Event-complementing Online Human Life Summarization based on Social Latent Semantic Analysis

Klimis S. Ntalianis[1] and Anastasios D. Doulamis[2]
[1]*Athens University of Applied Sciences, Department of Marketing, Division of Computing, 28, Agiou Spyridonos str., Egaleo 12243, Athens, Greece*
[2]*National Technical University of Athens, School of Rural and Surveying Engineering, 9, Iroon Polytechniou str., Zografou 157 73, Athens, Greece*

Keywords: Online Human Life Summarization, Events Detection, Social Media, Social Computing, Social Latent Semantic Analysis.

Abstract: In this paper, online human life summarization is performed, based on multimedia content, published on social media. The life summaries are also automatically annotated with events, persons, places etc. Towards this direction, initially a content preparation module is activated that includes an intelligent wrapper. The content preparation module scans social networks, extracts their pages and segments them into tokens, in an unsupervised way. Next multimedia content is kept and it is associated to its respective metadata. In the following step, a novel ranking mechanism puts multimedia content in order of importance based on user-content interactions. Finally the event-complementing summarization module produces a meaningful annotated video clip, based on a spectral visual clustering technique and the innovative Social Latent Semantic Analysis algorithm. Experimental results illustrate the promising performance of the proposed architecture and set some foundations for future research.

1 INTRODUCTION

The largest social networks such as Facebook and Twitter have expanded rapidly during the last decade. Users share more and more information (personal videos/pictures/documents, youtube videos, flickr images, pinterest content etc) and are increasingly in control of how and when they view, create and post their favourite content. This stimulates new applications in the area of leisure and entertainment, e-democracy and e-business (Kokkinos et al, 2013).

On the other hand, currently an interesting initiative towards digital preservation of memories, heads in the creation of virtual interactive museums. A characteristic example includes the V-MUST.NET project (see http://www.v-must.net/), an EU FP7-funded Network of Excellence that aims to provide the heritage sector with the tools and support to develop Virtual Museums that are educational, enjoyable, long-lasting and easy to maintain. But, what about a virtual museum containing summaries of peoples' lives ? Instead of opening albums and viewing old video tapes wouldn't it be better to keep digital summaries of the lives of our ancestors, so that we can follow their experiences, life events, professional moments etc. and better keep their memory ? For example imagine that we had a multimedia summary of the life of Socrates, Napoleon, Isaac Newton, Christopher Columbus or Albert Einstein. How influential could it be?

Some years ago we did not have these capabilities, but now things have changed. For example, since social networks currently contain extremely vast amounts of information posted by their users, this information could possibly be used to create personal event-driven summaries. In particular, several users regularly post personal images/videos/graphics/documents in albums, which include important events, activities and occasions of their lives. This content may include several metadata. In particular it is associated to a posting date, it may state the place where it was created, it may also tag people, activities, buildings, events etc. Furthermore friends usually interact with posts, which receive likes, positive/negative comments or they may be shared by friends and other users. And

Ntalianis K. and Doulamis A..
Event-complementing Online Human Life Summarization based on Social Latent Semantic Analysis.
DOI: 10.5220/0005456506110622
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (MMS-ER3D-2015), pages 611-622
ISBN: 978-989-758-090-1

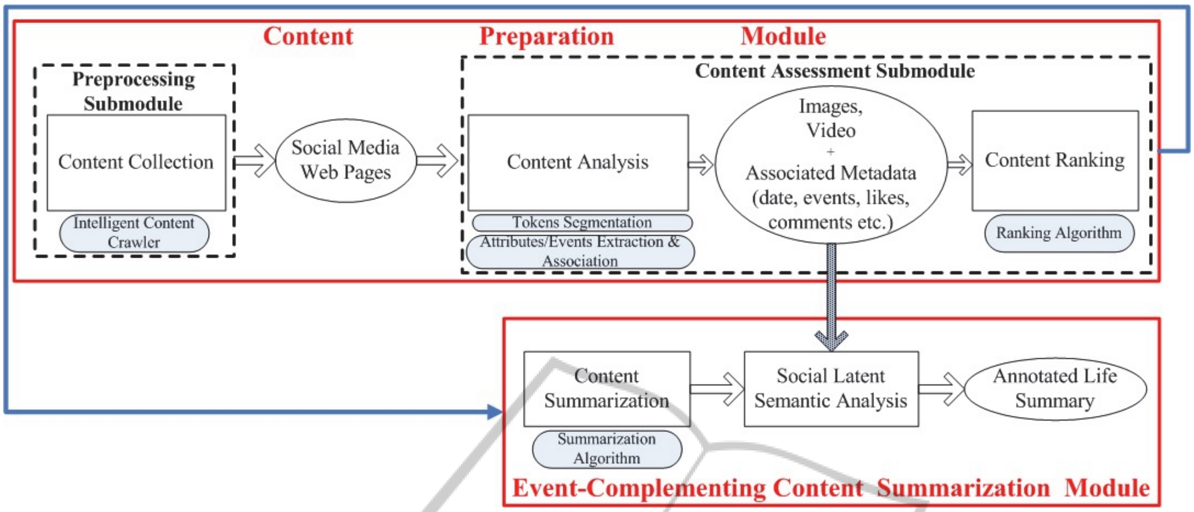

Figure 1: Overview of the proposed event-complementing life summarization scheme.

the question is: can we produce a meaningful searchable event-oriented multimedia summary of all this content by taking also into consideration the interactions, metadata and duration ?

This paper proposes an innovative event-complementing human life summarization scheme, which is based on social computing data over social media content. We aim at producing event summaries, where a compact and searchable overview of the life of each user is provided. Towards this direction, in this paper an innovative architecture is proposed, an overview of which is provided in Figure 1. In particular the proposed architecture for event-complementing human life summarization includes several novel components and it can be divided into two main modules: the content preparation and the event-complementing content summarization module. The content preparation module (CPM) scans the Web, finds social media web pages, analyzes them, detects multimedia content, extracts relevant metadata, associates the metadata to the multimedia content and orders the content according to its importance. On the other hand the event-complementing content summarization module receives at its input the ordered multimedia content, summarizes it and automatically annotates the summary. Experimental results on real life social networks content evaluate the robustness, scalability and flexibility of the proposed human life summarization scheme, revealing its advantages and limitations.

This paper is organized as follows: In Section 2 related work is presented. Section 3 discusses the content preparation module, while Section 4 focuses on the content summarization module and the innovative S-LSA algorithm. Experiments are presented in Section 5 and Section 6 concludes this work.

2 RELATED WORK

Regarding conventional video summarization, two extensive reviews of key-frame extraction and video summarization approaches are given in (Money and Agius, 2007), (Truong and Venkatesh, 2012). The presented interesting algorithms summarize single videos with selected still images or with a short summary video. However they do not consider social media content and implicit crowdsourcing metadata such as likes, comments and sharing.

On the other hand some works have been proposed in the literature, focusing on new summarization trends. In (Fabro, et. al., 2012) an algorithm for the summarization of real-life events based on community-contributed multimedia content is presented. The proposed event summarization algorithm uses photos from Flickr and videos from YouTube to compose summaries of well-known society events, which took place in the last three years. A summary is built according to search terms, specified by the user (e.g. Royal wedding of William and Kate). In (Wang, et. al., 2012) an approach for event driven web video summarization is proposed based on tag localization and key-shot mining. Initially the method localizes the tags that are associated with each video into its shots. Then the relevance of the shots is estimated with respect to the event query by matching the shot-level tags with the query. However, it cannot be straightforwardly applied to social media content and does not take into consideration user interactions.

In (Chua and Asur, 2013) a search and summarization framework is proposed to extract relevant representative tweets from a time-ordered sample of tweets to generate a coherent and concise summary of an event. Two topic models are introduced that take advantage of temporal correlation in the data to extract relevant tweets for summarization. The aforementioned approach focuses on text and does not consider any other kind of multimedia content. In (Wang, et.al, 2013) the task of personal profile summarization by leveraging both personal profile textual information and social networks is addressed. The use of social networks is motivated by the intuition that, people with similar academic, business or social connections tend to have similar experience and summaries. To achieve the learning process, the authors propose a collective factor graph model to incorporate all these resources of knowledge to summarize personal profiles with local textual attribute functions and social connection factors as is presented in (Doulamis, et al, 2013a,b) for personalized 3D navigation. However the work considers only textual information included in user profiles aiming mainly at building automatic resumes. An active learning algorithm for classifying user's preferences has been proposed in (Yiakoumettis et al, 2014).

The approach of (Yang, et. al., 2011) is based on modelling Web documents and social contexts into a unified framework. A dual wing factor graph (DWFG) model is proposed, which utilizes the mutual reinforcement between Web documents and their associated social contexts to generate summaries. An efficient algorithm is designed to learn the proposed factor graph model. Again this approach does not consider other multimedia content except of text. The work of (Hu et. al., 2011) performs social summarization by first employing the tripartite clustering algorithm to simultaneously discover document context and user context for a specified document. Then sentence relationships intra and inter documents plus intended user communities are taken into account to evaluate the significance of each sentence in different context views.

Finally, a few sentences with highest overall scores are selected to form the summary. This approach focuses only on text documents and does not analyze images or video. The work of (Meng, et. al.,2012) proposes a unified optimization framework to produce opinion summaries of tweets through integrating information from dimensions of topic, opinion and insight, as well as other factors. Their approach is limited to producing personalized summaries and does not provide audiovisual abstraction. In (Sinha, et.al., 2011) a framework for generating representative subset summaries from large personal photo collections is proposed. Three salient properties are defined that an informative summary should satisfy: quality, diversity and coverage. Methods are presented to compute these properties using multidimensional content and context data. This interesting approach does not consider video data. In (Gentili, et. al., 2012) events are defined as tuples (u, a, o, t), which mean that a user u performed the action a over the object o at time t. The authors aim to produce a concise summary of sequences of events related to time, based on the data size reduction obtained merging time intervals and collapsing the descriptions of more events in a unique descriptor or in a smaller set of descriptors.

However the proposed approach does not consider user interaction metadata and cannot be straightforwardly applied to social media content. In (Lee, et. al., 2012) a video summarization approach for egocentric or “wearable” camera data is proposed. Given hours of video, this method produces a compact storyboard summary of the camera wearer’s day. The resulting summary focuses on the most important objects and people with which the camera wearer interacts. This scheme is limited to video content coming from wearable cameras. Recent work also includes schemes proposed by (Ntalianis et. al., 2013, 2014). In that scheme, humans behavior understanding algorithms as in (Voulodimos et al, 2013) can be exploited.

Last but not least, Facebook has presented a very interesting application entitled «A Look Back» or «My Facebook movie» (Griggs, 2014). This service has been described as an experience that compiles your highlights since joining Facebook. Depending on how long you have been on Facebook and how much you have shared, you will see a movie, a collection of photos or a thank you card. The movie is about one minute long and includes the date when someone joined Facebook, their first moments and most liked posts and photos they have shared. However there are several limitations of this application: (a) it does not consider videos but only photos, (b) it does not annotate time instances of the summary with possible events, (c) it does not consider comments, and (d) it is limited to about one minute irrespectively of one’s activity.

3 CPM & CONTENT RANKING

In this section the CPM is described, the main product of which is an ordered list of multimedia items (MIs). Each item is associated to several metadata and its ranking among other items is calculated. The CPM, is divided into the Preprocessing and the Content Assessment Submodules. The Preprocessing Submodule gathers social media content, a task which is very challenging on rule-stringent social media like Facebook. For this reason we incorporate a middleware intelligent crawling architecture, which accomplishes content collection and analysis.

Aim of the Content Assessment Submodule is to evaluate content importance, associate it to its respective metadata and rank it. It consists of the content analysis and the content ranking components (CAC & CRC). The CAC initially segments a page into tokens and associates each posted MI to its related metadata. In case of a typical Facebook post, tokens include: the posted MI, the date when the MI was posted, the title of the MI, the "like" area, the shares area, the comments area and the person area of each comment.

3.1 The Proposed CRC

The CRC receives the MIs and associated metadata and attempts to meaningfully rank them, using a social computing approach. In this direction, social interactions may provide a very good clue regarding the "value" of a post. In particular, people tend to interact with few of their social media "friends" (Huberman, et. al., 2009), who are their actual friends. In this paper actual friends are explicitly considered. Towards this direction, the following definitions are made:

Definition 1: Let U_i be the i^{th} user of a social network.

Definition 2: The set FS_i of all friends of U_i is given by:

$$FS_i = \{F_i^1, F_i^2, ..., F_i^M\} \tag{1}$$

where F_i^M is the M^{th} friend of U_i.

Definition 3: An actual friend AF_i^k, k = 1, …, L of U_i, frequently interacts (likes, comments etc) with content posted by U_i or tagging U_i. At the same time U_i frequently interacts with content posted by AF_i^k or tagging AF_i^k.

Definition 4: Based on *Definition 3*, the set AFS_i of the actual friends of U_i is defined as:

$$AFS_i = \{AF_i^1, AF_i^2, ...AF_i^L\}, \tag{2}$$

where AF_i^L is the L^{th} actual friend of U_i, and $AFS_i \subseteq FS_i$.

Definition 5: For a multimedia item $MI^{i,m}$, m=1, …, G, posted by a user U_i, or tagging user U_i, three vectors are defined, $\mathbf{l}_{i,m}$, $\mathbf{p}_{i,m}$ and $\mathbf{c}_{i,m}$, corresponding to likes, shares and comments the MI has received respectively:

$$\mathbf{l}_{i,m} = [l^i_{F^1}, l^i_{F^2}, ..., l^i_{F^M}, l^i_{F^{M+1}}] \tag{3a}$$

$$\mathbf{p}_{i,m} = [p^i_{F^1}, p^i_{F^2}, ..., p^i_{F^M}, p^i_{F^{M+1}}] \tag{3b}$$

$$\mathbf{c}_{i,m} = [c^i_{F^1}, c^i_{F^2}, ..., c^i_{F^M}, c^i_{F^{M+1}}] \tag{3c}$$

where $l^i_{F^1}$ / $p^i_{F^1}$ equals to 1 if friend F_i^1 has liked/shared the respective MI, otherwise it equals to 0. $c^i_{F^1}$ equals to the number of comments friend F_i^1 has made to the respective MI, while $l^i_{F^{M+1}}$, $p^i_{F^{M+1}}$ and $c^i_{F^{M+1}}$ count the likes, shares and comments the MI has received from non-friends.

Definition 6: Let us denote as $L_{i,m}$, $P_{i,m}$ and $C_{i,m}$ three variables that count the total numbers of likes, shares and comments a MI has received respectively:

$$L_{i,m} = \sum_{r=1}^{M+1} l^i_{F^r} \quad P_{i,m} = \sum_{r=1}^{M+1} p^i_{F^r} \quad C_{i,m} = \sum_{r=1}^{M+1} c^i_{F^r} \tag{4}$$

Definition 7: Variable $DA_{i,m}$ over a multimedia item $MI^{i,m}$, m=1, …, G, denotes its duration of activity, capturing the first and last day the MI was shared or received a comment.

Definition 8: Variable $RA_{i,m}$ over a multimedia item $MI^{i,m}$, m=1, …, G, denotes how frequently a MI receives attention:

$$RA_{i,m} = \frac{L_{i,m} + P_{i,m} + C_{i,m}}{DA_{i,m}} \tag{5}$$

By taking into consideration the aforementioned measures, MIs are then ranked in seven steps:

Step 1: for a user U_i, i=1,…, N, and for a given time instance TP, gather all multimedia items $MI^{i,m}$, m=1, …, G, that have been posted by U_i or tag U_i.

Step 2: $\forall F_i^j \in FS_i$, j=1, …, M, calculate an interaction value $IV(F_i^j)$ between F_i^j and U_i. Gather all values $IV(F_i^j)$, j=1, …, M, to a vector $\vec{\mathbf{v}}_i$:

$$\vec{\mathbf{v}}_i = [IV(F_i^1), IV(F_i^2), ..., IV(F_i^M)] \tag{6}$$

***Step 3*:** Sort $\vec{\mathbf{v}}_i$ in descending order to produce $\vec{\mathbf{v}}_i^*$:

$$\vec{\mathbf{v}}_i^* = [IV(F_i^o), IV(F_i^q), ..., IV(F_i^p)] \tag{7}$$

with $IV(F_i^o) \geq IV(F_i^q) \geq, ..., \geq IV(F_i^p)$ and $o, q, p \in$ [1, …, M]. The top values of $\vec{\mathbf{v}}_i^*$ distinguish the AFS_i members.

***Step 4*:** Having estimated Eq. (7), an ordered set of U_i's friends is produced:

$$FS_i^* = [F_i^o, F_i^q, ..., F_i^p] \tag{8}$$

where F_i^o / F_i^p is the user who maximally/minimally interacts with U_i. Then for each $MI^{i,m}$, $L_{i,m}$, $P_{i,m}$ and $C_{i,m}$ are recalculated, by considering FS_i^*. In particular the ordering of FS_i^* is mapped to a weights vector $\mathbf{w}_i$ so that activities from actual friends are strengthened while activities from all others are weakened:

$$\mathbf{w}_i = [w_{F^o}^1, w_{F^q}^2, ..., w_{F^p}^M, w^{M+1}] \tag{9}$$

Eq. (9) contains M + 1 weights. The first M weights correspond to the list of M sorted friends of U_i (Eq. 8), while w^{M+1} corresponds to non-friends. Vectors $\mathbf{l}_{i,m}$, $\mathbf{p}_{i,m}$ and $\mathbf{c}_{i,m}$ are also sorted by following the ordering of set FS_i^*, forming $\mathbf{l}^*{}_{i,m}$, $\mathbf{p}^*{}_{i,m}$ and $\mathbf{c}^*{}_{i,m}$:

$$\mathbf{l}^*{}_{i,m} = [l_{F^o}^i, l_{F^q}^i, ..., l_{F^p}^i, l_{F^{M+1}}^i] \tag{10a}$$

$$\mathbf{p}^*{}_{i,m} = [p_{F^o}^i, p_{F^q}^i, ..., p_{F^p}^i, p_{F^{M+1}}^i] \tag{10b}$$

$$\mathbf{c}^*{}_{i,m} = [c_{F^o}^i, c_{F^q}^i, ..., c_{F^p}^i, c_{F^{M+1}}^i] \tag{10c}$$

Then $L'_{i,m}$, $P'_{i,m}$ and $C'_{i,m}$ are calculated by the dot product:

$$\begin{aligned} L'_{i,m} &= \boldsymbol{l}^*{}_{i,m} \cdot \boldsymbol{w}_i \\ P'_{i,m} &= \boldsymbol{p}^*{}_{i,m} \cdot \boldsymbol{w}_i \\ C'_{i,m} &= \boldsymbol{c}^*{}_{i,m} \cdot \boldsymbol{w}_i \end{aligned} \tag{11}$$

***Step 5*:** Estimate the average variable $RA'_{i,m}$ as:

$$RA'_{i,m} = \frac{L'_{i,m} + P'_{i,m} + C'_{i,m}}{DA_{i,m}} \tag{12}$$

Step 6: Estimate the importance $I_{i,m}$ of each multimedia item $MI^{i,m}$, m = 1,… G:

$$I_{i,m} = \left(w^L \cdot \frac{L'_{i,m}}{L_{i,m}} + w^P \cdot \frac{P'_{i,m}}{P_{i,m}} + w^C \cdot \frac{C'_{i,m}}{C_{i,m}} \right) \cdot \frac{RA'_{i,m}}{RA_{i,m}} \tag{13}$$

where w^L, w^P and w^C control the importance of likes, shares and comments in the ranking process. In Eq. (13) the division of terms of Eq. 11 by terms of Eq. 4 plays a normalization role, since the latter terms are not affected by the friends' ordering process.

Step 7: Gather all $I_{i,m}$'s , m = 1,… G, into set $SI_{i,m}$:

$$SI_{i,m} = [I_{i,1}, I_{i,2}, ..., I_{i,G}] \tag{14}$$

Finally sort $SI_{i,m}$ to produce $SI^*{}_{i,m}$:

$$SI^*{}_{i,m} = [I_{i,w}, I_{i,z}, ..., I_{i,y}] \tag{15}$$

with $I_{i,w} \geq I_{i,z} \geq ... \geq I_{i,y}$ and $w, z, y \in$ [1, …, G].

$SI^*{}_{i,m}$ contains all measures of importance for all MIs posted by U_i or tagging U_i. The order of the measures of importance determines the order of importance for each MI. More/less important MIs are summarized in finer/coarser detail and presented first/last.

4 EVENT-COMPLEMENTING CONTENT SUMMARIZATION BASED ON THE SOCIAL LSA

The event-complementing content summarization module attempts to create and unsupervisedly annotate the most representative summaries, exploiting the output of the CRC and the visual characteristics of each MI. Here, clustering of MIs based on their visual features is very important, since "uncorrelated" content that covers the whole storyline should be included.

Let us denote as $\mathbf{d}_m^i$ a descriptor vector that represents the visual content of $MI^{i,m}$. There are several ways to estimate $\mathbf{d}_m^i$ based on global/local features. Global descriptors provide an average of the visual information, whereas local descriptors are more suitable for describing specific areas. Local descriptors include FAST (Rosten and Drummond, 2006), SURF and SIFT or recently ORB (Rublee, et. al., 2011). In this paper, the extended MPEG-7 descriptors are used (Spala, et. al. 2012).

For creating a representative summary, a graph-based partitioning algorithm is adopted to form key-representative clusters. Spectral graph partitioning is incorporated instead of e.g. k-means, since it can simultaneously localize both intra-cluster coherence and inter-cluster separation. In addition, it can partition the space into complex regions allowing the extraction of more sufficient summaries than other conventional approaches.

4.1 Graph-based Representation

Let $G=\{V,E\}$ be a graph. A vertex $v\in V$ represents a MI, while the edges $e_{m,j}$ the correlation degree between two MIs. In particular, $e_{m,j}$ is defined as the correlation coefficient of the visual descriptors $\mathbf{d}_m^i$ and $\mathbf{d}_j^i$ respectively.

$$e_{mj} = Corr(\mathbf{d}_m^i, \mathbf{d}_j^i) = \frac{{\mathbf{d}_m^i}^T \cdot \mathbf{d}_j^i}{\sqrt{{\mathbf{d}_m^i}^T \cdot \mathbf{d}_m^i}\sqrt{{\mathbf{d}_j^i}^T \cdot \mathbf{d}_j^i}} \tag{16}$$

Cross-correlation presents advantages compared to the Euclidean distance, which is sensitive to feature vector scaling and/or translation (Doulamis and Doulamis, 2006). For this reason, normalized cross-correlation has been widely used as it remains unchanged after feature vector scaling or translation.

4.2 Spectral Visual Clustering

Using the graph representation, we estimate M mutually exclusive clusters, which are as "uncorrelated" as possible with samples belonging to different clusters and as coherent as possible with samples of the same cluster. "Uncorrelation" means that the M clusters are able to represent the whole storyline and it is formulated as:

$$\begin{aligned} \hat{C}_r :&\min \sum_{i=1}^{M} P_r = \\ &= \sum_{m\in C_r, j\notin C_r} e_{m,j} \quad \max \sum_{i=1}^{M} Q_r = \sum_{m\in C_r, j\in C_r} e_{m,j} \end{aligned} \tag{17}$$

In Eq. (17), $\hat{C}_r$ is the optimal r-th partition of the relevant set C among the M requested, while $e_{m,j}$ is a metric distance between two MIs as defined by Eq. (16). The left hand of Eq. (17) minimizes the overall correlation between clusters, satisfying the concept of "uncorrelation", while the right hand maximizes coherence within a cluster. The main limitation of Eq. (17) is that optimization favors the creation of small clusters, since as the number of elements of a class increases, the respective cost $\sum_{i=1}^{M} P_r$ also increases. To face this difficulty, normalization factors are included in Eq. (17), resulting in the following optimization problem:

$$\begin{aligned} &\hat{C}_r \; :\max Q = \sum_{r=1}^{M} NQ_r = \sum_{i=1}^{M} \frac{\sum_{m\in C_r, j\in C_r} e_{m,j}}{\sum_{m\in C_r, j\in C_r} e_{m,j}} \\ &\text{and } \min P = \sum_{r=1}^{M} NP_r = \sum_{i=1}^{M} \frac{\sum_{m\in C_r, j\notin C_r} e_{m,j}}{\sum_{m\in C_r, j\in C_r} e_{m,j}} \end{aligned} \tag{18}$$

where Q and P are the normalized quantities of Q_r and P_r respectively, and $C = \cup_{r=1}^{M} C_r$. Since it is easy to prove that $P+Q=M$, the aforementioned optimization problem can be solved only by minimizing variable P.

Then Eq. (18) can be written in matrix form as:

$$\mathbf{e}_r : \min P = \min \sum_{r=1}^{M} \frac{\mathbf{e}_r^T \cdot (\mathbf{L}-\mathbf{E}) \cdot \mathbf{e}_r}{\mathbf{e}_r^T \cdot \mathbf{L} \cdot \mathbf{e}_r} \tag{19}$$

where $\mathbf{E}$ is the graph's adjacent matrix, that is $\mathbf{E}=[e_{m,j}]$, while $\mathbf{L}$ is a diagonal matrix $\mathbf{L} = diag(\cdots l_i \cdots)$, the elements of which equal $l_i = \sum_{j\in C} e_{m,j}$. Vector $\mathbf{e}_r$ is equal to 1 if the m-th MI belongs to the r-th partition or zero otherwise.

Minimization of Eq. (19) can be obtained only under the assumption that the elements of $\mathbf{e}_r$ receive continuous values instead of binary. The concept is to initially estimate the continuous version of $\mathbf{e}_r$ and then discrete the solution to take binary values. Under the assumption of continuity regarding $\mathbf{e}_r$, optimization of Eq. (19) is obtained through the estimation of the generalized eigenvectors of the matrices $\mathbf{L}$ and $\mathbf{E}$. In this way, we estimate the continuous version of the index vector denoted as $\mathbf{e}_r^c$. Then, the problem is how to round vector $\mathbf{e}_r^c$ to obtain discrete values. A simple rounding process is to set the maximum value of each row of $\mathbf{e}_r^c$ equal to one and the remaining values equal to zero.

4.3 Creation of Summaries

Next, the M clusters are created in a way so that each element contains as uncorrelated MIs as possible. Therefore, in order to create a spherical view of the storyline of a user, we need to extract one or more items from each cluster. Initially, for every cluster, a score I_{C_r} is assigned as the average ranking criterion of all MIs belonging to C_r :

$$I_{C_r} = \sum_{i \in C_r} \frac{I_{i,m}}{|C_r|} \tag{20}$$

In Eq. (20), $|\cdot|$ expresses the cardinality of C_r and I_{C_r} is the importance of a cluster. So, the higher the score, the more significant a cluster is. Therefore, the score I_{C_r} indicates the percentage of MIs extracted from each cluster. Let us denote by S the scale of a summary. S expresses a summary's level of detail and when it increases, more MIs are included in the summary. Then, C_r contributes to a summary by $I_{C_r} \cdot S$ MIs. It is clear that within a cluster C_r, each MI has a score $I_{i,m}$. Therefore, for each cluster the $I_{C_r} \cdot S$ highest scored MIs are extracted. By collecting data for every of the M clusters, we construct the multimedia summary at scale S.

4.4 Event-complementing Annotation of Summary

Latent semantic analysis (LSA) is a technique in natural language processing that analyzes relationships between a set of documents and the terms they contain, by producing a set of concepts related to the documents and terms (Landauer and Dumais, 1997). LSA assumes that words that are close in meaning will occur in similar pieces of text. A matrix containing word counts per paragraph is constructed from a large piece of text and singular value decomposition (SVD) is used to reduce the number of rows, while preserving the similarity structure among columns. Words are then compared, by taking the cosine of the angle between the two vectors, formed by any two rows. Values close to 1 represent very similar words, while values close to 0 represent very dissimilar words.

In this paper and in the framework of social media, the Social LSA (S-LSA) is introduced, in order to also consider interactions among users and content. In particular, in our case specialized analysis is performed per user, since the friends of a user may use their own vocabulary, expressions etc. Additionally, the title of a post as well as comments made by friends of a user, also receive likes (meaning that they are approved). Thus the keywords of this kind of social dialogue should be further strengthened. Towards this direction let X be a matrix where element $x_{i,j}$ describes the occurrence of term i in the associated text area of the j^{th} MI:

$$X = \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \dots & \dots & \dots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix} \tag{21}$$

Let also Y be a matrix, where element $y_{i,j}$ describes the total likes a comment has received, which also contains term $x_{i,j}$:

$$Y = \begin{bmatrix} y_{1,1} & \dots & y_{1,n} \\ \dots & \dots & \dots \\ y_{m,1} & \dots & y_{m,n} \end{bmatrix} \tag{22}$$

Then in our case the strength of each term in a social framework is defined by the Hadamard product:

$$\begin{gathered} X \circ Y = \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \dots & \dots & \dots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix} \circ \begin{bmatrix} y_{1,1} & \dots & y_{1,n} \\ \dots & \dots & \dots \\ y_{m,1} & \dots & y_{m,n} \end{bmatrix} = \\ = \begin{bmatrix} x_{1,1}y_{1,1} & \dots & x_{1,n}y_{1,n} \\ \dots & \dots & \dots \\ x_{m,1}y_{m,1} & \dots & x_{m,n}y_{m,n} \end{bmatrix} = \begin{bmatrix} z_{1,1} & \dots & z_{1,n} \\ \dots & \dots & \dots \\ z_{m,1} & \dots & z_{m,n} \end{bmatrix} = Z \end{gathered} \tag{23}$$

Now a row t_i^T in Z will be a vector corresponding to a term, providing its extended relation to each SVP, while a column in Z will be a vector, giving its relation to each term contained in the associated textual information $d(MI)$ of an MI:

$$t_i^T = [z_{i,1} \dots z_{i,n}] \tag{24(a)}$$

$$d(MI^j) = \begin{bmatrix} z_{1,j} \\ \dots \\ z_{m,j} \end{bmatrix} \tag{24(b)}$$

where, for simplicity of notation, the page index has been eliminated from the MI.

Now the dot product $t_i^T t_p$ gives the correlation between terms i and p over all MIs, while ZZ^T provides dot products for all terms. Furthermore let us assume that a decomposition of Z exists such that U and V are orthogonal matrices:

$$Z = U \Sigma V^T \tag{25}$$

while Σ is a diagonal matrix of the form:

$$\Sigma = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_l \end{bmatrix} \tag{26}$$

Date: 06 August 2012
Title of Album: Cover Photos
Title of Picture: pagaki of skiathos — together with Olga Chrysafogeorgou
Likes: 16 (**from** Katerina Gkaravella, Elizabeth Karahanidi, Αλεξάνδρα Σκαρμέα, χάρης χάρης, PanagiotisViper Vlachos A, Ioanna Tsami, Mixalis Zaranis, Babar Hussain, Johanna Vassilopoulou, Raf Trifon, Ειρηνη Γκικακη, Hara Barka, Όλγα χρυσαφογεώργου, Vanessa Boukoura, ΝΤΑΣΙΩΤΗ ΜΑΡΙΑ, Yannis Pappas).
Total Number of Comments: 4 (**from** Όλγα χρυσαφογεώργου → 2, from Katerina Gkaravella → 2).
Shares: 0
First day of Activity: 06 August 2012
Last day of Activity: 06 August 2012

Figure 2: Output of the content analysis component (image & associated metadata).

The matrix products giving us the term and textual information of *MI*s correlations then become:

$$ZZ^T = U\Sigma\Sigma^T U^T \qquad (27a)$$

$$Z^T Z = V\Sigma^T\Sigma V^T \qquad (27b)$$

Since $\Sigma\Sigma^T$ and $\Sigma^T\Sigma$ are diagonal, U contains the eigenvectors of ZZ^T, while V contains the eigenvectors of Z^TZ. Both products have the same non-zero eigenvalues given by the non-zero entries of $\Sigma\Sigma^T$ and $\Sigma^T\Sigma$ respectively. Additionally when the k largest singular values among $\sigma_1, \ldots, \sigma_l$ and their corresponding singular vectors from U and V are selected, the rank k approximation of Z is accomplished and can be written as:

$$Z_k = U_k \Sigma_k V_k^T \qquad (28)$$

Based on Eq.(28) terms i and p can be compared, by comparing the vectors $\Sigma_k \hat{\mathbf{t}}_i^T$ and $\Sigma_k \hat{\mathbf{t}}_p^T$. In this paper *MI*s are associated to the terms that best approximate them, so that each *MI* is enriched with events, places, persons, time etc, providing better content understanding.

Figure 3: Top 5 MIs for U_{78} (ranking mechanism).

Now the CRC receives 261 MIs, aiming at putting them into an order from the most to the less important. Set

5 EXPERIMENTAL RESULTS

In order to evaluate the proposed scheme, on 03/02/15 we have recorded the "Albums" information of 120 Facebook friends of the Online Computing Group that can be found at: www.facebook.com/klimis.ntalianis.7. In total 611 videos and 26,004 pictures were gathered, providing on average 5 videos and 216.7 pictures per friend. In parallel, the preprocessing submodule gathered and associated to each *MI* its respective metadata. For visualization purposes, the results over U_{78} are presented, whose albums contained 2 videos and 259 pictures. Next the CAC is applied, providing in total 261 combinations of MIs and associated metadata.

One such combination is provided in Figure 2, where L=16, P=0, C=4 (Eq. 4) and DA=1 (Def. 7). FS_{78} of all friends of U_{78} contains 703 persons. In order to calculate $L'_{i,m}$, $P'_{i,m}$ and $C'_{i,m}$ (Eq. 11), FS_i^* (Eq. 8) should be estimated, which shorts the friends of U_{78} according to their interaction values ($IV(F_i^j)$). Interaction value between U_{78} and her 703 friends have been estimated for the data recorded on 03/02/15. Furthermore T_{AF} was set equal to 2 %. Based on $IV(F_i^j)$, $\vec{\mathbf{v}}_{16}^*$ (Eq. 7) is estimated and then FS_i^* is calculated. In this case the top 2% of U_{78}'s friends are considered as actual friends, or 14 persons in total. Based also on the set of actual friends, $L'_{i,m}$, $P'_{i,m}$ and $C'_{i,m}$ were calculated, where weights vector $\mathbf{w}_{16}$ (Eq. 9) were experimentally set to take values in the interval [3, 0.01]. Finally $SI^*_{78,m}$, m = 1, …., 261, is estimated, containing all MIs from the most to the less important. For visualization purposes the top 5 are presented in Figure 3. As it can be observed, all of them contain U_{78} in different poses. Having this into mind, the proposed summarization algorithm tries to unsettle this kind of theme monotony by visually clustering

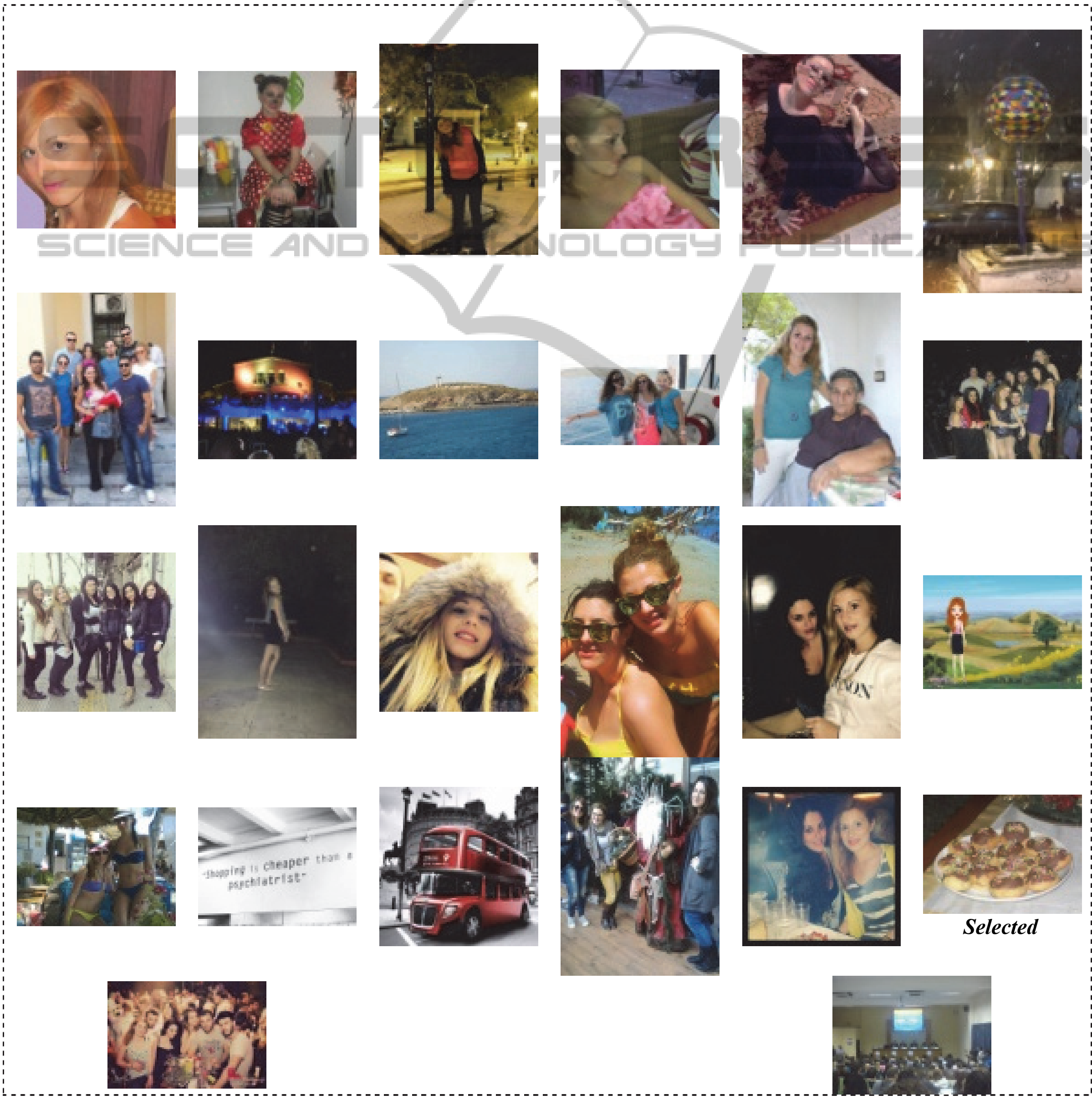

Figure 4: The online life summary of user U_{78} (3rd of February 2015).

content. In our experiments and in case of U_{78}, 5 clusters were created and, based on scores I_{C_r} , 25 images and 1 video key-frame were extracted. All 26 MIs are integrated into a video, similarly to the «My Facebook movie» application. The summary is provided in Fig. 4.

Table 1: Parts of the *d*(*MI*) vector for the selected *MI* of Figure 4. The second, fourth and sixth column contain terms, while the first, third and fifth column contain the respective $z_{i,j}$ values.

Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6
0	**Absorbed**	0	**King**	4	Sweet
0	**Beer**	0	**Lonely**	0	**Talk**
0	**Candle**	31	New	0	**Tired**
5	Double	4	Order	7	Unfair
0	**Dozen**	0	**Parrot**	0	**Ultimate**
4	Eat	6	Piece	31	Vanish
0	**Frightened**	0	**Query**	0	**Vine**
31	Girls	31	Set	8	Want
0	**Hide**	4	Small	4	Yesterday
7	Jealous	0	**Star**	0	**Zircon**

Finally the S-LSA algorithm has annotated (per image) the automatically produced summary, by taking into consideration both titles and comments of all 261 MI. In particular, the associated text of each MI was analyzed to its words and stop words have been removed, using the Page Analyzer's list (http://www.ranks.nl/). As a result 857 unique terms have remained, while the mean number of terms per MI was equal to 3.28. Analysis for one MI of Figure 4, marked by "selected" is provided. In particular this MI, which contains a plate of sweets, has a translated title "Girls the new set will vanish". Furthermore L=31, P=0 and C=22 respectively (Eq. 4), while it had 62 unique terms. Now regarding vector $d(MI)$ of Eq. 24(b) for the *MI* under consideration, parts of it are presented in the first, third and fifth column of Table I. The vector has size 857 × 1 and since the specific *MI* has only 62 unique terms, it is very sparse. The respective terms (in alphabetical order) are also presented in the second, fourth and sixth column. As it can be observed, the S-LSA takes into consideration also user interactions (likes made to the title and comments), which strengthen specific terms. Among the terms that gain more strength are words that are included in the title of the *MI*. For the *MI* under consideration the top 14 annotation terms, according to score, were (Greek terms translated to English): girls, new, set, vanish, want, unfair, jealous, piece, double, sweet, eat, order, small, yesterday.

6 CONCLUSIONS

In this paper we have presented an innovative event-complementing human life summarization scheme, based on a social computing methodology over social media content. In particular, 120 summaries have been composed, corresponding to members of the Online Computing Group. The proposed scheme, except of achieving information reduction, it also provides sufficient summaries. The only major complain from users was focused on the duration of the summary (in some cases more than 7 minutes). This issue could be confronted e.g. by multiresolution summaries, where a user would be able to zoom in or out to content of interest.

Future work can take many directions. First of all an intelligent mechanism could be implemented to gather and integrate information of a user from as many online sources as possible. This would provide a much better profile of one's network life (their interests, habits, activities etc.) and maybe lead to a more inspired summary. Secondly a mechanism to take into consideration also time would reveal new dimensions of the problem. Currently content is gathered for a specific time instance without taking into consideration the life cycle of a multimedia item. Furthermore results should be normalized (e.g. by taking into consideration the percentage of friends that interact with a post) since a user may currently have 100 friends and a year later may have 1,000 friends. The more the friends, the more the interactions. Thus old time moments may be considered insignificant.

Additionally a sentiment analysis module could also be integrated to check the polarity of comments (positive, negative, neutral), so that polarity is also included into the ranking mechanism. Another interesting research direction has to do with distinguishing actual from non-actual friends (and setting threshold T_{AF} and weights $\mathbf{w}_i$). To do so, statistics and formulas based on the interaction values could be introduced. Noise detection algorithms could also be incorporated for excluding irrelevant content from the summary. Finally methods that analyze web pages based on their visual appearance can be incorporated so that the proposed scheme can be applied also to other types of web sites.

ACKNOWLEDGEMENTS

The research leading to these results has been supported by European Union funds and National funds (GSRT) from Greece and EU under the project JASON: Joint synergistic and integrated use of eArth obServation, navigatiOn and commuNication technologies for enhanced border security funded under the cooperation framework.

REFERENCES

Chua, F. C. T. and Asur, S., 2013 Automatic Summarization of Events From Social Media. *7th International AAAI Conference on Weblogs and Social Media*, Cambridge, USA.

Doulamis, N., and Doulamis, A., 2006. *Evaluation of Relevance Feedback Schemes in Content-based in Retrieval Systems. Signal Processing: Image Communication,* Vol. 21, No 4, April.

Doulamis, N., Yiakoumettis, C., Miaoulis, G. 2013. Personalised 3D navigation and understanding of Geo-referenced Scenes. *IEEE 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks,* WoWMoM.

Doulamis, N., Yiakoumettis, C., Miaoulis, G., Protopapadakis, E., 2013. A constraint inductive learning-spectral clustering methodology for personalized 3D navigation. *Lecture Notes in Computer Science.*

Fabro, M., Sobe, A. and Boszormenyi, L., 2012. Summarization of Real-Life Events Based on Community-Contributed Content. *Fourth International Conferences on Advances in Multimedia,* pp. 119 – 126, France, May.

Gentili, E., Milani A., and Poggioni, V., 2012 Data Summarization Model for User Action Log Files. In Proc. of the International Conference on Computational Science and Its Applications (ICCSA), vol. 7335, pp. 539–549.

Griggs, B., 2014. Your Facebook life in 62 seconds," CNN, February 6, http://edition.cnn.com/2014/02/04/tech/social-media/facebook-look-back-video/.

Hu, P., Sun, C., Wu, L., Ji, D., and Teng, C., 2011. Social Summarization via Automatically Discovered Social Context. In Proc. of the International Joint Conference on Natural Language Processing, p.p. 483-490, 2011.

Huberman, B., Romero, D., and Wu, F., 2008. Social networks that matter: Twitter under the microscope," arXiv:0812.1045v1, December.

K. Raftopoulos, K. Ntalianis, D. Sourlas and S. Kollias, 2013. Mining User Queries with Markov Chains: Application to Online Image Retrieval. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 2, p.p. 433 – 447, Feb. 2013.

Kokkinos, P., Koumoutsos, K., Doulamis, N., et al 2013. PERIKLIS - Electronic democracy in the 21st century using mobile and social network technologies. *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI).*

Landauer T.K., and. Dumais, S.T, 1997. A solution to Plato's problem: the latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, Vol. 104, p.p. 211–240, 1997.

Lee, Y.J., Ghosh, J., and Grauman, K., 2012. Discovering important people and objects for egocentric video summarization. In Proc. of the Computer Vision and Pattern Recognition Conference (CVPR), p.p. 1346-1353.

Meng, X., Wei, F., Liu, X., Zhou, M., Li, S., and Wang, H., 2012. Entity-Centric Topic-Oriented Opinion Summarization in Twitter. In Proc. of the International Conference of Knowledge Discovery and Data Mining (KDD), p.p. 379-387, 2012.

Money, A. G., and Agius, H., 2008. Video summarisation: A conceptual framework and survey of the state of the art. *Journal of Visual Communication and Image Representation*, Vol. 19, No. 2, p.p. 121–143.

Ntalianis, K. S., Doulamis, A., and Doulamis, N., 2014. An Automatic Event-Oriented Human Life Summarization Scheme based on Social Computing over Social Media Content. Multimedia Tools and Applications, Springer.

Ntalianis, K.S., Doulamis, A., Doulamis, N., and Tsapatsoulis, N., 2013. Video Abstraction in Social Media: Augmenting Facebook's Edgerank Algorithm in Video Content Presentation. In Proc. of the IEEE International Conference on Image Processing (ICIP '13), Melbourne, Australia, September.

Rosten E., and Drummond, T., 2006. Machine Learning for High-Speed Corner Detection. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 430–443.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. *In Proc. the IEEE International Conference on Computer Vision (ICCV)*, 2011.

Sinha, P., Mehrotra S., and Jain, R., 2011. Summarization of personal photologs using multidimensional content and context. In Proc. of the 1st ACM International Conference on Multimedia Retrieval, p.p. 4.1 – 4.8, New York, USA.

Spala, P., Malamos, A.G., Doulamis, A., Mamakis, G., 2012. Extending MPEG-7 for efficient annotation of complex web 3D scenes. *Multimedia Tools and Applications,* Springer, vol. 52, no. 2, pp. 463-504.

Truong, B. T., and Venkatesh, S., 2007. Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications and Applications*, Vol. 3, No. 1.

Voulodimos, A., Doulamis, N., Tsafarakis, S. 2013. Behavior recognition from video based on human constrained descriptor and adaptable neural networks. *ARTEMIS 2013 - Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream.*

Wang, M., Hong, R., Li, G.,. Zha, Z.-J, Yan, S. and Chua, T.-S., 2012. Event Driven Web Video Summarization by Tag Localization and Key-Shot Identification.

IEEE Transactions on Multimedia, Vol. 14, No. 4, p.p. 975 – 985.

Wang, Z., Li, S., Fang, K. and Zhou, G., 2013. Collective Personal Profile Summarization with Social Networks. *In Proc. of the Conference on Empirical Methods in Natural Language Processing,* Seattle, Washington, p.p. 715–725, October.

Yang, Z., Cai, K., Tang, J., Zhang, L., Su, Z., and Li., J., 2011. Social Context Summarization. *In Proc. of the Special Interest Group on Information Retrieval* (SIGIR), p.p. 255 – 264, 2011.

Yiakoumettis, C., Doulamis, N., Miaoulis, G., Ghazanfarpour, D., 2014. Active learning of user's preferences estimation towards a personalized 3D navigation of geo-referenced scenes. GeoInformatica, 18 (1), pp. 27-62.