D4.2 - Module for cross-media linking of personal events to web content, v1

Project arconym:	ALIAS
Project name:	Adaptable Ambient Living Assistant
Strategic Objective:	ICT based solutions for Advancement of Social Interaction of Elderly People
Project number:	AAL-2009-2-049
Project Duration:	July, 1st 2010 – Juni, 30th 2013 (36months)
Coordinator:	Prof. Dr. Frank Wallhoff
Partners:	Technische Universität München Technische Universität Ilmenau Metralabs GmbH Cognesys GmbH EURECOM Guger Technologies Fraunhofer Gesellschaft pme Familenservice

D4.2

Version:	1.0
Date:	30-09-2011
Author:	X. Liu R. Troncy B. Huet

Dissemination status: PU

This project is co-funded by the Ambient Assisted Living (AAL) Joint programme, by the German BMBF, the French ANR, the Austrian BMVIT.

Once completed please e-mail to WP leader with a copy to
eric.bourguignon@tum.de and **frank@wallhoff.de.**

D4.2	Executive Summary
The aim of the ALIAS project is to build a system within a mobile robot to assist elderly people in their daily life. Besides the expected functionalities of the robot such as intelligent positioning, user interaction, day-to-day cognitive assistance, the entertainment part proposed and described in this documents will help users enjoy their retired life and feel closer to their family and friends. In details, we propose a framework to support users to choose their social events of interest based on the semantic web dataset, EventMedia (see D4.1). Events are a natural way for referring to any observable occurrence grouping persons, places, times and activities. They are also observable experiences that are often documented by people through different media (e.g. videos and photos). This intrinsic connection between media and experiences is explored in two aspects. Firstly, a method for finding medias hosted on Flickr that can be associated to a public event is presented. It will show the benefits of using linked data technologies for enriching semantically the descriptions of both events and media, so that people can search and browse through content using a familiar event perspective. Secondly, another method will be present-ed to automatically detect and identify events from social media sharing web sites. The approach is based on the observation that many photos and videos are taken and shared when events occur. The approach focuses on detecting events from the spatial and temporal labeled social media, and an algorithm is proposed to retrieve those media, perform event detection and identification, and finally enrich the detected events with visual summaries.	

Dissemination Level of this deliverable (*Source: Alias Technical Annex p20 & 22***)**	
PU	Public
Nature of this deliverable (*Source: Alias Technical Annex p20 & 22***)**	
P & R	Prototype and Report

Due date of deliverable	M12
Actual submission date	30/09/2011
Evidence of delivery	Web Service for Event Media Enrichment Operational

Authorisation			
No.	**Action**	**Company/Name**	**Date**
1	Prepared	EURECOM	

1. Introduction

The aim of the ALIAS project is to build a system within a mobile robot to assist elderly people in their daily life. Besides the expected functionalities of the robot such as intelligent positioning, user interaction, day-to-day cognitive assistance, the entertainment part proposed and described in this documents will help users enjoy their retired life and feel closer to their family and friends. In details, we propose a framework to support users to choose their social events of interest based on the semantic web dataset, EVENTMEDIA. Events are a natural way for referring to any observable occurrence grouping persons, places, times and activities. They are also observable experiences that are often documented by people through different media (e.g. videos and photos). This intrinsic connection between media and experiences is explored in two aspects. Firstly, a method for finding media hosted on Flickr that can be associated to a public event is presented. It will show the benefits of using linked data technologies for enriching semantically the descriptions of both events and media, so that people can search and browse through content using a familiar event perspective. Secondly, another method will be present to automatically detect and identify events from social media sharing web sites. The approach is based on the observation that many photos and videos are taken and shared when events occur. The approach focus on detecting events from the spatial and temporal labeled social media, and an algorithm is proposed to retrieve those media, perform event detection and identification, and finally enrich the detected events with visual summaries.

This document describes the proposed approaches to mine the relationship in the two aspects, which will support users browsing, querying the past social events, as well making their decision on upcoming events.

2. Related Work

Recent years are witnessing the success of content-sharing web site such as Wikipedia, Flickr, YouTube, Last.fm etc. Research related to and using social media has become a hot topic in multimedia research in the past years [1][2][3].

In recent years, research on how to better support the end user experience when searching and browsing multimedia content has drawn lots of attention in the research community. A tremendous amount of work has been done in very different areas. Among the possible directions, the usage of low-level visual features for improving content-based multimedia retrieval systems has made great progress in the past ten years [4]. The drawback of content-based retrieval systems is often the lack of manually labeled data for training systems. Our approach propagates the rich semantic description of events to the media, thus it contributes to semi automatically build reliable large training datasets. We exploit the rich "context" associated with social media content, including user-provided annotations (e.g., title, tags) and automatically generated information (e.g., content creation time) and we use linked data technologies for realizing large scale integration. A natural extension of our work would benefit from [5]. For ALIAS, a system is proposed to present the media content from live music events, assuming a series of concerts by the same artist such as a world tour. By synchronizing the music clips with audio fingerprint and other metadata, the system gives a novel interface to organize the user-contributed content. We did not yet consider audio fingerprint for tracking down series of events but rely only on semantic metadata so far.

Compared with traditional media data, social media is often accompanied with metadata such as tags, description and geographic location. The study of media content itself as well as the metadata are drawing much attention. In [6], Ames et al. studied the users behavior when they provide tags on Flickr. They found out that the intrinsic reason for users to tag their content is to make it better understood by others. However, user generated tags are often imprecise and incomplete or even irrelevant. Tag quality improvement is receiving important attention. In [7], Liu et al. proposed a tag ranking framework to automatically rank the tags associated with a given image according to their relevance to the image content. To generate knowledge from Flickr, the authors in [1] proposed a framework to fuse tag relevance with location information and visual cues, and aimed at mining the patterns from them. In [8], Cao et al. tried to leverage the contextual information of photos in social media to benefit the annotation process. They firstly mined the event relation from the spatial and temporal metadata of photos, and then used the hierarchical structure to improve the simple annotation. In [9], the authors proposed an image clustering framework to group the images based on geographical location and visual feature to depict different views of a location. In the work proposed here, we leverage on multiple metadata associated to user generated content to mine events.

While most work on event detection concerns the recognition and localization of special spatio-temporal patterns from video, a challenging topic for computer vision/video surveillance researchers, the identification of more general events such as vacation, concerts or conferences has received attention very recently. In [2], a framework to detect landmark and events to improve the user

browsing and retrieval is proposed. The work presented in [10] attempts to identify public events using both spatio-temporal context and photo content. Our work differs from the previously cited research in that we achieve event detection by mining social media sharing web sites using geographic and semantic properties simultaneously. Additionally, our event detection approach is not based on user query but rather on monitoring localized media uploading behavior of users.

3. LODE and EventMedia

Large numbers of web sites contain information about scheduled events, of which some may display media captured at these events. This information is, however, often incomplete and always limited on the sites. In previous work, we carried out user studies in order to collect end-user experiences, opinions and interests while discovering, attending and sharing events, and user insights about potential web-based technologies that support these activities. The results of this study support the development of an environment that merges event directories, social networks and media sharing platforms. We argue that linked data technologies are suitable for doing this integration at large scale given they naturally are based on URIs for identifying objects and a simple triple model (RDF) for representing semantic descriptions. We present the LODE event model [11] and explain how we populate this ontology by scraping three large event directories: last.fm, eventful and upcoming.

LODE ontology provides a minimal model that encapsulates the most useful properties for describing events. The goal of this ontology is to enable interoperable modeling of the "factual" aspects of events, where these can be characterized in terms of the "four Ws": "What" happened, "Where" did it happen, "When" did it happen, "Who" was involved. This ontology contains a formal set of axioms defining mapping between different metadata formats for multimedia.

Figure 1 depicts the metadata attached to the event identified by id = 350591 on last.fm. In details, it indicates that an event with type of "Concert" has been holden on the *13th of July 2007 at 20:30 PM* in the *Nouveau Casino theater* in Paris. The link between the media and the event is realized through the **lode:illustrate** property, while more information about the **sioc:UserAccount** can be attached to the URI that identifies this account. Hence, we see that the video hosted on YouTube has for **ma:creator** the user *aghorrorag*.

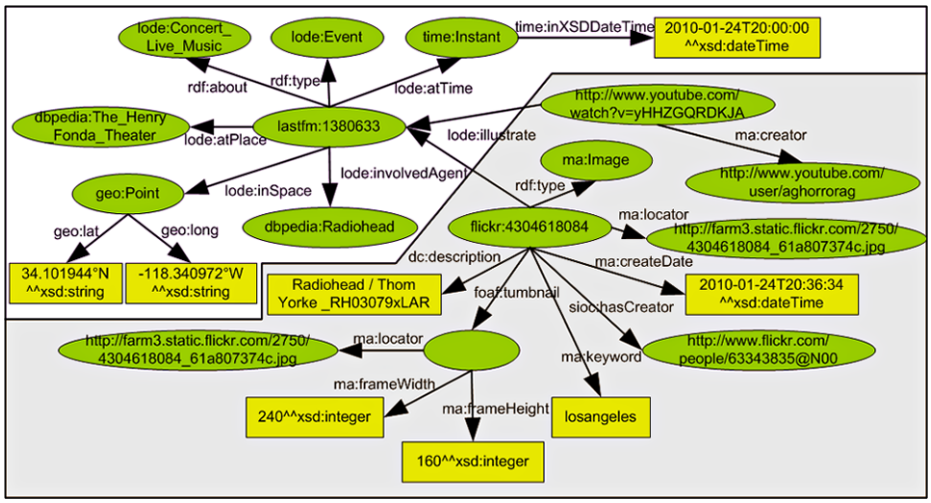

Figure 1: The Radiohead Haiti Relief Concert described with LODE (top) and illustrated with media described by the Media Ontology (bottom)

The EventMedia dataset is built on LODE ontology, which is composed of more than 1,7 million photos explicitly associated to more than 110,000 events. This dataset has been built from the

overlap in metadata between four popular web sites, namely Flickr as a hosting web site for photos and Last.fm, Eventful and Upcoming as a documentation of past and upcoming events. Explicit relationships between scheduled events and photos are looked up using special machine tags such as *lastfm:event=XXX* or *upcoming:event=XXX*. This dataset will also be used as the semantic knowledge source in the entertainment part in ALIAS project, and our proposed approach is performed on 110 events in EventMedia dataset with the details described in the following sections.

4. Find Media Illustrating Events

The set of photos and videos available on the web that can be explicitly associated to a Last.fm event using a machine tag is generally a tiny subset of all media that are actually relevant for this event. Our goal is to find as much as possible media resources that have NOT been tagged with a *lastfm:event=xxx* machine tag but that should still be associated to an event description. In the following, we investigate several approaches to find those photos and videos to which we can then propagate the rich semantic description of the event improving the recall accuracy of multimedia query for events[12].

Starting from an event description, three dimensions from the LODE model can easily be mapped to metadata available in Flickr and be used as search query in these two sharing platforms: the "what"-dimension that represents the title, the "where"-dimension that gives the geo-coordinates attached to a media, and the "when"-dimension that is matched with either the taken date or the upload date of a media. Querying Flickr or YouTube with just one of these dimensions bring far too many results: many events took place on the same date or at nearby locations and the title are often ambiguous. Consequently, we will query the media sharing sites using at least two dimensions. We also find that there are recurrent annual events with the same title and held in the same location, which makes the combination of "title" and "geo tag" inaccurate. In the following, we consider the two combinations "title"+ "time" and "geotag"+ "time" for performing search query and extending media that could be relevant for a given event.

4.1. Temporal Media Uploading Trend

We first investigate the time difference between the start time of an event and the upload time of Flickr photos attached to this event. For the 110 events in our dataset, we analyze the 4790 photos that are annotated with the Last.fm machine tag in order to compute the time delay between the event start time and the time at which the photos were captured according to the EXIF metadata. Figure 2 shows the result: the y-axis represents the number of photos uploaded on a day to day basis, while the x-axis represents the time (in days) after the event occurred.

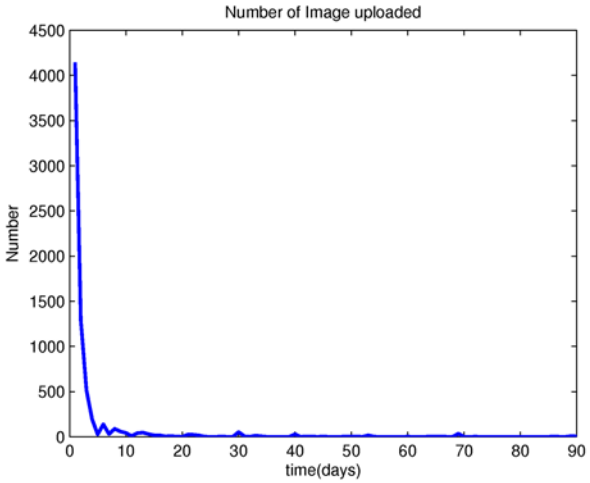

Figure 2: Image uploading tendency along time

The trend is clearly a long-tail curve where most of the photos taken at an event are uploaded during or right after the event took place and within the first 5 days. After ten days, only very few photos from the event are still being uploaded. In the following, we choose a threshold of 5 days when querying the photos using either the title or the geotag information.

4.2.Photo Query by GeoTag

Geotagging is the process of adding geographical identification metadata to media documents and is a form of geospatial metadata. These data usually consist of latitude and longitude coordinates, though they can also include altitude, bearing, distance, accuracy data, and place names. They are extremely valuable for application to structure the data according to location and for users to find a wide variety of location-specific information. Considering that a place is generally a venue, we assume that at any given place and time there is a single event taking place.

For all events of our dataset, we extract the latitude and longitude information from the LODE descriptions and we perform search query using the Flickr API applying a time filter of 5 days following each event date.

Figures 3 show the distribution of the number of retrieved photos for the 110 events in our dataset. We observe that the data is centralized in the left bins which means that for most of the events (n=95), the number of photos (resp. videos) retrieved with geotag is within the 0-100 range (resp. 0-20 range). The largest bin is composed of 45 events that have each between 1 and 50 photos retrieved.

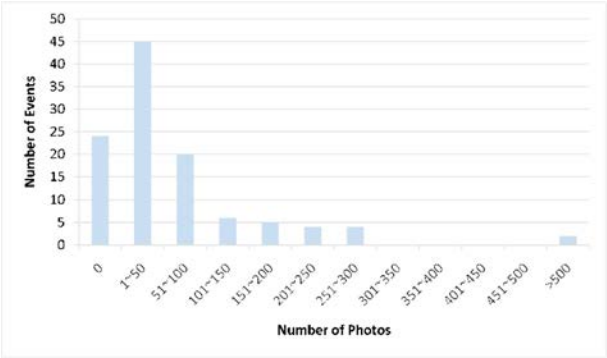

Figure 3: Number of photos per event in geotag based query

4.3.Photo Query by Title

The title is often the most useful information for describing the events. Similarly to geo-tagged queries, we perform full text search queries on Flickr based on the event title that is extracted from the LODE description. The photos retrieved are also filtered using a time interval of five days following the time of the event. When performing search query using the Flickr API query, we use the “text mode” rather than the ”tag mode” since the latter is missing in many photos. The number of photos retrieved at this stage is however in an order of magnitude greater than with geo-tagged queries. Due to the well-known polysemy problems of textual-based query, the title-based query brings a lot of irrelevant photos. We describe in the following section a heuristic for filtering out those irrelevant media.

The distribution of the number of retrieved photos and videos for the 110 events in our dataset is depicted in Figure 4. Generally, the results of query by title have a similar distribution like the results of query by geotag. For most of the events, a lower number of photos are obtained. Out of the 110 events under investigation, there are 80 events with less than 150 photos. However, for some events, a large number of media is retrieved: 12 events (resp. 15) with more than 500 photos.

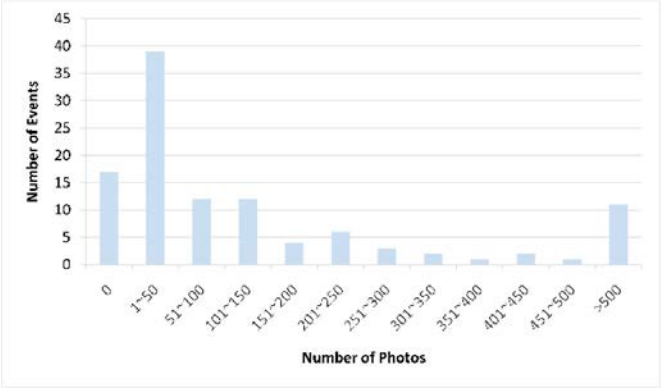

Figure 4: Number of photos per event in title based query

4.4.Pruning Irrelevant Media

Images and videos with specific machine tags such as lastfm:event=207358 can be unconditionally associated to events. We consider that media retrieved with geotag queries during a correct time frame should also be relevant for those events. The problem arises with the media retrieved by text-based queries (using the event title) where one can find many irrelevant media because of the polysemy problems.

In order to filter out this noise and to avoid propagating rich event descriptions to those media, we propose a method for pruning the set of candidate photos using visual analysis. The photos captured at a single event are already very diverse, depicting the artist, the scene, the audience or even the tickets. The diversity of the data makes it difficult to remove all the noisy images that should not be associated with the event considered, while keeping as much as possible the good ones. We address this issue in two steps to ensure high precision and recall ratio. First, we build a training dataset composed of the media containing either the event machine tag or a combination of geo-coordinates and time frame corresponding to the event. The photos resulting from query by title compose the testing dataset. The visual features employed are 225D color moments in Lab space, 64D Gabor texture, and 73D Edge histogram. For each image in the training data, the nearest neighbors using the L1 distance measure in the training set are found and the smallest distance taken as threshold. Second, images originating from the title query are compared with training images. Images for which the distance to images in the test set is below the threshold are candidates for illustrating the event. The algorithm can be formalized as followed:

Table 1: Prune function

```
1: INPUT: TrainingSet, TestingSet
2: OUTPUT: PrunedSet
3: for each img in TrainingSet do
4:         D = [ ]
5:         for each imgj in TrainingSet− {img} do
6:                 D.append ( dist L1 ( img, imgj ) )
7:         end for
8:         Threshold = min ( D )
9:         for each imgt in TestingSet do
10:                if dist L1 ( imgt, img ) < Threshold then
11:                        PrunedSet.append ( imgt )
12:                end if
13:        end for
14: end for
15: return PrunedSet
```

We adopted an adaptive threshold because of the visual diversity within the training dataset. Even for the images belonging to the same event, the concept can vary from the musicians, singer to venue, or event ticket. In order to remove noisy images in the testing data, the threshold should be adjusted respectively. Figure 5 shows the values of the threshold used in the experiments which range from 0.01 to 0.346.

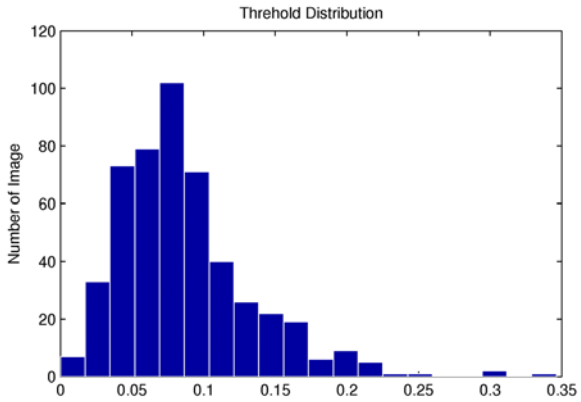

Figure 5: The distribution of threshold in the pruning process

For evaluating our pruning algorithm, we take the top 20 events from our 110 events dataset. For these 20 events, there are 785 images in the training set (photos containing either an event machine tag or a geotag) and 1766 photos in the testing set (photos retrieved by event title). We build manually the ground truth for those 1766 photos selecting which ones should be attached to an event and which ones should not (Table 2). The 20 events were all concert events and photos are often depicting artists, venues, stages or audience. Some photos were, however, sometimes hard to judge but the manual assessor used all metadata available around each photo such as the entire list of tags or the albums in which the photos were gathered to decide whether the photo should be discarded or not. In the end, we manually remove 193 irrelevant images by their visual appearance and metadata. The remaining 1593 images are used as ground truth dataset.

The results of the pruning algorithm detailed in the Section 4.4 applied to the 1766 photos are shown in Table 2. The threshold used is quite strong in order to guarantee a precision of 1 for most of the events. However, this causes about 80% of the candidate images to be excluded, including many relevant photos.

Table 2: Number of photos for 20 events, results of the pruning algorithm and the simple heuristic extension

ID	DataSet (nb of photos)			Pruning Result			Extended Heuristic	
	TrainingData	TestingData	GroundTruth	Pruned	Precision	Recall	Extend	NewRecall
346054	2	24	2	1	1	0.5	1	0.5
158744	3	48	48	23	1	0.479	44	0.917
371981	4	16	6	4	1	0.667	4	0.667
341832	7	0	0	0	1	1	0	1
362195	7	0	0	0	1	1	0	1
235445	10	1	1	0	1	0	0	0
42644	13	85	81	13	1	0.16	13	0.16
165697	23	1	1	0	1	0	1	1
137530	24	9	4	0	1	0	1	0.25
517159	24	0	0	0	1	1	0	1
222241	36	204	180	33	0.97	0.183	72	0.4
234649	45	35	4	1	1	0.25	1	0.25
207358	54	68	4	4	1	1	4	1
429517	60	171	169	27	1	0.16	41	0.243

437747	65	144	142	8	1	0.056	13	0.092
117886	68	99	97	4	1	0.041	11	0.113
150390	71	16	16	1	1	0.063	1	0.063
350591	79	85	85	6	1	0.071	66	0.776
472733	93	500	478	8	1	0.017	18	0.038
176257	97	260	255	47	1	0.184	147	0.576

4.5. Human computer Interface

Based on our proposed approach, a vivid web service interface is developed for the user to query and browse the media data. On a given event with id=XXX, the interface can finish all of the query and process. At first it will do the query from EventMedia dataset to obtain the event information, such as the taken place, time, title, etc. And then it will search the photos from Flickr automatically with query parameters of machine tags, title+time, as well as geotag+time. After the visual pruning process, the final results will be displayed in the web browser. Figure 6 gives an example of the final illustration results on event lastfm:event=915311, which is the "**Peter Fox feat. Cold Steel**" holden in Berlin on 12 June 2009. And the inferface shows all of the queried photos that are related to this event.

Now the web service is still under construction, and a more vivid and friendly interface will be delivered in the following version.

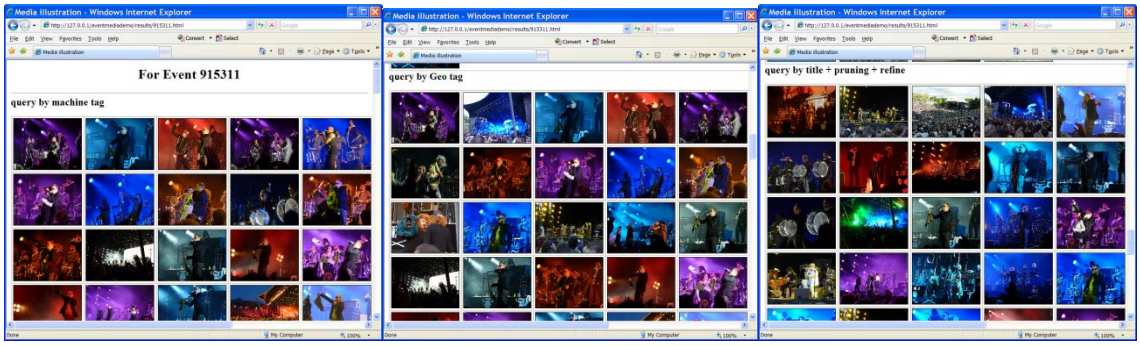

Figure 6. Events illustration from lastfm Event :915311

5. Event Detection on social photos

In this section, we address the problem of structuring social media activity into events and in particular how to automatically detect those events and their properties (location, time and participation). Event directories such as Last.fm, Eventful and Upcoming publish information about scheduled events in order to help users in tagging and structuring multimedia material and ease their future retrieval. While such services are becoming increasingly popular they are often incomplete, sometimes inaccurate in terms of the information they provide and always work as silos that lock information into the sites. On the other hand, they largely overlap in terms of coverage, but they generally fail to give a good feeling of the atmosphere of an event, while this feature is considered as of primary importance to support users in deciding to attend or not an upcoming event. The problem we tackle is therefore how can we make use of metadata attached to media and events (tags, description, geographic location) to create the missing link between an event description and its illustrating media documents.

5.1.Event Detection based on social media distribution

We are interested in detecting events by monitoring the social media sharing activity at specific locations. Our approach is based on the observation that many media documents are taken and shared when events occur.

We focus on detecting events from the spatial and temporal labeled social media, and propose an algorithm to perform event detection and identification. Media sharing websites are regularly providing novel means for users to semantically enrich their photo and video collections. Geo-tagging adds location information to media in the form of latitude and longitude coordinates. Such information is either captured by the device itself (when featuring GPS functionalities) or through user input (via textual input or location identification on a digital map). It is expected that when an event takes place, there will be many persons taking picture or videos and later uploading them on sharing platforms. As shown in Figure 7, an event is detected when the number of taken photos reaches maximum values.

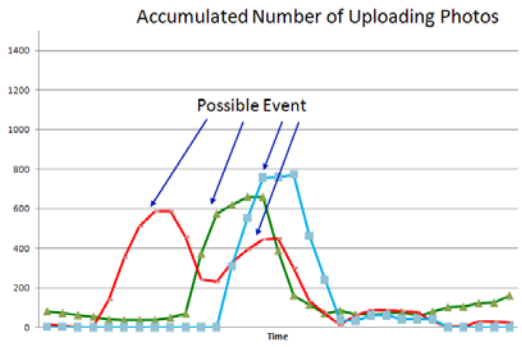

Figure 7: Event Detection

5.2.Venue Bounding-Box Location Estimation

Our approach consists in measuring upload peaks at known venues in order to detect an event occurrence. It is therefore necessary to model the venue location. The Flickr API allows to query photos based on their geographical location. Given region parameters, in the form of center and radius, or rectangle bounding box, the photos taken within a specified location can be retrieved. However, it is not so easy to obtain the geographical covering area for a place, since there are no public data for the size of a venue. We address this issue by leveraging on the event model provided by Last.fm and used by Flickr users. Using the Last.fm API, events (EventID = E) which took place at a given venue (VenueID = V) are retrieved. Their machine tags "lastfm:event=E" are then used to search for geotagged media on Flickr. A bounding box is then computed using the GPS coordinates of the retrieved photos. The algorithm 1 details the processing steps leading to the venue's location estimation. The final bounding box is estimated as the minimized rectangle of the GPS coordinates after re- moving the outliers (photos which are located further than twice the variance of the set in either direction (longitude or latitude)).

Table 3. Bounding-box Estimation

```
1: INPUT: VenueName
2: OUTPUT: BoundingBox
3: PhotoSet = [ ]
4: EventSet=GetPastEvent ( VenueName )
5: for each eventid in EventSet do
6:         photos = GetFlickrPhotos ( eventid, hasGeo =
True )
7:         PhotoSet.append ( photos )
8: end for
9: GeoSet = GetGeoInfo ( PhotoSet )
10: GeoSet.filter ( )
11: return MinRect ( GeoSet )
```

5.3.Analyzing the Flickr Activity around Venues

We aim at mining events automatically based on photo upload activity at particular locations. We are interested in detecting events by monitoring the social media sharing activity at specific locations. Our objective in terms of event detection is to identify the date and title of the event given its venue or location. Our approach consists of carefully selecting the dates with high number of uploads and consider those candidate dates as events. More formally, let us consider the time series $\{d_i, i \in [1,T]\}$ that represents the temporal evolution of the photo upload characteristic at a given venue *v*. The event *e* starting at time *t* is detected when the photo upload characteristic is greater than a given threshold *THD*.

$$e_t = arg_i(t_i > THD)$$

5.4.Detection Experiments and Results

The proposed approach has been validated on 9 venues in EventMedia dataset. The EventMedia dataset has knowledge of more than 13,000 different venues for which at least one description of event explicitly associated to at least one photo is available. From this very large dataset, we selected 9 venues that proved to have a significant activity the last three years. Table 4 shows the number of events, photos and distinct users for those venues during this period according to the EventMedia dataset. The ranking value corresponds to the popularity of the venue in the entire dataset when the sorting criterion is the number of distinct users that have uploaded at least one photo on Flickr taken during an event hosted at those venues.

Table 4: Number of events, photos and distinct users for 9 venues

Venue	NbEvents	NbPhotos	NbUsers	rank
Melkweg	352	6912	266	1
Koko	151	3546	155	5
HMV Forum	106	2650	130	8
111 Minna Gallery	24	1369	105	14
Hammersmith Apollo	79	2124	96	20
Circolo Magnolia	79	2190	76	40
Rotown	204	3623	49	48
Ancienne Belgique	212	7831	56	83
Circolo degli Artisti	784	2590	86	43

We define the bounding boxes for the 9 venues selected and we crawl all photos taken at those locations during a one month period from May 1st 2010 to May 31st 2010. Hence, a collection of 4604 geo-tagged photos has been obtained. The number of geo-tagged photos hosted on Flickr represents still a tiny fraction of all photos shared on this web site. Many other relevant photos will not be retrieved using a pure location-based search. In order to obtain more relevant photos, we use the venue name as the keywords and query photos without geo-tags during the same period. We extend the dataset with another 4589 relevant photos. Figure 8 depicts the Flickr upload activity during May 2010 for two of the nine venues selected. Looking at both curves, one can see that the number of pictures taken and uploaded varies temporally over the month. The 9178 photos gathered by crawling Flickr during May 2010 using either the geographical bounding boxes or the venues name will be the basis for the event detection experiments.

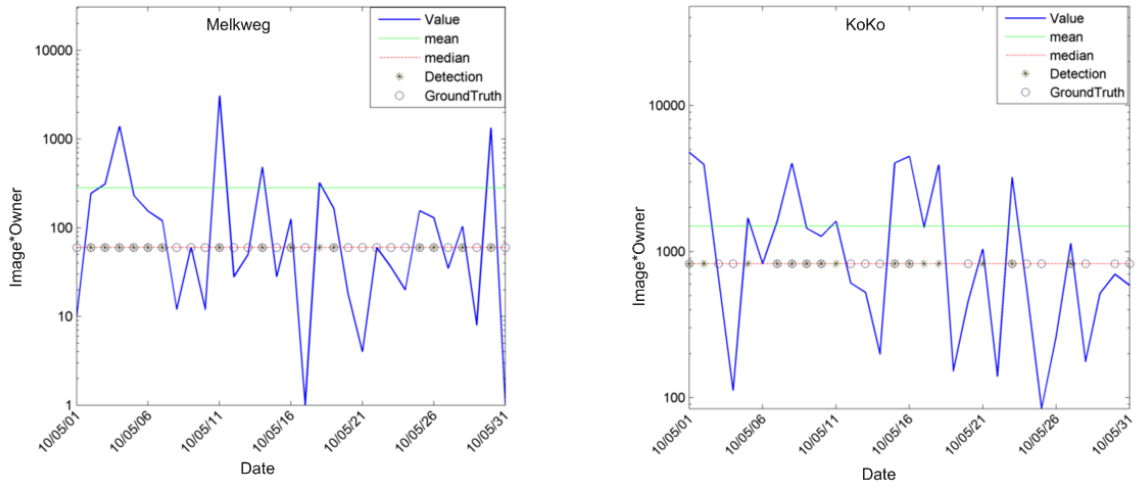

Figure 8: Event detection during May 2010 in the Melkweg (Amsterdam, NL) and Koko (London, UK) venues

The detection method described is run on the nine selected venues. The Figure 8 shows results for 2 of these venues: Koko and Melkweg respectively. In those figures, the main blue curve shows the number of photos multiplied by the number of photo uploaders per day, the green curve corresponds to the mean while the red one is the median of the blue curve, the stars is the date for which events are detected (based on median thresholding). The ground truth is shown using circles over the median. One can see that many events would be missed using the mean threshold while the median is able to capture many more events while keeping the false detection rate low.

In May 2010, a total of 242 events are reported from the official schedules of the nine selected venues. In order to evaluate the performance and accuracy of our proposed approach, we need to align the detected events with official events. Once event start times are identified, we use the tags of the corresponding photos to deduce the topic of each event. All words from tags and titles of photos taken on that day are parsed and sorted by their occurrences. The top 15 keywords are kept to infer the topic of individual events. Detected events are manually matched with ground truth events based on date and title. Matching events are those sharing a common starting date and for which at least one non stop-word can be found in both the top 15 keywords list of detected events and the title of the ground truth events.

The details of the detection results for each venue are shown on Table 5. In this table, the Recall column requires further discussion. Since the detection of events is based on photo upload characteristics, any events for which no or too few photos have been uploaded on Flickr cannot be detected. The recall values reported are in fact currently floored and should be understood as such. As the current uploading trend continues to evolve, we expect the number of “detectable" events using our proposed approach to grow accordingly.

Table 5: Event Detection Results

Venues	GroundTruth	Prediction	Matched	Precision	Recall
Melkweg	69	15	12	0.800	0.174
Koko	20	15	8	0.533	0.400
HMV Forum	14	12	9	0.750	0.643

111 Minna Gallery	23	15	2	0.133	0.087
Ancienne Belgique	38	15	9	0.600	0.237
Rotown	16	15	8	0.533	0.500
Circolo degli Artisti	22	15	8	0.533	0.364
Circolo Magnolia	25	3	1	0.333	0.040
Hammersmith Apollo	15	15	10	0.667	0.667
In total	**242**	**120**	**67**	**0.558**	**0.277**

To provide the vivid interface, a visual cluster is generated for each event [13]. From the final photo set identified in the previous section, a visual cluster is created to show a vivid interface for users. Figure 9 shows the visual cluster result for a detected event. During the enrichment phase, we expect to bring more diverse photos into the collection. For example, the Figure 9 depicts the event 2, which is held in Hammersmith Apollo with the title iggy stooges. Figure 9(A) is generated from the relevant photos from the detection set that corresponds to photos that are either geo-tagged or tagged with a venue name. Figure 9(B) shows the collection of images resulting from our enriching and visual pruning method. We can clearly see the increased visual diversity of the scenes between the two sets. The final set of images illustrating the iggy stooges event will be composed of both sets.

(A) Before enrichment (B) After enrichment & Pruning

Figure 9: Visual cluster for an event , which was held on 03/05/2010, in the venue Hammersmith Apollo with the title iggy stooges

6. EURECOM at MediaEval'11

Here, we present our approach and results of the MediaEval 2011 social event detection (SED) task. We solve the event detection problem in three steps. First, we query all event instances that happened given some condition. Then, an event identification model is proposed to measure the relationship between events and photos. Finally, visual pruning and owner refining heuristics are employed to improve the results.

6.1. Approach Description

The challenge of the social event detection task is to find the photo clusters that are relevant to events held on a given location during a particular period of time. We tackle this problem in two steps: first, we attempt to retrieve all of the events that occurred at a given place and time; second, we use the extracted information about these events and attempt to match them to the photos metadata in the dataset. All of the photos that are matched to the same event can be grouped in one cluster. Besides these two main steps, we also improve the detection results with visual feature and ``owner'' metadata.

6.1.1. Prior knowledge acquisition

We know that it is easier and more accurate for the computer to identify specific pattern compared with abstract concept. To find concert or soccer events that may be hidden in the dataset, we first look for all instances of these two kinds of events held in a given place and time.

Soccer games and concerts are types of favorite activities in people's daily life and one can find substantial information online about such scheduled events. For example, BLeague (http://www.fbleague.com) provides the official football games that registered in FIFA (http://www.www.fifa.com) and UEFA (http://www.www.uefa.com). From this web site, we obtained 461 football games that occurred in May 2009, among which 6 took place in Roma and Barcelona. These 6 soccer events are our prior knowledge for the challenge 1.

For challenge 2, we extract concerts information from event directories such as Last.fm (http://www.last.fm), Eventful (http://www.eventful.com), and Upcoming (http://upcoming.yahoo.com). After manual check, only Last.fm contains descriptions of events held on the given conditions. Last.fm is a popular music web site that records concert events held in more than 190 countries. In addition, Last.fm provides an API for the developer to build their algorithm based on its data. Using its public API, we found 68 events that took place in the Paradiso and 3 events in Parc del Forum in May 2009.

6.1.2. Event Identification Model

With the prior knowledge of scheduled events description, the event detection task changes to a matching problem where a model can be used to measure the relationship between events and

photos. Here, we consider events as something happening in some place during some time. Therefore, the title, time and location are three key factors that identify an event. The corresponding photo metadata are text description, taken time and place. Since these three factors are independent, we can measure the probability of a given photo P to be relevant to an event E by

$$p(P|E) = p(P.text|E.title)p(P.time|E.time)p(P.geo|E.geo) \quad (1)$$

where the first item measures the similarity of a photo text description with an event title. Since both of them are short and sparse, the most straightforward way to measure them is:

$$p(Text1|Text2) = \frac{|Text1 \cap Text2|}{|Text2|} \quad (2)$$

where the function $|.|$ is the total number of words in a text vector.

The second item in Equation 1 measures the difference between photos taken time and event held time. Here, we measure the difference using the Dirac function.

$$p(Time1|Time2) = \delta\left(\frac{date(Time2 - Time1)}{N}\right) \quad (3)$$

Where the function $|.|$ calculates the number of days for a time span, δ is the Dirac delta function that takes the value 1 when and only when the input parameter is zero, and N is used for scaling (its value will be discussed in the Result section).

The third item in Equation 1 measures the distance between photo geo tags and event locations. The best distance measure to use seems the L2 distance between the two locations. However, an important amount of photos do not have geo tags and when provided, GPS data in the Flickr dataset can be inaccurate. Consequently we just use the city/venue name to measure the location feature and we use the textual metric formalized in the Equation 2.

This method finds many photos with a clear description and association to events. However, text-based matching brings also noise and it cannot deal with photos without any text description. We employ visual features to remove the noisy photos and ``owner'' metadata to find out relevant photos without text description.

6.1.3. Visual Pruning

Visual pruning is employed to remove the noisy photos from the results of the event identification model [12]. We assume that the photos that are corresponding to the same event should be similar visually. The method used here is quite straightforward. Given a set of the photo features $\{f_i, i \epsilon [1, N]\}$, the distance between each feature f_i and its mean vector m is measure by the L1 distance.

$$d_i = sum(|f_i - m|) \quad (4)$$

Photos are then sorted according to the distance d_i. The bigger the distance and the less similar the photo is with the photo cluster, so we prune the photos with such a large distance. Experimentally, we remove the 5% photos that are far from the center in the visual feature space.

6.1.4. Owner Refinement

Owner refinement is another way to improve the detection results [12]. We assume that a person cannot attend more than one event simultaneously. Therefore, all the photos that have been taken by the same owner during the event duration should be assigned to the same cluster. Using this heuristic, it is possible to retrieve photos which do not have any textual description.

6.2. Experiments and Results

Based on the proposed approach and the event instances obtained previously, we design our runs as follows:

Challenge 1:

run1: The parameter N in Equation 3 is set to 3, and the basic **Event Identification Model** is run.

run2: **Owner Refinement** is performed on the results of **run1**.

Challenge 2:

run1: The parameter N in Equation 3 is set to 1, and the basic **Event Identification Model** is run.

run2: **Owner Refinement** is performed on the results of **run1**.

run3: The parameter N in Equation 3 is set to 3 to reduce the impact from erroneous taken time, and the basic **Event Identification Model** is run.

run4: **Owner Refinement** is performed on the results of **run3**.

run5: Visual Pruning and **Owner Refinement** are performed on the results of **run3**.

A summary of the results is shown in Table 6.

Table 6: Event Detection Results

Challenge	Run	Results		Evaluation			
		Events	Photos	P(%)	R(%)	F(%)	NMI
1	run1	2	216	97,69	41,21	57,97	0,242
	run2	2	222	97,75	42,38	59,13	0,2472
2	run1	18	1133	70,79	48,9	57,84	0,4516
	run2	18	1172	71,13	50,49	59,06	0,4697
	run3	24	1502	70,51	64,57	67,41	0,5987
	run4	24	1556	70,99	67,01	68,95	0,6171
	run5	24	1546	71	66,59	68,72	0,6139

As shown in the Table 6, two events are found for challenge 1 with 216 photos identified by the event identification model. Six additional photos are found by the ``Owner Refinement'' approach.

For the challenge 2, there are mainly two groups of runs. The first group (run1, run2) used the parameter N=1, and 18 events are found from the 69 events set previous obtained. In the second group (run3, run4, run5), 24 events are found with the parameter N=3. In general, the results for the challenge 1 are just average since only 6 football games were found as prior knowledge and we suppose that several other games have been missed. For challenge 2, the results are more promising and competitive.

7. Summary and Conclusion

This document reports on the techniques that are currently being developed within the ALIAS project for enabling an easy access to social events. In particular, it identified the objective and implementation details of the entertainment part of the ALIAS project. The work will be helpful to assist the elder people to identify new potentially interesting social events that they may wish to attend or relive (in the case of a past event). Besides the public social events, it is well known that private events [14] (such as birthday parties, weddings) also play important role in our daily life. In the future, we aim at extending our current framework so as to integrate private events detection/illustration.

Reference

[1] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How flickr helps us make sense of the world: context and content in community-contributed media collections," *15th ACM International Conference on Multimedia (ACM MM'07)*, Augsburg, Germany: 2007, pp. 631-640.

[2] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali, "Cluster-Based Landmark and Event Detection for Tagged Photo Collections," *IEEE Multimedia*, vol. 18, 2011, pp. 52-63.

[3] J. Hays and A.A. Efros, "IM2GPS: estimating geographic information from a single image," *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.

[4] R. Datta, D. Joshi, J. Li, James, and Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (CSUR)*, vol. 40, 2008.

[5] L. Kennedy and M. Naaman, "Less talk, more rock: automated organization of community-contributed collections of concert videos," *18th ACM International Conference on World Wide Web (WWW'09)*, Madrid, Spain: 2009, pp. 311-320.

[6] M. Ames and M. Naaman, "Why we tag: motivations for annotation in mobile and online media," *25th SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*, San Jose, CA, USA: 2007, pp. 971-980.

[7] D. Liu, X. Hua, M. Wang, and H. Zhang, "Image retagging," *18th ACM International Conference on Multimedia (ACM MM'10)*, Firenze, Italy: 2010, pp. 491-500.

[8] C. Liangliang, L. Jiebo, H. Kautz, and T.S. Huang, "Image Annotation Within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models," *IEEE Transactions on Multimedia*, vol. 11, 2009, pp. 208-219.

[9] Y. Avrithis, Y. Kalantidis, G. Tolias, and E. Spyrou, "Retrieving landmark and non-landmark images from community photo collections," *18th ACM International Conference on Multimedia (ACM MM'10)*, Firenze, Italy: 2010, pp. 153-162.

[10] M. Gao, X. Hua, and R. Jain, "WonderWhat: Real-time Event Determination from Photos," *20th World Wide Web Conference (WWW'11)*, Hyderabad, India: 2011.

[11] R. Shaw, R. Troncy, and L. Hardman, "LODE: Linking Open Descriptions Of Events," *4th Asian Semantic Web Conference (ASWC'09)*, 2009.

[12] X. Liu, R. Troncy, and B. Huet, "Finding Media Illustrating Events," *1st ACM International Conference on Multimedia Retrieval (ICMR'11)*, Trento, Italy: 2011.

[13] T. Wang, T. Mei, X. Hua, X. Liu, and H. Zhou, "Video Collage: A Novel Presentation of Video Sequence," *IEEE International Conference on Multimedia and Expo*, 2007, pp. 1479-1482.

[14] D. Joshi and J. Luo, "Inferring generic activities and events from image content and bags of geo-tags," *Proceedings of the 2008 international conference on Content-based image and video retrieval - CIVR '08*, New York, USA: 2008, p. 37.

[15] 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.