series of search and annotation functionalities in order to manage and annotate large collections of user-contributed photos with event information. CrEve offers a thorough list of event dimensions by taking different parts of data into consideration as opposed to previous studies that only support a subset of that list. We demonstrate through our user study that the use of CrEve results in annotations of higher quality and larger coverage compared to a standard "browse-and-annotate" interface (i.e. interfaces that support mere text search and simple sorting in terms of date).
- **Understanding the event annotation process:** Through the conducted user study, insights are gained with respect to the event dimensions used by annotators to discover event-related photos. We confirm the great significance of textual queries in the process of seeking event-related photos, as well as the important contribution of the temporal and spatial aspects when used in composite queries. Furthermore, we demonstrate that automatic photo clustering can be a valuable facility in the event annotation process.
- **Maintaining event-based annotations for photo collections:** CrEve is not limited to the creation of event-based annotations, but offers facilities to support their maintenance. Making annotators aware of the annotations of other users and providing web-based viewing and editing capabilities for them (list of events and associated photos) in a large photo collection enables the continuous improvement of the ground truth quality and its verification by a large number of individuals.

CrEve could be also useful to anyone who is interested in gathering and organizing media content for events. More specifically, we see a potential use case in communities of users ranging from event enthusiasts (e.g. fans of a rock band) to event professionals (e.g. organizers of an event) that systematically collect photos from different sources regarding specific events. The fact that the related content for an event is likely to be distributed in more than one owners adds extra value to CrEve which can bring all that data together. In addition, it is expected that these communities would want to do much of their work in a collaborative manner in order to get a good result which means a high-quality dataset of photos from the events they have in mind. Apart from the aforementioned groups of people, even single users may benefit from CrEve especially in the case of organizing the photos of their personal collection in events.

The rest of the article is structured as follows. Section 2 discusses the prior art related to different research aspects of this work. Section 3 presents the proposed annotation tool. Section 4 describes the user-based evaluation of the tool and discusses the obtained results. Section 5 concludes the paper and discusses future work.

2 Background

2.1 Related work

The advent of social media sites has brought not only the opportunity for users to share media content with others but also new ways of organizing and discovering content through the use of free-form keywords, named tags. This annotation process, known as Folksonomy [39], serves both personal and social purposes. Exploring these purposes would give us an answer to the question why people make use of tags. Authors in this study [1] presented a taxonomy of motivations for annotations which includes two dimensions, sociality and function. The first one has to do with the tag's intended usage, whether it is used by the the user who uploaded the photo or by the other users of a community (i.e. self and social). The second dimension refers to a tag's intended uses where users tagged photos either to facilitate later organization and retrieval or to communicate some additional context to the viewers of the photo. In a different work, Marlow et al. [16] present a list of user incentives for annotation in order to give insight how these motives can influence the use and utility of tags in tagging systems. The introduction of GPS technology enabled users to attach location-based information to media content which along with time information have proved to be useful contextual metadata for retrieving purposes. Naaman et al. [18], apart from using location and time to organize photos, they utilize these kind of data as generators of additional contextual metadata coming from external data sources in order to accompany photos with more context information.

While tags seem very promising source of information, they have some limitations that restrict their usability [11]. Invalid metadata, tag synonymy and redundancy are some of the problems that come with the tagging processes. Therefore, many studies from different research fields have been conducted in order to provide and efficient organization of all this information. Begelman et al. [4] made use of clustering techniques in order to overcome some limitations in terms of search and exploration in social tagging systems with the goal to enhance user experience and improve the use of tag space in general. Rattenbury et al. [25] presented a generalizable approach for extracting event and place semantics based on distribution of individual tags. Their technique can be seen as a classifier through which it is decided whether a tag refers to a place of an event. Tag recommendations is another active research field that aims to help the user in the annotation process. In this study [32] the authors provide an automatic way through which tags are suggested to users when they add a photo. They proceeded to an analysis of how people tag photos and what information is contained in the tagging in order to evaluate their recommendation strategies.

Beyond simple tagging, events have recently emerged as a convenient means for organizing multimedia content. Related works in event-based annotation mainly deal with three event aspects: (a) event representation, (b) event detection, and (c) event annotation. In this section, we briefly review some important works in each of these areas in order to delineate the paper scope. Apart from the aforementioned topics, a brief overview of studies on conventional tagging is included in order to present the motivations behind the use of tags.

Event representation: The paper focuses on real-world events that are captured in user generated media content and shared through social networking sites. There have been many efforts to capture and model the semantics of events ranging from wide-scope models, such as the one by Westermann and Jain [40] and the Event-Model-F [28], to more pragmatic models, such as the 5W1H journalistic event model by Xie et al. [42], and the lightweight LODE ontology [29] for easily publishing events in the form of Linked Data. In all cases, the following important aspects are considered: (a) time (When), (b) location (Where), (c) participants (Who), (d) information (What). Furthermore, many of the models consider causality (Why) and hierarchical structure (events consisting of sub-events).

The event annotation tool presented in this paper supports the four universal event dimensions, as well as a simple event structuring mechanism (event consisting of sub-events). In addition, CrEve supports *annotation provenance*, i.e. explicit representation of the association between an annotator and the event-to-media linking, which is particularly important in the case of ground truth generation. Finally, the proposed tool introduces the concept of *candidate sets* of media items, i.e. item sets that are constructed in an ad hoc manner (e.g. as a result of a user query) and need to be reviewed in order to be linked to a target event. The notion of Candidate Set has been used in recommender systems and especially in Cascade-Hybrid Algorithms [5] where a staged procedure is involved. These techniques produce an initial set of items as a first step (i.e. candidate set) and then, these items are filtered in order to keep the most suitable ones for the target user. We can relate our methods of creating candidate sets with the rationale of the cascade-hybrid procedures with respect to the fact that both approaches are responsible to search into a wide space of objects and select the ones that seem to be most appropriate. In line with the first stage of a cascade-hybrid algorithm is the problem of preference-based multi-criteria item search where narrowing down to a subset of items is vital when it comes to finding the best item. In that case, reducing user's effort is the ultimate objective [24].

Event detection: Recently, significant interest has been shown in the detection of content of real-life events in multimedia collections. For instance, a benchmarking contest, named MediaEval, has been held in order to evaluate new algorithms for multimedia access and retrieval. In this contest, a specific task was dedicated to the detection of social events [22] [36]. In [23], the authors used a composite approach consisting of a candidate photo selection, clustering and event expansion methods in order to cope with the challenges of this task.

In order to deal with the massive amounts of media content involved in large media annotation tasks, automatic schemes for event detection in media content are indispensable. For instance, the scheme by Sayyadi et al. [27] identifies events in streams of online articles by organizing their text content around topically related keyword clusters. Recent approaches [26], [20] identify landmarks and events in large tagged photo collections by clustering photos based on their textual and visual similarity and then classifying the resulting clusters as landmarks or events. Wu et al. employ a user-contributed video mining framework involving burst detection, keyword mining, and near duplicate keyframe detection, with the goal of organizing large video content collections in events [41]. Finally, Liu et al. employ and evaluate the effectiveness of simple query schemes based on the basic event dimensions (when, where, what, who) in order to identify media items that are associated with a given event [15].

CrEve makes use of the photo clustering framework presented in [20] in order to present annotators with a list of candidate events that can be used as a starting place for the annotation process. Furthermore, the tool offers query mechanisms similar to the ones of [15] with the goal of enriching the set of media items that are potentially relevant to the event of interest.

Event annotation tools: At its core, the paper focuses on the problem of event annotation. Shneiderman et al. envision a combination of annotation, browsing, and sharing of photos as an effective means of addressing the content organization needs that emerge in large scale personal photo collections [31]. Appan et al. [3] propose a collaborative media annotation tool with sophisticated personalization, recommendation and visualization capabilities aiming at facilitating the review and annotation of large photo collections. In [2], the authors specialize this interface for event-focused media. In a different work, Suh and Bederson proposed a generic media annotation interface that makes use of event detection with the goal of enabling bulk photo annotation by users [34]. In another work [45] that deals with video content the authors presented a narrative system where users can upload their personal recordings from social events and later, through a number of configuration parameters, they can filter and take the most appropriate content from a shared repository. In line with these works, CrEve adopts the "search and explore" paradigm for event annotation, offering different options for searching in the content collection of interest for media items associated with a given event.

Most event annotation tools are restricted to a monolithic content exploration paradigm by relying on a single criterion to browse event-related content. This is clearly insufficient when it comes to discovering numerous event aspects in large content collection. The different search options offered by CrEve enable the discovery of event-related content following a variety of content exploration paths. In addition to state-of-the-art query mechanisms, the tool offers capabilities for *reviewing the query history* related to a specific event (i.e. the set of queries issued by annotators with the goal of finding candidate media items for associating with events), thus providing prospective annotators with insights regarding the event structure of the collection. Finally, in contrast to previous studies that evaluate only the user satisfaction from the proposed annotation tool, this paper evaluates the contribution that the different query mechanisms have to the discovery of media items associated with the event of interest.

2.2 Problem Definition and Notation

Here, we provide a formal definition of the research problem we deal with and the associated notation that we employ throughout the paper. In this paper we restrict our discussion to photo collections denoted by $\mathbb{P} \triangleq \{p\}$, where p is a tuple $(\theta_p, l_p, t_p, u_p)$ containing a unique photo ID, θ_p, geotagged with location information l_p (consisting of a pair of latitude-longitude coordinates), captured at time t_p and uploaded by user u_p. The second element in the input dataset is the set of tags associated with each photo. We use x to denote a tag. Each photo p can be associated with multiple tags, i.e. with the set $\mathbb{X}_p$ of tags. For convenience, we define the subset of photos associated with a specific tags as $\mathbb{P}_x \triangleq \{p \in \mathbb{P} | x \in \mathbb{X}_p\}$. Although we include only Flickr photos in our study, it is possible to apply the processes presented in this paper to videos as well by incorporating key frame selection techniques. For instance, in this study [17] the same clustering framework to both Flickr collections and Youtube videos is applied. Using a key-frame selection method will bring us to a frame-photo collection, thus there is no difference from the collection we made use here.

In addition to the available media content and associated metadata, we consider a collection of events $\mathbb{E} \triangleq \{e\}$, each of which is captured in at least one photo in $\mathbb{P}$. Each event is represented by a tuple $(\theta_e, T_e, l_e, t_{p,0}, t_{p,1})$, where θ_e denotes the event id, T_e the event title, l_e the geographical centre of the event location, and $t_{p,0}$, $t_{p,1}$ mark the beginning and the end time of the event. Similar to photos, an event can be associated with a set of tags $\mathbb{X}_e$. Finally, we consider the set of users-annotators $\mathbb{U} \triangleq \{u\}$ who make use of the proposed tool with the goal of producing a set of annotations $\mathbb{A} \triangleq \{a\}$. Each annotation is represented as a tuple (p, e, u, q), expressing the association of a photo p with an event e by annotator u who used query q to retrieve the photo from the original collection. The latter is particularly helpful in evaluating the utility of the query facilities offered by the framework. Given the notation introduced above, the two tasks supported by the proposed event annotation framework can be formally described as follows:

Definition 1 (Event Discovery in Media Collections)
Given a large collection of media items $\mathbb{P}$, identify the set of events $\mathbb{E}$ that are captured by at least a photo in the collection.

Definition 2 (Event-Media Association)
Given a large collection of media items $\mathbb{P}$ and an event specification e, find all photos in the collection that illustrate the given event, i.e. the set $\mathbb{P}_e$ of photos.

3 The CrEve Annotation Framework

The proposed event annotation framework can be conceptually described as a process of six modules, each of which addresses a certain aspect of the data retrieval, analysis and management needs of the problems specified in Definitions 1 and 2. Each of these modules is described below, while the framework is illustrated in Figure 1.

1. **Collection Creation:** The first step in the annotation process is the creation of a photo collection. CrEve offers support for collecting a portion of photos available through an online photo sharing service according to certain constraints, such as text, time, and place. To support place-centric content collection, the framework enables users to provide a bounding box for an area through a map-based interface (Figure 2). Apart from collecting external resources by means of the query facilities offered by this module, CrEve makes it possible to create a collection by importing an existing set of tagged photos in the local collection.
2. **Collection Indexing:** This module performs a set of offline indexing operations on the photo collection in order to enable a variety of browsing paradigms for discovering event-related content offered by the subsequent module. There are two indexing structures related to typical searches regarding text, geo and time information. In addition, there are two indexing schemes to support cluster-based photo retrieval: (a) a standard temporal clustering, and (b) the graph-based clustering scheme proposed in [20] incorporating visual and textual similarity between photos. Last but not least, there is a structure dedicated to the indexing of visual similarities between photos.
3. **Queries and Browsing:** As part of the collection exploration step, CrEve offers seven ways to search for photos that are similar with a selected input photo along certain dimensions. Search facilities include text, time, location and photo owner.

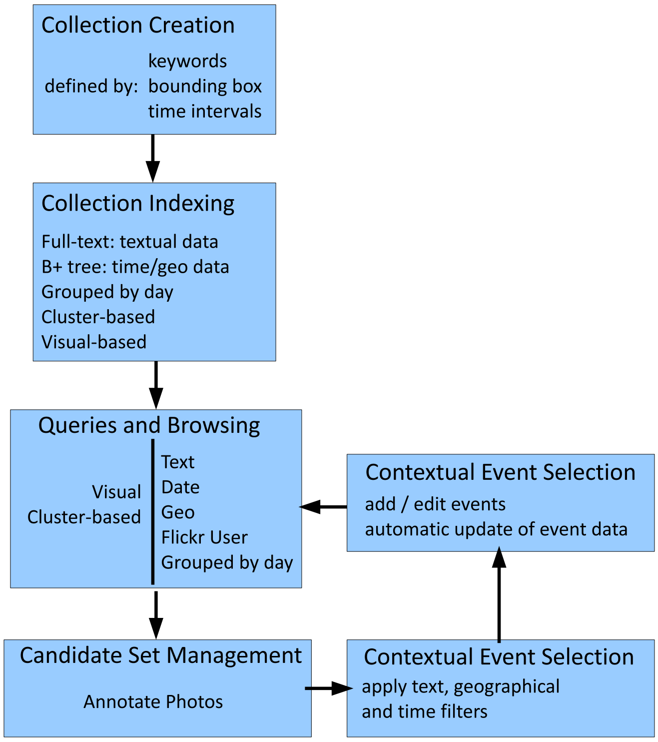


Fig. 1 Event Annotation Framework.

In addition, visual similarity and photo clusters derived either from temporal data or from image similarity graphs are supported.

4. **Candidate Set Management:** Both the automatically generated sets of photos and the results of queries on the collection of interest are characterized as candidate sets including photos related to some event. In this step, the user has to decide whether the photos of a candidate set depict an event or not. The scope of this component is to gather potentially related photos from an event together and give the opportunity of batch annotations to the users. Apart from that, this grouping also has a conceptual purpose due to the fact that gives the users some general indications of what photos to look for the event in question.
5. **Contextual Event Selection:** Once a user selects a set of photos to annotate, a photo-event matching process is carried out with the goal of providing a list of candidate events for annotating the photos. The event matching is based on the metadata of the selected photos (location, time, title).
6. **Event Management:** Annotators are able to create their own events providing the necessary metadata, such as title, description, location and time. To speed up the event creation process, the annotator does not need to enter all metadata (only the title is mandatory). Instead, the event metadata are automatically computed and progressively adjusted once one or more photos are associated with the event.

In addition, event hierarchies are supported by enabling users to link events to each other through a parent-child relation.

In the following paragraphs, each of the aforementioned steps is described in detail.

3.1 Collection Creation

Through a map-based interface (see Figure 2) a user can specify an area to download the photos that have been captured there. Besides this location-based filter, users may restrict their search in terms of time (i.e. begin and end date) and text (i.e. keywords). In the current version, CrEve establishes connections with the Flickr API in order to post the queries of the users. There are certain restrictions in both the size of the bounding box and the interval between begin and end date arising from a limitation in the maximum number of photos that can be returned from the photo sharing service. Another way of creating a collection is by importing a custom set of tagged photos such as a personal collection. The imported photos are presented to the user with the option of creating a collection (ℙ) populated with them. As for the structure of metadata from any external application, any differences or missing values will cause no problem due to the fact that CrEve can act even without metadata (i.e. visual similarity clustering is supported). CrEve will make use of whatever metadata are available.

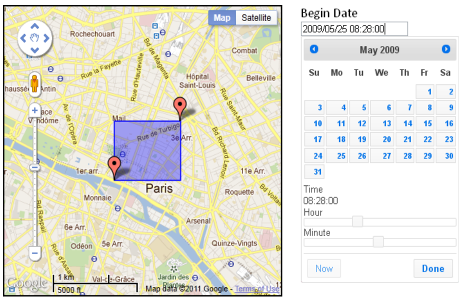


Fig. 2 Collection Creation.

3.2 Collection Indexing

Photos returned from the aforementioned retrieval process constitute a collection where five different ways of indexing are supported.

- **Full-text:** CrEve performs standard full-text indexing on the textual fields of the metadata that accompany the photos of the collection, i.e. titles, descriptions, tags and photo owners' usernames. Such indexing speeds up the retrieval of relevant content in text-based searches on these fields.
- **B+ tree:** For spatial and temporal data, B+ trees are used. This enables rapid response to interval-based (temporal) and bounding-box (spatial) queries.

– **Visual similarity index:** An additional index is built based on the visual characteristics of the photos. For each downloaded photo, CrEve computes the most similar photos in terms of visual content. To this end, SIFT descriptors [14] are extracted and a bag-of-visual-worlds feature vector is formed based on the software implementation of [37].
– **Temporal cluster index:** This type of indexing comes from a clustering technique taking into account only temporal data (i.e. $\forall t_p, p \in \mathbb{P}$). In the current implementation, temporal clusters are created by partitioning the photo collection by day. Apart from this simple grouping, there seems to be a potential in grouping photos according to certain parts of a day regarding discovering events or even arbitrary time intervals extracted by temporal clustering. However, our rationale was not to perform such a fine-grained time-based clustering on photos but to provide the users with some precomputed groups (i.e. grouping by day) in order to help them find event-related photos more easily. We believe that with this practice it is more likely to bring photos from the same events together and, at the same time, this approach was easy to implement. Nevertheless, an appropriate time-based clustering in different granularities would be useful for a future extension of our framework (e.g. by use of methods as the ones presented in [8]).
– **Hybrid cluster index:** Through the clustering framework presented in [20], CrEve provides another type of indexing by showing precomputed photo clusters regarding events. A hybrid similarity graph is built where the nodes represent the photos of the collection and the edges the scores of similarity between them. There are two types of similarity: visual and tag-based. For the visual similarity, the same process as in the case of the visual similarity index is used, while the tag co-occurrences between photos are used to derive the tag similarity. More specifically, each edge on the graph is weighted by the number of tags shared between the two photos. Very frequent tags are not taken into account in this process and very weak edges are discarded in order to increase noise resilience and to reduce the computational needs of the clustering algorithm. Subsequently, a community detection procedure is applied on the hybrid graph with the goal of identifying sets of nodes (i.e. photo clusters) that are more densely connected to each other than to the rest of the network. The resulting clusters are classified as landmarks or events [20]. CrEve makes use of the event clusters to build the Hybrid Cluster index.

3.3 Queries and Browsing

An effective way to manage a large collection of photos is to provide a rich set of query options that covers many different aspects (i.e. dimensions) of event-based media. Textual information of an event is supported by text search (dimension T), while geographic location (dimension G) and date/time information (dimension D) are supported by the respective indexing structures. In addition, the social aspect of events is supported by enabling search by photo owner ($u \in \mathbb{U}$, dimension U) given that that the same user is likely to have captured more than one photos of the event of interest. To increase usability of the aforementioned query capabilities, an auto-complete functionality is offered enabling CrEve users to click on specific buttons located under each photo (see Figure 3) and fill in the search fields on the page with information obtained from the photo. Along with the above types of queries, a user is provided with three other types of search. The visual similarity search (dimension V) and cluster-based

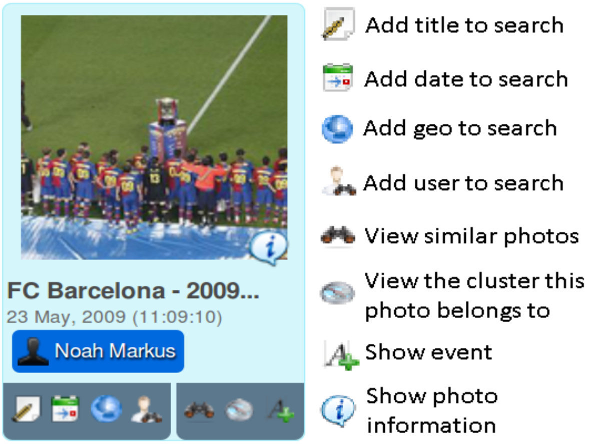


Fig. 3 Photo block facilities.

Table 1 Search Dimensions.

Dimension	**Description**
Text search (T)	Supports search on titles, descriptions and tags of photos
Date search (D)	Supports search between time intervals given by the user having as input the date when photo has been taken (t_p)
Geographical search (G)	Supports search inside a bounding box created by the geodata of a photo (l_p)
Photo Owner search (U)	Supports search by the username of the photo owner (u_p)
Search by day (DL)	Access to sets of photos grouped by day
Cluster-based search (CL)	Access to the cluster where a photo belongs to
Visual search (V)	Returns the m most similar photos based solely on visual information

search, including both the temporal (dimension DL) and hybrid similarity clusters (dimension CL). All these query facilities help event annotators gain insights into the photo collection by interactively presenting restricted views. Each search dimension acts as a means of targeted browsing through the photos of a collection in order to facilitate the process of discovering event-related content. Table 1 lists all search dimensions supported by CrEve.

3.4 Candidate Set Management

The scope of this component is to proceed to a grouping of some potentially related photos from an event. This turns to be useful in two ways. It gives the opportunity of batch annotations to the users, and further it gives some general indications of what photos to look for the event in question. Both precomputed photo clusters and the sets of photos that come from user queries are considered by CrEve to be candidate sets.

The association between photos and events is accessible through a candidate set page where users may decide if the photos belonging to the set are related to the event of interest. For better organization of the photos inside a candidate set, CrEve provides four ways of sorting: Users may sort photos by title, date, photo owner, and based on whether the photos have already been annotated. The first three sorting options speed up the process of quickly selecting related photos for batch annotation in case there is prior knowledge about the time, title or photo contributor of the event. The last option brings forward the photos without annotation, making the annotation process more efficient, since once proceeding with the annotation task, the annotator should not be obstructed by reviewing the already annotated photos again.

3.5 Contextual Event Selection

After completing several annotation tasks, a large number of available events are accumulated in the CrEve database. In case all of these events were presented to the annotator, the process of selecting the right event would become cumbersome. Therefore, a set of event matching functions are employed to recommend the events that are strongly related to the photos selected for annotation ($SP = \{p_1, p_2, ..., p_k\}, p_i \in \mathbb{P}$). Given an event E and the selected set of photos SP, each of the matching functions results in a score indicating the relevance of the event to the selected photos, $f : (SP, E) \rightarrow [0, 1]$. The product of the individual function scores is used to derive the ranking of matching events. The employed matching functions are the following:

- **Text-based matching:** In this type of matching a set of tokens is built by taking into account the tags of each selected photo ($p \in SP$) and the words included in their titles. From this set, the n most popular words ($n = 30$) in terms of frequency are selected and they are compared against those appearing in the tags ($\mathbb{X}_e$) or the title (T_e) of the event. The Jaccard coefficient between the two sets of tags is the result of this function.
- **Geo-based matching:** In case the selected photos contain geographical data ($\forall l_p, p \in SP$), their geographical median point is calculated. The geodesic distance D between the geographical median of SP and the event position ($l_e, e \in \mathbb{E}$) is computed. To normalize the matching score, the geodesic distance is transformed according to the formula: $d = 1 - \frac{D}{D_N}, D \leq D_N$, where D_N is the maximum allowed distance (e.g. $D_N = 500m$). For $D \geq D_N, d = 0$.
- **Time-based matching:** The time median is calculated from the dates of photos ($\forall t_p, p \in SP$), and the matching function results in full match ($f_T = 1$) for events with temporal overlap with the computed interval.

3.6 Event Management

CrEve enables users to add and edit their own events. Textual information such as title, description, tags/categories and information regarding location (i.e. place, venue, geo-coordinates) and time (i.e. begin and end date) are inserted in the system through a web form. Autocomplete functionalities are offered for several of the fields (place, venue) by making appropriate requests to third-party web services (Wikimapia, GeoNames). However, some of the fields may be unknown to users when they create the event.

To facilitate event metadata completion, CrEve offers automatic updating of event-related information based on new annotations. More specifically, when new photos are associated with an event, the event location (l_e), begin/end date ($t_{p,0}, t_{p,1}$), and tags ($\mathbb{X}_e$) are updated in case they have not been explicitly provided by the user. This mechanism greatly simplifies the task of creating new events.

CrEve also supports a hierarchical structure for events. Since an event can be a part of another event (e.g. live performance of a band is part of a music festival), CrEve enables the declaration of a parent-child relationship between two events. Each event can only have one parent but there is no limitation to the number of children. Thanks to this relationship, retrieving event-related content results in richer results since, apart from the direct association between the event of interest and content items, the employed retrieval model takes into account associations between content items and the parent/child/sibling events to the event. Note that the users themselves connect events together under the parent-child relationship. CrEve does not currently support any automatic way of inserting such relations.

4 User studies

We conduct three evaluation tasks in order to assess different aspects of the proposed annotation framework and to gain insights in the annotation process. Furthermore, the evaluation compares CrEve against standard annotation functionalities with respect to their impact on creating high-quality event annotations. In the first task, we evaluate the added value of using CrEve when creating event annotations for large-scale collections (Problem 1 in subsection 2.2). This task was conducted by a single expert annotator with access to the results of non-expert annotators produced in the other two tasks. The other two tasks were performed by 12 non-expert users after a brief presentation of the tool in order to evaluate its ease of use and effectiveness for finding photos related to specific events (Problem 2). In addition, these two tasks assess the extent to which different event dimensions contribute to the discovery of event-related content. In Figure 4 one may take a look at the three steps that describe the general picture of how this experiment was conducted with respect to the final outcome of the ground truth. In the first step (dotted box 1) the expert made use of the baseline tool in order to create P_{base} ground truth set while in step two (dotted box 2) the 12 participants made use of CrEve to complete the tasks given to them. These two steps are not connected whatsoever and the one does not depend on the results of the other. On the contrary, in the third step (dotted Box 3) the expert user proceeded to a verification of users' annotation in order to produce P_{CrEve} ground truth. Note that the expert user annotated the same collection of photos using first the baseline tool and then CrEve. This gave an advantage to CrEve due to the fact that the expert user spent some time to learn the dataset, so he could use this knowledge when he examined users' annotations coming from CrEve.

Dataset: In terms of the dataset, the evaluation was conducted on a collection of $36,675$ geotagged photos located within the metropolitan areas of Barcelona and Paris and captured in a single month (May 2009).

Users: The participants were able to access CrEve from any browser since it is a web-based tool. After a brief tutorial on how to use the interface they were free to choose

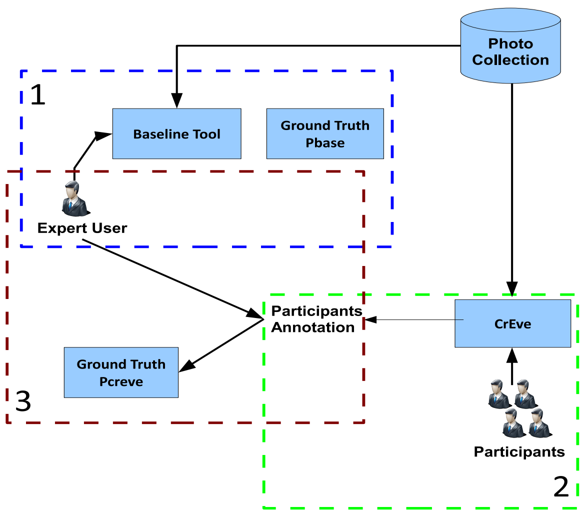
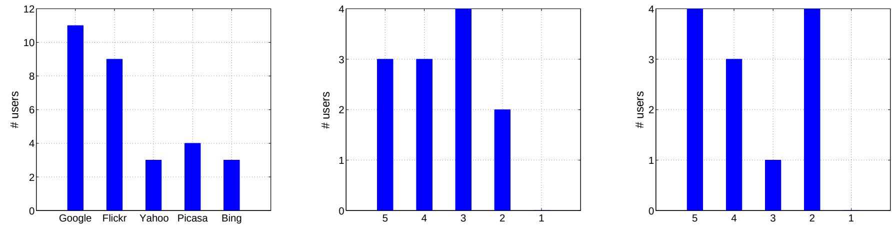


Fig. 4 Ground Truth generation process. Regarding the tasks of this study which are described later, Box 1 and 3 include Task 1 and Box 2 includes Tasks 2 and 3

performing the tasks either at home or in the lab. The 12 participants who know each other were kindly asked to work in isolation and definitely not to share any hints regarding the experiment. The participants were not paid for their effort and at the end of the experiment they had to answer to a number of questions giving some personal information and providing feedback about CrEve. As for the 'Personal Information' part of the questionnaire we gathered information about the gender of the participants (7 males and 5 females), their age (20-25: 2 users, 25-30: 2 users, 30-35: 6 users and 35-40: 2 users) and their educational level (Bachelor: 2, Master: 8, PhD: 2). Apart from the aforementioned demographics, we asked the users what image search tools they have used and the level of their familiarity with both tagging processes (e.g. adding descriptive keywords to photos in Flickr) and soccer. Such kind of information would be quite useful since the annotation tasks they had to accomplished were soccer-related. Figure 5(a) depicts a list of existing image search tools and the number of users that have used them while Figure 5(b) and Figure 5(c) show users' familiarity with tagging processes and soccer respectively.

Task 1: ***Event discovery and annotation***
In this task, an expert annotator is expected to discover all events of a specific category that appear in a given collection. For each discovered event, the annotator is expected to find all photos in the collection that are related to the particular event. In our study, the annotator was asked to find all events related to soccer. This task was performed using two different tools: a baseline annotation tool and CrEve. With the baseline annotation tool, an effort was made to simulate the capabilities of a standard annotation tool, i.e. enabling the exploration of the collection by means of browsing



(a) Image Search Tools (b) Tagging processes (c) Soccer

Fig. 5 Users' feedback regarding familiarity with existing image search tools, tagging processes and soccer. The 1-5 scale in figures (b) and (c) indicates the familiarity of users with the respective subject (1-completely unfamiliar, 5-very experienced)

through pre-calculated photo sets (e.g. all photos taking place at the date of the event, all photos including the term "soccer" in their tags, etc.). The baseline tool gave no possibility for interactive exploration of the collection. In addition, CrEve provided the expert annotator with precomputed photo sets populated with the annotations created by the 12 users of our study in Tasks 2 and 3 (for events in Table 2). In this way, the annotator could benefit from the collaborative nature of CrEve, since the annotated photo sets act as recommendations to the annotator. Note that apart from textual information, the baseline tool does not leverage location, time and photo owner metadata while CrEve takes all of them into consideration.

It is important to mention that the baseline tool was used only by the expert user and CrEve was used by both the 12 participants and him. The choice of using only one expert user lies in the fact that we had to tackle with a task that has not a great level of ambiguity to use more than one annotation experts. Given a sufficient amount of time we believe that with one fully committed expert a ground truth dataset pretty close to the ideal can be created. Ideally, in order to decrease the expert's load of work more than one expert users would be needed. However, it would be difficult to ensure such commitment from more people.

As for the user interfaces of both Creve and the baseline tool, Figures 6 and 7 contain two representative screenshots of them where the most important differences in terms of the UI are depicted. As opposed to CrEve, the baseline tool does not support search functionalities, photo block facilities and multiple photo selection for bulk annotation while it does not keep any information about both previous queries and annotations of users.

Task 2: *"Cold start" event detection*
The 12 users were given a title and a short description for three soccer events (Events 1-3 in Table 2 presented in random order). Having only this information, users were asked to make use of the CrEve functionalities (see Section 3.3) in order to discover the maximum number of relevant photos in the collection. No time limit was set for the task, but the time spent by each user on finding event-relevant photos was recorded.

Task 3: *Seed-based event detection*
In this task, users were provided with the title and short description of three other events (Events 4-6 in Table 2 presented in random order), but in addition they were also provided with two photos already associated with the event and were asked to find

all other photos related to the event of interest. Similar to Task 2, no time limit was set, but the time spent by each user on finding event-relevant photos was recorded. This task evaluates the value of collaborative annotation, since another user's annotations (i.e. seeds) seem to be indicators of useful information about the event, e.g. time, place, and visual characteristics of the event in question.

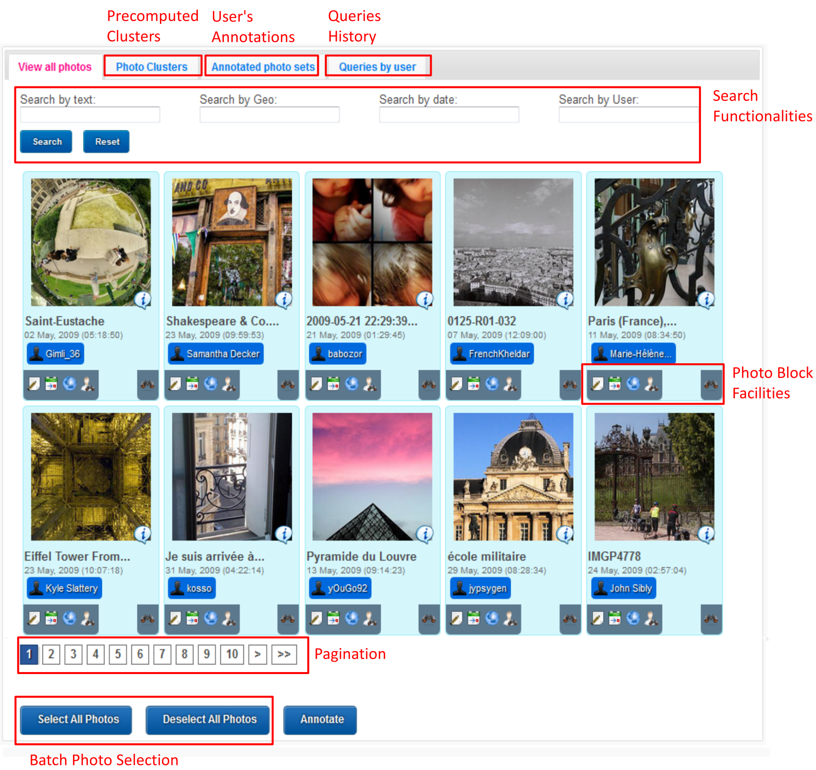


Fig. 6 CrEve User Interface

4.1 Evaluation Measures

In order to quantify the effectiveness of CrEve application we make use of the following measures:

1. **Ground Truth Improvement:** Given the ground truth sets generated by the baseline tool and CrEve, the measure of Ground Truth Improvement (GTI) is expressed by the number of additional photos included in the ground truth created by CrEve, both in absolute terms (Δ) and as a percentage (δ) of the total number of photos in the baseline ground truth.

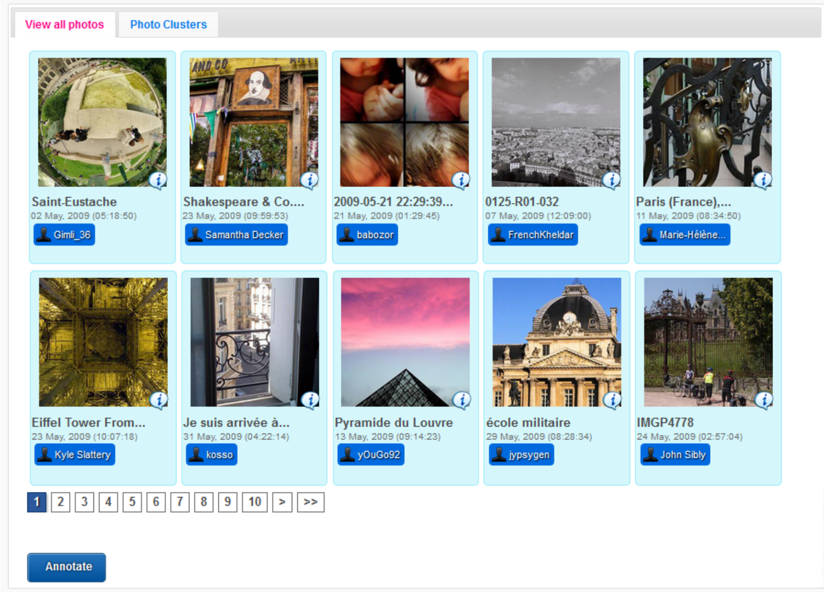


Fig. 7 Baseline Tool User Interface

Table 2 Events used in user studies.

Events	Title	Description	photos
Ev1	Celebration for Barcelona winning the title 2009	Barcelona won three titles (championship, cup, and Champions League) in the season 2008-2009. This event is about the celebration, not the winning matches.	304
Ev2	Barcelona - Osasuna	Soccer match between Barcelona and Osasuna	24
Ev3	Barcelona - Villareal	Soccer match between Barcelona and Villareal	19
Ev4	Pauleta Award	A match and ceremony for honoring Pauleta (football player)	200
Ev5	PSG - Monaco	Soccer match between PSG and Monaco	138
Ev6	PSG - Auxerre	Soccer match between PSG and Auxerre	99

2. **Precision - Recall:** Comparing the users' annotations with the ground truth generated in Task 1, it is possible to compute Precision (P), Recall (R) and F-measure (F) for each user.
3. **Normalized Mutual Information:** This is an alternative measure for quantifying the extent to which two groupings of objects match to each other. Considering the ground truth generated by the expert annotator as one grouping (P^a), and the annotation generated by a CrEve user as the second (P^b), the Normalized Mutual Information between them is computed as follows:

$$\text{NMI}(P^a, P^b) = \frac{-2 \cdot \sum_{i=1}^{K_a} \sum_{j=1}^{K_b} n_{ij}^{ab} log(\frac{n_{ij}^{ab} \cdot n}{n_i^a \cdot n_j^b})}{\sum_{i=1}^{K_a} n_i^a log(\frac{n_i^a}{n}) + \sum_{j=1}^{K_b} n_j^b log(\frac{n_j^b}{n})} \tag{1}$$

where $K_a = K_b$ is the number of groups (events) in the groupings, n is the total number of objects (photos) in the ground truth grouping, n_i^a, n_j^b are the number of photos contained in events i, j of groupings P^a, P^b respectively, and $n_{i,j}^{a,b}$ is the number of photos that are common between events i, j of groupings P^a, P^b.

4. **Overall time:** No time constraints were imposed on Tasks 2 and 3, but instead the time elapsed (T) between the start and the end of each event annotation process was recorded in order to quantify the effort spent in each annotation task. Due to the importance of time in the evaluation process, users were asked to be completely focused during the experiment.
5. **Contribution of search dimensions:** This measure quantifies the utility of the query facilities offered by CrEve (see Section 3.3) to the result of the annotation process. In particular, for each query facility we compute the percentage of correct annotations (P_C) that were created as a result of a query of this type. Since queries can be combined in composite queries, the contribution of search dimensions is quantified both in isolation and in combination with other dimensions.

4.2 Results

Ground truth improvement: Table 3 presents the GTI results for CrEve. In total, the ground truth created with the help of CrEve contains 147 more photos compared to the one created by the baseline tool, which corresponds to an 18.7% improvement on average. For six events, the sets of photos produced by the two tools were exactly the same. For four events CrEve led to the discovery of more photos associated with each one of the events (all photos found by the baseline tool were contained in the ground truth created with the help of CrEve). There were events, for which the improvement was dramatic, e.g. for the celebration of the Champions League win by Barcelona, 80 photos were found by CrEve compared to the 6 photos found with the help of the baseline tool. Another very important result is that the use of CrEve led to the discovery of three events that were not discovered at all with the use of the baseline tool. One may have noticed that most improvements occurred at those soccer events that include celebrations. We believe that this is due to the fact that these events cover a wider aspect of an event including many different scenes. Apart from the main soccer event there might be scenes with celebrating in the stadium, in the streets or at some local cafeterias. Identifying these different parts of data as content of the same event is not a trivial task and the appropriate use of metadata, especially location and date/time, can be used to find the right content and present it to the users. Most of the participants were able to find these different scenes and associate them with the right event. As a whole, CrEve was found to potentially improve the quality and coverage of event annotations created by experts.

Tool effectiveness and ease of use: Table 6 presents the results obtained from conducting the user study consisting of Tasks 2 and 3. On average, the participants'

Table 3 Comparison of ground truth created with the help of a baseline annotation tool and with the help of CrEve.

Title	P_{base}	P_{CrEve}	Δ	δ (%)
PSG vs Monaco (Ev5)	138	138	0	0
PSG vs Auxerre (Ev6)	99	99	0	0
Jubile Pauleta (Ev4)	200	200	0	0
PSG vs Rennes	49	49	0	0
Bayern United - Yahoo	0	4	4	NaN
Barcelona vs Villareal (Ev3)	9	19	10	111.1
Barcelona vs Osasuna (Ev2)	19	24	5	26
Celebration of Barcelona FC progressing to CL finals	1	1	0	0
Barcelona vs Atletic de Bilbao	1	3	2	200
La Liga Celebration (Ev1)	0	7	7	NaN
Celebration of CL win	6	80	74	1233.3
Triple Celebration for La Liga, Copa del Rey and CL	261	304	47	18
Sabadell vs Real Union	3	3	0	0
Espanyol vs Athletico Bilbao	0	2	2	NaN
Total	786	933	147	18.7

annotation performance with CrEve is translated to a 90% precision and 72% recall ($F = 80\%$). Given the fact that users had no familiarity with the collection at hand (i.e. some of them are not familiar with soccer) and no information for the target events, the annotation accuracy is considered satisfactory. When looking into more detail in the individual user behaviour, we found that two of them performed very poorly in the task (Users 4 and 11 achieved an F-measure of 66% and 63% respectively). We found that these two users had only basic understanding of the soccer domain (e.g. one user annotated the photos of an American football game as relevant to a soccer match). It is noteworthy that six of the users achieve very high precision ($\geq 95\%$) and other six of them achieve very satisfactory recall ($> 75\%$). In terms of effort, the users managed to complete the annotation of six events in a total of less than an hour. Given the size of the content collection ($> 35,000$ photos) and the fact that the users had no prior experience with the tool, this performance is very satisfactory. There are some examples of correct annotations from all the events included in our studies in Table 4. On the contrary, Table 5 illustrates some examples of erroneous annotations and provides possible explanations for their occurrence.

Figure 8 reveals additional insights with respect to the event annotation. According to Figure 8(a), there appears to be no connection between the effort spent on the annotation and the achieved annotation quality. For instance, the best annotated event (Ev4) was annotated in the least time. Also, it appears that when users are provided with positive examples of the target events, they achieve higher annotation quality with less effort: the events annotated in the context of Task 2 (blue squares, Figure 8(a)) have significantly lower F-measure than the events annotated within Task 3 (red dots) indicating that the existence of prior annotations can potentially help in the generation of high-quality ground truth. Moreover, the time required to annotate an event is not related with the number of photos associated with it as illustrated by Figure 8(b) for the observed range of values. One possible explanation for this surprising finding is that the batch annotation functionalities of CrEve in combination with its querying and sorting capabilities facilitated the annotation process thus making it independent of the number of photos to be annotated (i.e. the time spent on the annotation is mostly determined by the time needed to find the photos to be annotated). Once

Table 4 Examples of correct annotations of events described in Table 2.

Ev1 	Ev2 
Ev3 	Ev4 
Ev5 	Ev6 

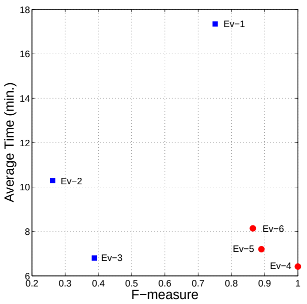
Table 5 Examples of incorrect annotations. In the first set of examples (Ev2) the user was misled by date information and annotated photos of an American football game as relevant to Barcelona - Osasuna soccer match. In the second set of examples (Ev5) users took only visual information into account and annotated photos of other soccer events as relevant to PSG - Monaco match.

Ev2

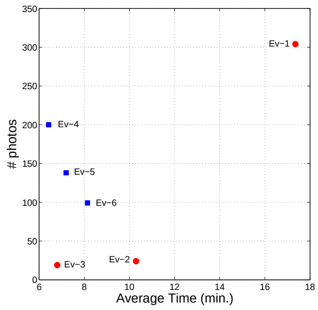
Ev5


Table 6 Evaluation results for Tasks 2 and 3.

Users	**IR performance**				**Effort** (mm:ss)					
	P	R	F	NMI	Ev1	Ev2	Ev3	Ev4	Ev5	Ev6
User-1	0.95	0.76	0.84	0.8	5:22	14:18	5:22	4:35	4:09	2:13
User-2	0.89	0.72	0.8	0.76	27:16	10:00	5:55	11:00	5:55	4:50
User-3	0.93	0.96	0.95	0.93	29:50	6:51	8:44	7:16	6:24	37:17
User-4	0.75	0.59	0.66	0.62	11:00	24:04	10:00	8:46	7:29	8:57
User-5	0.91	0.88	0.89	0.85	19:40	11:52	3:33	5:51	3:51	9:29
User-6	0.96	0.61	0.75	0.70	19:17	14:50	12:30	3:18	15:46	7:00
User-7	0.96	0.76	0.85	0.83	14:12	9:37	7:28	9:22	4:57	3:38
User-8	0.97	0.57	0.72	0.70	16:45	9:44	10:00	5:33	9:00	6:30
User-9	0.97	0.61	0.75	0.79	7:07	4:15	2:43	6:03	3:29	4:49
User-10	0.90	0.87	0.89	0.86	20:04	7:49	6:53	4:51	4:06	4:33
User-11	0.69	0.58	0.63	0.48	21:10	7:17	1:15	5:36	16:53	5:25
User-12	0.97	0.77	0.86	0.77	16:27	2:54	7:10	4:54	4:24	3:03
Average	0.90	0.72	0.80	0.75	17:20	10:17	6:47	6:26	7:12	8:08



(a) Scatter plot of T versus F.



(b) Scatter plot of T versus N.

Fig. 8 Event scatter plots.

more, it is noteworthy to point out that the above independence holds for the range of values taken by the event size variable (number of related photos) in the particular experiment.

Next, we examine the individual annotation performance of users as illustrated in Figure 9. At first look, there appears to be no correlation between effort spent and annotation quality achieved. However, if we consider the fact that the events of our study were football-related we have to take into consideration users' familiarity with football. Along with the latter, users' performance was certainly affected by any prior experience with searching/annotating tools or the lack of that. Having in mind these two factors we could say that there seems to be three clusters of users which depict the three following groups: (i) users that can be characterized as 'experts' due to the fact that spent little time with great performance, (ii) users that seem to be unfamiliar with both football and searching tools so they spent much more time than the others but with low performance, (iii) users where time and performance follow a linear relation.

Contribution of event dimensions: Figure 10 depicts the contribution of the different event dimensions, as expressed through the offered CrEve query functionalities, on the effectiveness of the annotation process. According to the Figure 10, the most useful query facility is the combination of date with location (DG), which ac-

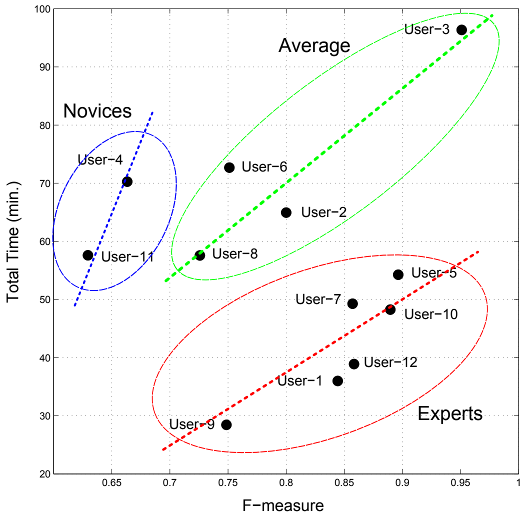


Fig. 9 Scatter plot of T versus F per user.

counts for 24.64% of the correct annotations. Apart from confirming the results of previous studies that consider time and location as the most fundamental dimensions of an event, this finding also points to the effectiveness of CrEve in exploiting these dimensions for helping users find event-related content.

The date-location dimension is followed in terms of contribution by the text query (T), which accounts for 24.18% of the correct annotations. A very valuable contribution is provided by the automatic clustering (CL) that is responsible for 16.55% of correct annotations. Other important query types are text and photo owner ($TU \rightarrow$ 9.05%), text, date, location and user ($TDGU \rightarrow$ 7.73%), text, date and location ($TDG \rightarrow$ 3.28%), date ($D \rightarrow$ 3.06%) and location ($G \rightarrow$ 2.9%). When we consider the use of an event dimension also in the context of composite queries, then the most important primitive dimension is text used for the generation of 47.8% of correct annotations, followed by location (43.1%), date (42.2%), and photo owner (21.3%). The high contribution of the photo owner (i.e. social) dimension demonstrates the added value of CrEve for the event annotation of content coming from social media.

Participants' feedback on CrEve: At the end of the experiment, each participant had to answer to a number of question about CrEve. We gathered positive answers (5 or 4 in most cases) in these two questions: (a) "Did CrEve help you find relevant photos (as specified by the tasks)?" and (b) "Were the tasks clear to you?". In addition we asked users whether CrEve was easy to use and the results were mostly positive (rate 5: 3 users, rate 4: 6 users and rate 3: 3 users). With the respect to Task 3 the users were asked if they found helpful the two photos already associated with the events. Their answers were positive again (rate 5: 7 users, rate 4: 4 users and rate 3: 1 user). Furthermore, we asked our users to check from a list what functionalities of CrEve they used in the experiment. Figure 11 depicts the list of available functionalities and the number of users that used them. Last but not least, some users provided us with additional comments on CrEve which can be considered as useful suggestions

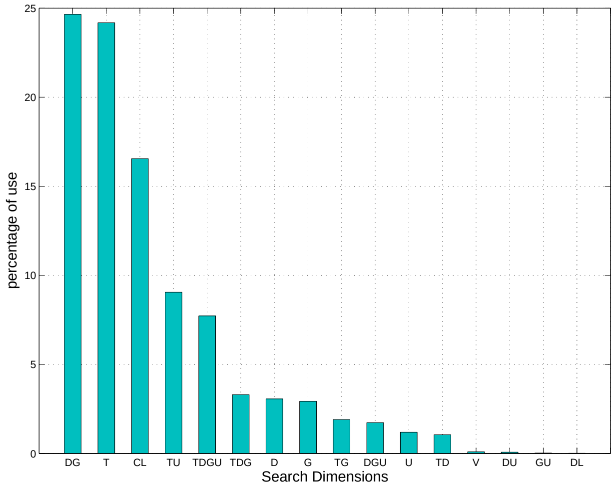


Fig. 10 Value of search dimensions in annotation process.

for future work. Some users suggest us adding new facilities in the interface (right click menu) whilst others recommend putting the photo block facilities (see Figure 3) in the Event View page as well.

5 Conclusions

The paper presented CrEve, a collaborative event annotation framework with the goal of facilitating event annotation in large photo collections. CrEve provides enhanced means for the creation and interactive exploration of the collection, including textual, temporal, spatial and user filters, as well as visual similarity and cluster-based search. CrEve speeds up event annotation by enabling bulk photo selection and automatic event suggestion based on the selected photos. Finally, CrEve reinforces the collaborative aspect of annotation by making users aware of the annotation of other users.

We conducted three evaluation tasks with the goal of assessing the ease of use and effectiveness of CrEve for event annotation, and of comparing the proposed framework with standard annotation capabilities with respect to their contribution in generating high-quality event annotations. Our main findings are the following: (a) the use of CrEve led to significant improvement ($> 18\%$) in the annotations generated by expert annotators compared to the use of a baseline annotation interface, (b) even with no experience in the use of the tool and no knowledge of the photo collection, CrEve users managed to achieve satisfactory annotation quality (both in terms of precision and recall), (c) users achieve better annotation quality and in less time when they are provided with positive examples of the target event, (d) there is no correlation between the number of photos related to an event and the time required to find them, (e) the more time users spend on event annotation, the better annotation quality they achieve, (f) text queries are the most widely used query facility, either on its own or

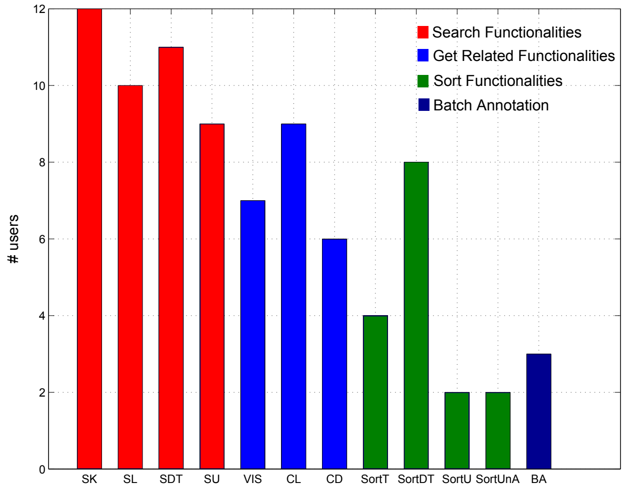


Fig. 11 Users' feedback regarding what functionalities of CrEve they used. Labels. (a) **Search Functionalities:** Search photos by keyword (SK), Search photos by location (SL), Search photos by date/time (SDT), Search photos by photo owner (SU). (b) **Get Related Functionalities:** Get visually similar photos (VIS), Get photos of the same cluster (CL), See photos captured in the same day (CD). (c) **Sort Functionalities:** Sorting photos by title (SortT), Sorting photos by date/time (SortDT), Sorting photos by username (SortU), Place un-annotated photos at the top (SortUnA). (d) **Batch annotation** (BA).

in the context of a composite query, (g) photo clustering is a promising mechanism for integrating in the event annotation process.

In the future, we plan to endow CrEve with additional features and evaluate its effectiveness in annotations problems of larger scales. An important facility is the possibility to export the generated event annotation as linked data, e.g. in the form specified by LODE[1] or Event-Model-F[2]. In addition, we would like to integrate external event sources, such as *last.fm*, *eventbrite* and *upcoming* to avoid the need for manual insertion of events by users.

We stated earlier that it is possible to apply the processes presented in this paper to videos as well by incorporating key frame selection techniques. Taking into consideration that video collections are usually richer than photo collections regarding events, it would be interesting to apply the proposed framework to videos. New insights regarding annotating events and possible drawbacks of our framework may be revealed and lead us to useful alterations and additions to CrEve. Finally, a longer term research goal would be to devise gamification mechanisms (e.g ESP game [38]) in order to incentivize users to spend more time on the annotation task at hand.

[1] LODE: An ontology for Linking Open Descriptions of Events `http://linkedevents.org/ontology/`

[2] Ontology Event Model F, Formal model of events `http://ontologydesignpatterns.org/wiki/Ontology:Event_Model_F`

Acknowledgements Christos Zigkolis's work has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. The work of Symeon Papadopoulos was supported by the GLOCAL and SocialSensor projects, partially funded by the European Commission, under contract numbers FP7-248984 and FP7-287975.

References

1. Ames, M., Naanman, M. (2007) Why We Tag: Motivations for Annotation in Mobile and Online Media. In Proceedings of the SIGCHI conference on Human factors in computing systems CHI '07, 971-980.
2. Appan, P., Sundaram, H. (2004) Networked multimedia event exploration. In Proceedings of the 12th annual ACM international conference on Multimedia (MULTIMEDIA '04). ACM, New York, NY, USA, 40-47
3. Appan, P., Shevade, B., Sundaram, H., Birchfield, D. (2005) Interfaces for networked media exploration and collaborative annotation. In Proceedings of the 10th international conference on Intelligent user interfaces (IUI '05). ACM, New York, NY, USA, 106-113
4. Begelman, G., Keller, P., Smadja, F. (2006) Automatic Tag Clustering: Improving search and exploration in the tag space. In Proc. of the Collaborative Web Tagging Workshop at WWW06
5. Burke, R. (2002) Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction 12, 4 (November 2002), 331-370.
6. Chandramouli, K., Izquierdo, E. (2010) Semantic structuring and retrieval of event chapters in social photo collections. In Proceedings of the international conference on Multimedia information retrieval (MIR '10). ACM, New York, NY, USA, 507-516
7. Chua, T.S., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.T. (2009) NUS-WIDE: A Real-World Web Image Database from National University of Singapore. In ACM International Conference on Image and Video Retrieval. Greece. Jul. 8-10, 2009
8. Cooper, M., Foote, J., Girgensohn, A., Wilcox, L. (2005) Temporal event clustering for digital photo collections. ACM Transactions on Multimedia Computing, Communications and Applications 1, 3 (August 2005), 269-288.
9. Elliott, B., Özsoyoglu, Z. M. (2008) Annotation suggestion and search for personal multimedia objects on the web. In Proceedings of the 2008 international conference on Content-based image and video retrieval (CIVR '08). ACM, New York, NY, USA, 75-84
10. Fialho, A., Troncy, R., Hardman, L., Saathoff, C., Scherp, A. (2010) What's on this evening? Designing User Support for Event-based Annotation and Exploration of Media. 1st International Workshop "EVENTS 2010 - Recognising and tracking events on the Web and in real life" at SETN 2010, Athens, Greece
11. Golder, S. and Huberman, A. (2006) The Structure of Collabo-rative Tagging Systems. Journal of Information Science.
12. Huiskes, M. J., Lew, M. S. (2008) The MIR Flickr Retrieval Evaluation. ACM International Conference on Multimedia Information Retrieval (MIR'08), Vancouver, Canada
13. Lim, J.-H., Tian, Q., Mulhem, P. (2003) Home photo content modeling for personalized event-based retrieval. IEEE Multimedia Magazine 10(4), 28-37
14. Lowe.,D.G. (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91110.
15. Liu, X., Troncy, R., Huet, B. (2011) Finding Media Illustrating Events. In ACM International Conference on Multimedia Retrieval (ICMR 2011), Trento, Italy
16. Marlow, C., Naaman, M., Boyd, D., Davis, M. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In Proceedings of the seventeenth conference on Hypertext and hypermedia (HYPERTEXT '06). ACM, New York, NY, USA, 31-40.
17. Mirkovic, M., Culibrk, D., Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., McArdle, G., Crnojevic, V. (2011) A comparative study of spatial, temporal and content-based patterns emerging in YouTube and Flickr, International Conference on Computational Aspects of Social Networks (CASoN), pp.189-194
18. Naaman, M., Harada, S., Wang, Q., Garcia-Molina, H., Paepcke, A. (2004) Context data in geo-referenced digital photo collections. In Proceedings of the 12th annual ACM international conference on Multimedia (MULTIMEDIA '04).

19. Papadopoulos, S., Vakali, A., Kompatsiaris, Y. (2010) Community Detection in Collaborative Tagging Systems. In Book Community-built Database: Research and Development, Springer
20. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A. (2011) Cluster-based Landmark and Event Detection on Tagged Photo Collections. IEEE Multimedia Magazine 18(1), 52-63
21. Papadopoulos, S., Zigkolis, C., Kapiris, S., Kompatsiaris, Y., Vakali, A. (2010) ClustTour: City Exploration by use of Hybrid Photo Clustering. In Technical Demonstration session of ACM Multimedia 2010 (Florence, 2010)
22. Papadopoulos, S., Troncy R., Mezaris, V., Huet B., Kompatsiaris, I. (2011) Social Event Detection at MediaEval 2011: Challenges, Dataset and Evaluation. In MediaEval Workshop, September 1-2, 2011, Pisa, Italy
23. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A. (2011) CERTH @ MediaEval 2011 SED Task In Proceedings of MediaEval Workshop, Pisa, Italy, 1-2 September 2011
24. Pu, P. Chen, L., Kumar, P. (2008) Evaluating product search and recommender systems for E-commerce environments. 8, 1-2 (June 2008), 1-27.
25. Tye Rattenbury, Nathaniel Good, and Mor Naaman. 2007. Towards automatic extraction of event and place semantics from flickr tags. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '07).
26. Quack, T., Leibe, B., Van Gool, L. (2008) World-scale mining of objects and events from community photo collections. In Proceedings of the 2008 international conference on Content-based image and video retrieval (CIVR '08), ACM, New York, NY, USA, 47-56
27. Sayyadi, H., Hurst, M., Maykov, A. (2009) Event Detection and Tracking in Social Streams. Proceedings of International AAAI Conference on Weblogs and Social Media, AAAI Press
28. Scherp, A., Saathoff, C., Franz, T., Staab, S. (2011) A Core Ontology on Events for the Semantic Representation of Human Experiences in the Real World. In Multimedia Tools and Applications, doi:10.1007/s11042-010-0667-z
29. Shaw, R., Troncy, R., Hardman, L. (2009) LODE: Linking Open Descriptions of Events. 4th Asian Semantic Web Conference (ASWC'2009). Shanghai, China
30. Shevade, B., Sundaram, H., Xie, L. (2007) Modeling personal and social network context for event annotation in images. In Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries (JCDL '07). ACM, New York, NY, USA, 127-134
31. Shneiderman, B., Bederson, B.B., Drucker, S.M. (2006) Find that photo!: interface strategies to annotate, browse, and share. Communications of the ACM 49, 4 (April 2006), 69-71
32. Sigurbjörnsson, B., Zwol, R. van, (2008). Flickr tag recommendation based on collective knowledge. In Proceedings of the 17th international conference on World Wide Web (WWW '08), 327-336.
33. Smeaton, A. F., Over, P., Kraaij, W. (2006) Evaluation campaigns and TRECVid. In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (Santa Barbara, California, USA, October 26 - 27, 2006). MIR '06. ACM Press, New York, NY, 321-330
34. Suh, B., Bederson, B.B. (2007) Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition. Interacting with Computers 19(4), pp. 524-544, Elsevier
35. Troncy, R., Fialho, A., Hardman, L., Saathoff, C. (2010) Experiencing events through user-generated media ISWC'10, International Workshop on Consuming Linked Data (COLD'10), CEUR Proceedings, Vol. 665, November 7-11, 2010, Shanghai, China
36. Troncy, R., Malocha, B., Fialho, A. (2010) Linking events with media I-SEMANTICS'10, 6th International Conference on Semantic Systems, colocated with "In the Open Track of the Linked Data Triplification Challenge" ,Graz, Austria, September 1-3, 2010
37. Van de Sande, K.E., Gevers, T., Snoek, C.G. (2009) Evaluating color descriptors for object and scene recognition. IEEE Trans. on Pattern Analysis and Mach. Intelligence, 99
38. Von Ahn, L., Dabbish, L. (2004) Labeling images with a computer game, In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '04).
39. T. Vander-Wal. Explaining and showing broad and narrow folksonomies, February 2007. http://vanderwal.net/folksonomy.html.
40. Westermann, U., Jain, R. (2007) Toward a Common Event Model for Multimedia Applications. IEEE Multimedia Magazine 14(1), 19-29
41. Wu, X., Lu, Y.-J., Peng, Q., Ngo, C.-W. (2011) Mining Event Structures from Web Videos. IEEE Multimedia Magazine 18(1), 38-51

42. Xie, L., Sundaram, H., Campbell, M. (2008) Event Mining in Multimedia Streams. Proceedings of the IEEE 96(4), 623-647
43. Zhao, M., Teo, Y.W., Liu, S., Chua, T.-S., Jain, R. (2006) Automatic Person Annotation of Family Photo Album. In CIVR 2006, LNCS 4071, pp. 163-172
44. Zhao, Q., Mitra, P., Chen, B. (2007) Temporal and information flow based event detection from social text streams. In Proceedings of the 22nd National Conference on Artificial intelligence (Vancouver, British Columbia, Canada, July 2007), AAAI Press, 1501-1506
45. Zsombori, V., Frantzis, M., Guimaraes, R.L., Ursu, M.F., Cesar, P., Kegel, I., Craigie, R., Bulterman, D.C.A. (2011) Automatic generation of video narratives from shared UGC, In Proceedings of the 22nd ACM conference on Hypertext and hypermedia, pp. 325-334