Semantic Structuring and Retrieval of Event Chapters in Social Photo Collections

Krishna Chandramouli, Ebroul Izquierdo
Multimedia and Vision Research Group
School of Electronic Engineering and Computer Science
Queen Mary, University of London, Mile End Road, E1 4NS, London, UK
{krishna.chandramouli, ebroul.izquierdo}@elec.qmul.ac.uk

ABSTRACT

The phenomenal growth of multimedia content on the web over the last couple of decades has paved the way for content management systems integrating intelligent information retrieval and indexing techniques. Also, in order to improve the performance of retrieval techniques while searching and navigating the database, many relevance feedback algorithms are implemented, in which the subjective semantics of individual users are included in the image search. Following the recent developments in social networking, there is an emerging interest to share experiences online with friends using multimedia data. As the experiences to be shared among social peers vary from a simple social gathering to a tourism visit with a group of peers, there is a critical need for intelligent content management tools driven by a social perspective. Addressing the challenges related to socially-driven content management, the objective of this paper is twofold. First, we investigate techniques to intelligently structure multimedia content to enable efficient browsing of photo albums. The proposed structuring schemes exploit EXIF metadata, visual content and social peer relationships. Second, we propose a retrieval model based on social context to identify users with similar interests. The retrieval model aims to allow increased interaction among social peers. The proposed techniques have been evaluated against tourism pictures captured across Europe.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Clustering, Retrieval models, Information filtering

General Terms

Algorithms, Experimentation

Keywords

User generated content, Particle Swarm Optimisation, EXIF metadata, Social context, Image Retrieval, Metadata fusion

MIR'10, March 29–31, 2010, Philadelphia, Pennsylvania, USA.

1. INTRODUCTION

On one end of the technology spectrum, the widespread use of the World Wide Web (WWW) has inevitably led to the Social Networking paradigm (SN). In this paradigm, users are connected with each other using relationships like friendship, kinship or relationships of belief. On the other end of the spectrum, due to the increasing availability of inexpensive digital capturing equipments, there is an exponential growth of user generated content (UGC). The designation UGC is used to refer to content generated by day-to-day users wanting to share experiences with social peers. Some of the characteristics of UGC (as opposed to professional content generators such as broadcasters and movie producers) include uncontrolled environment capture of targeted sets of semantic concepts and distortions during capture. Popular platforms enabling social networking include Facebook[1], Flickr[2], MySpace[3], Twitter[4] and LinkedIn[5]. Of these social networking platforms, Facebook has been the most visited social network, with nearly 65 million users [6] currently accessing Facebook through their mobile devices.

Consistent with the saying "A picture's worth 1000 words", users prefer to share experiences via multimedia, which has resulted in an explosive growth of multimedia data. For example, Facebook contains more than 3 billion pictures generated mostly by common users. Similarly, Flickr, an online photo management and social networking website, contains more than 200 million pictures. The number of multimedia items generated, accessed and shared is also evident through the social peers an individual is linked with on the web. For example, Facebook contains 69 user links on average for each individual user. Although technological developments in computer vision community have been visible, the application of these technologies is not evident.

Addressing the problem of the "Semantic Gap" which is succinctly defined as the gap between low-level features and high-level semantic features, a number of indexing and retrieval algorithms have been reported in the literature. Although the performance of the machine learning techniques has largely been improved, the results are still far away from results generated by human cognition. Addressing this problem, recent developments in optimisation techniques have been inspired by problem solving abilities of biological or-

[1]http://www.facebook.com
[2]http://www.flickr.com/
[3]http://www.myspace.com/
[4]http://twitter.com/
[5]http://www.linkedin.com/
[6]As of 15th December 2009

ganisms such as bird flocking and fish schooling. One such technique developed by Eberhart and Kennedy is called Particle Swarm Optimisation (PSO). The PSO algorithm has two main assertations as listed below [7].

- Mind is Social: Learning from experience and emulating the successful behaviours of others, people are able to adapt to complex environments through discovery of relatively optimal patterns of attitudes, beliefs and behaviours.
- Particle swarms are a useful computational intelligence methodology: Central to the concept of computational intelligence is system adaptation that enables or facilitates intelligent behaviour in complex and changing environments. Swarm intelligence comprises of three steps namely evaluate, compare and imitate. Each particle goes through these stages by performing simple mathematical operations in solving a more complex optimisation problem.

The objectives of this paper are twofold. First, we propose techniques to intelligently segment an individual album into event chapters. The term "event chapter" refers to a collection of images which are segmented using a temporal context. As opposed to an event (only) segmentation, a chapter could potentially consist of one or many events. We reserve the term "event" for specifying a particular semantic action that could be visually extracted from the multimedia content. The chapter segmentation schemes are based on three different parameters namely (i) events defined using Exchangeable image file format (EXIF) metadata [7]; (ii) content analysed using the PSO algorithm and (iii) friend-of-a-friend (FOAF) relationships. Second, we propose a retrieval model based on social context for identifying users sharing similar interests. The retrieval model exploits the visual coherence of the images shared by the individual user with image albums created by a social peer.

The remainder of the paper is organised as follows. In Section 2 an overview of literature review is presented for event segmentation from albums and image retrieval based on social context. A framework integrating both approaches proposed in this paper is discussed in Section 3, followed by different event chapter segmentation techniques in Section 4. In Section 6 the proposed retrieval model based on social context is discussed. The experimental results are presented in Section 7, followed by conclusion, remarks and future work in Section 8.

2. RELATED RESEARCH

In general, the information available from a photo can be classified into two categories: content-based and context-based. Content-based information includes low-level features like colour, texture, shape, etc. and high-level features that are derived from the pixel intensity. Context-based metadata refers to the information about the photo. Time and location mainly provide temporal and spatial information. Although they can provide some cues about scenes (for example, time can tell day/night and location can tell indoor/outdoor) the cues are limited. On the other hand, content-based features provide more information about the nature of scenes. Different content-based features describe different aspects of the scene photographed. Automatic event detection from personal photos still remains a challenging issue and has received most attention from the researchers. Typically, an event is defined as the group of photos captured in relative proximity in time. Google Picasa [14] organises the photos only by the date information. In PhotoTOC [17] a locally adaptive threshold is applied to time interval to group photos into a table of contents. But the results are sensitive to the predefined thresholds and content-based clustering is only used as a backup in post-processing.

In [11], a probabilistic framework for event based photo clustering was presented based on latent semantic concept. The photo event is considered as a latent semantic concept, while the generation process of captured photos is modelled by a generative model. The multimodal metadata of the photo belonging to the same event are assumed to exhibit coherence. The Expectation Maximization (EM) algorithm is employed to estimate model parameters and the number of events is determined by Minimum Description Length (MDL) principle. Platt [16] described an algorithm that generates a new event if a new photo is taken more than a certain amount of time since the last photo. Later this algorithm was improved by introducing an adaptive threshold scheme [17]. Loui and Savakis [10] described another time-based algorithm for automatic event segmentation. In this work, the authors first extract the time information and compute the time difference histogram. Then they divide the histogram into two parts using a K-means clustering algorithm. The time differences in the cluster with higher values are considered as separations between events. Graham et al [4] proposed a hierarchical time-based event segmentation algorithm. This approach first creates initial clusters between consecutive photographs. Then the initial clusters are split into finer clusters based on time difference. Finally, initial clusters are merged into higher levels by time difference or according to the year-month-day hierarchy. Event segmentation methods using content-based features normally group photos according to similarity. For example, Cooper et al [2] developed a segmentation algorithm based on similarity of discrete cosine transform coefficients. In addition to time-based and content-based some methods combine both time and content-based information. Research in [15], [13] also investigated grouping photos by locations.

Context based image retrieval has been an active topic over the last decade, with the term context being used in several senses. In the conventional cases, context-based search models specific user requirements in order to satisfy user information requirements. Context-based image retrieval methods are largely based on annotations that are manually added for disclosing the images using either keywords or descriptions or on collateral text that is accidentally available with an image (captions, subtitles, nearby text). From these texts, indexes are created using standard text retrieval techniques [6]. The similarity between images is then based on the similarity of the associated text which is often based on similarity between the associated texts which in turn is often based on similarity of word use. As an extension of these methods, several disambiguation techniques were also developed in order to extract the usage context within the

[7]EXIF is a specification for the image file format used by digital cameras. The information stored includes the date of digital content capture, GPS location if available, and manufacturing specifications of the camera

query. One such popular example, often used in disambiguation would be "jaguar", which could either mean a cat or the automobile. As the popularity and availability of the online thesaurus increases (for example, Wikipedia[8]) recent context disambiguation is able to exploit these resources in order to improve disambiguation [8].

With respect to social context, in [18], users are clustered to make possible a search engine that provides results tailored to users' interest moderated by the interests of their social groups. The authors argue that reordering of images based on implicit consensus of relevance between users in the same cluster was found to be useful. Therefore, taking social context of a user into account when trying to interpret their information need possibly improves relevant results. All of these approaches share a common goal with that of this paper, which is to segment photo streams into events. However, none of these earlier works studied segmenting photo streams into event chapters by combining camera parameters and visual information.

In comparison to other evolutionary computation techniques such as Genetic Algorithms (GA) [12], the advantages of PSO optimisation techniques include: (i) PSO does not suffer from some of GA's difficulties in interacting with the group members and often detracts from progress towards the solution. (ii) a particle swarm system has memory, which the genetic algorithm does not have. Change in genetic populations results in destruction of previous knowledge of the problem, except when elitism is employed, in which case usually one or a small number of individuals retain their identities. In PSO, individuals who bypass the optima are tugged to return towards them; and also the knowledge of good solutions are retained by all particles. A thorough evaluation of PSO and GA optimisation techniques has been presented in [1].

3. PROPOSED FRAMEWORK

John is a user in Facebook and is linked to 50 friends in facebook. As a tourist, his trips range from 1 day to 4 days. The average number of pictures he captures in any trip range from 50 to 250 depending on the interval. He creates a new album in Facebook after every trip and shares his experience with his friends. As a frequent traveller, he is not interested in individually tagging pictures after the creation of every album as it would consume a lot of time. Therefore, John would appreciate a system that would automatically create chapters and organise his album collection for his friends to quickly browse. Also, John is very interested in networking with his friends and likes to identify similar interests, in order to make travel plans according to common interests.

In Facebook, a set of albums are presented [9] as shown in Fig. 1. The "organise photos" option provides features to drag and drop images for sequential browsing of albums as shown in Figure 2. Although the feature makes it possible to organise the pictures in any order the user prefers, this also means that the user needs to manually shuffle individual pictures around in order to obtain a meaningful sequence. Also, the reordering of images from an album is static. Such a process could be time-consuming and would often discourage everyday users from using these features.

[8]http://www.wikipedia.org

[9]As Facebook has experienced an enormous growth, it has been taken as a reference in this paper

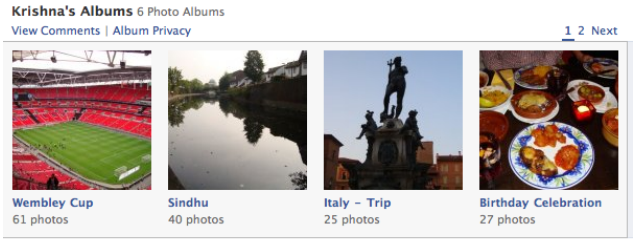

Figure 1: Albums organised and presented in Facebook

Figure 2: Facebook Album Organiser

The framework proposed below directly addresses the challenges raised in this use-case. First, event chapter segmentation of the albums is achieved using three types of metadata, based on EXIF, content and FOAF relationship. In addition, a retrieval model based on social context is proposed to identify similar interests by analysing the images uploaded by individual users. Depending on the image sequence, a metric for common interests is derived. In the following subsection the proposed framework is described in detail.

Addressing the above challenge, an event-based chapter segmentation framework as depicted in Fig. 3 is presented to enable efficient retrieval and/or organising the multimedia items from social networking sites. In the framework, the users on the top represent the content generators. The users capture content with the intention of sharing experience with their social peers. The content thus generated, often from visiting a place, would be published online using social networking websites such as Facebook. The user at the bottom of the framework is an every-day user who may share a direct relationship with the content generators. The proposed algorithms would directly benefit the end-user (represented at the bottom of the framework) for efficient and quick navigation of the created albums. Also, depending on the content created, a measure of relevance would be derived between the end-user and the content generator to represent common interests shared between them.

4. EVENT CHAPTER SEGMENTATION

In this section, a detailed description of different algorithms for achieving event chapter segmentation is precedented. These algorithms are mainly based on EXIF metadata and low-level MPEG-7 features.

4.1 Exchangeable image file format based Event Clustering

With the rapid progress of digital photography, even segmentation of photo streams has become one of the important and fundamental components in photo management systems. It provides semantically meaningful segments of the photo stream and therefore helps further management of the photos. For instance, it can help photo annotations

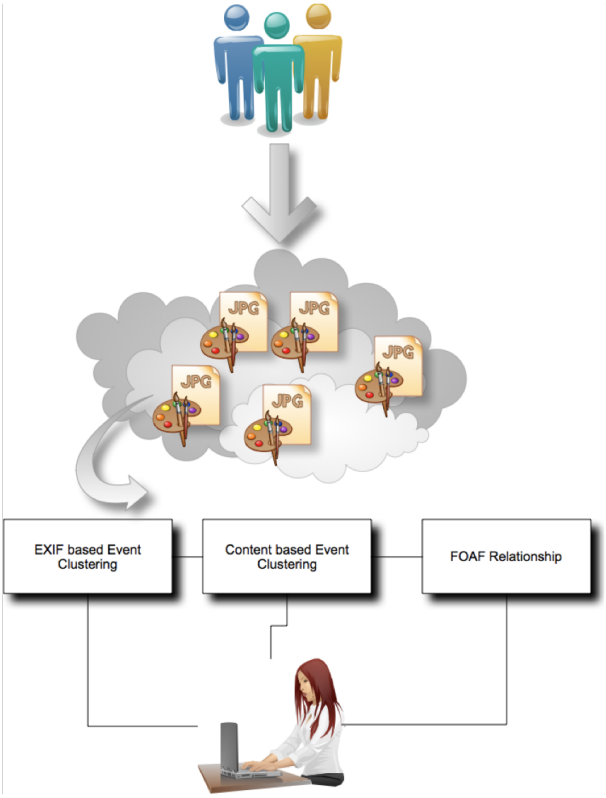

Figure 3: Framework representing the social content generation process along with ranking schemas

by providing event context information of the visual consistency within an event [23]. It can also help bulk annotation, where multiple photos are annotated at once, to reduce users' burden when making annotations [21]. Another photo management task that can benefit from event segmentation is photo retrieval. Recent research has shown evidence for exploiting event based photo navigation and browsing.

In order to distinguish our approach from other previous attempts, we define an "album" in terms of the abstract high-level participation of a group of people. Examples could include attending a 3-day rock concert, a vacation trip in summer, a ski trip in winter etc. These albums are temporally distinct and last for a considerable time. The amount of user-generated content during these trips is large and in many cases these multimedia items are largely unstructured. Therefore, the primary objective of EXIF-based clustering is to cluster or group these images into smaller chapters using "temporal" features. As reported in the related research section, hierarchical clustering of images based on EXIF leads to the problem of selecting a threshold that can cluster corresponding images into meaningful clusters. A low threshold could result in a large number of clusters and a high threshold could merge multiple clusters into one. In this paper, we address this problem by combining EXIF clustering with visual similarity obtained from content-based clustering. The results from the fusion of EXIF and visual similarity are presented in Section 7. In the following section, individual clustering techniques are discussed.

4.2 Content based Clustering

Visual coherence reflects semantic relationship between images and to this effect many machine learning algorithms have been developed to classify images according to their semantic relationship. As previously mentioned, UGC consists of special characteristics as opposed to content generated by professionals. Therefore, recently developed machine learning techniques based on biologically inspired algorithms are studied for visually clustering user generated albums. In this section, the clustering is initially performed by Self Organising Maps whose performance is improved using Particle Swarm Optimisation (PSO) technique. In the following, both techniques are briefly discussed with references to previous publication.

Particle Swarm Optimisation

In the PSO algorithm [3], the birds in a flock are symbolically represented as particles. These particles are considered to be "flying" through the problem space searching for the optimal solution [19]. A particle's location in the multidimensional problem space represents one solution for the problem. When a particle moves to a new location, a different solution to the problem is generated. This solution is evaluated by a fitness function that provides a quantitative value of the solution's utility. The velocity and position of each particle moving along each dimension of the problem space will be altered with each generation of movement. The particles at each time step are considered to be moving towards particle's personal best ($pbest$) and swarm's global best ($gbest$). The motion is attributed to the velocity and position of each particle. Acceleration (or velocity) is weighted with individual parameters governing the acceleration being generated for c_1 and c_2. The equations governing the velocity and position of each particle are presented in Equations 1 and 2.

$$v_{it}(t+1) = v_{id}(t) + c_1(pbest_i(t) - x_{id}(t)) + c_2(gbest_d(t) - x_{id}(t)) \quad (1)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (2)$$

- $v_{id(t)}$ represents the velocity of particle in dimension at time t
- $pbest_i(t)$ represents the personal best solution of particle i at time t
- $gbest_d(t)$ represents the global best solution for $d-$ dimension at time t
- $x_{id}(t)$ represents the position of the particle x in $d-$ dimension at time t
- c_1, c_2 constant parameters

The trajectory of each individual in the search space is adjusted by dynamically altering the velocity of each particle; according to the particle's own problem solving experience and the problem solving experience of other particles in the search space. The first part of Equation 1 represents the velocity at time t, which provides the necessary momentum for particles to move in the search space. During the initialization process, the term is set to '0' to symbolize that the particles begin the search process from rest. The second

part is known as the "cognitive component" and represents the personal memory of the individual particle. The third term in the equation is the "social component" of the swarm, which represents the collaborative effort of the particles in achieving the globally best solution. The social component always clusters the particles toward the global best solution determined at time t.

Self Organising Maps

The network architectures and signal processes used to model nervous systems can be categorised as feedforward, feedback and competitive. Feedforward networks [20] transform a set of input signals into a set of output signals. The desired input-output transformation is usually determined by external, supervised adjustment of the system parameters. In feedback networks [5], the input information defines the initial activity state of the feedback system, and after state transitions the asymptotic final state is identified as the outcome of the computation. In competitive learning networks, neighbouring cells in a neural network compete in their activities by means of mutual lateral interactions and develop adaptively into specific detectors of different signal patterns.

The basic idea underlying "competitive learning" is briefly presented here: Assume a sequence of statistical samples of a vectorial observable $x = s(t)$ where t is the time coordinate and a set of variable reference vectors $m_i(t) : m_i, \quad i = 1, 2, ..., k$. Assume that the $m_i(0)$ have been initialised in some proper way such as random initialization. If $x(t)$ can be simultaneously compared with each $m_i(t)$ at each successive instant of time, taken here to be integer $t = 1, 2, 3...$, then the best matching $m_i(t)$ is to be updated to match even more closely the current $x(t)$. If the comparison is based on some distance measure $d(x, m_i)$ altering m_i must be such that if $i = c$ is the index of the best-matching reference vector, then $d(x, m_c)$ is decreased, and all the other reference vectors m_i with $i \neq c$ are left intact. In this way, the different reference vectors tend to become specifically "tuned" to different domains of the input variable x.

In competitive neural networks, active neurons reinforce their neighbourhood within certain regions, while suppressing the activities of other neurons [22]. This is called on-centre/off-surround competition. The objective of SOM is to represent high-dimensional input patterns with prototype vectors that can be visualized in a usually two-dimensional lattice structure [9]. Each unit in the lattice is called a neuron, and adjacent neurons are connected to each other which results in a clear topology of how the network fits itself to the input space. Input patterns are fully connected to all neurons via adaptable weights, and during the training process, neighbouring input patterns are projected into the lattice, corresponding to the adjacent neurons. SOM enjoys the merit of input space density approximation and independence of the order to input patterns.

In the basic SOM training algorithm the input training vectors are trained with Equation 3

$$m_n(t+1) = m_n(t) + g_{cn}(t)[x - m_n(t)] \tag{3}$$

where m is the weight of the neurons in the SOM network, $g_{cn}(t)$ is the neighbourhood function that is defined as in Equation 4,

$$g_{cn}(t) = \alpha(t) exp(\frac{||r_c - r_i||^2}{2\alpha^2(t)}) \tag{4}$$

where, $\alpha(t)$ is the monotonically decreasing learning rate and r represents the position of the corresponding neuron.

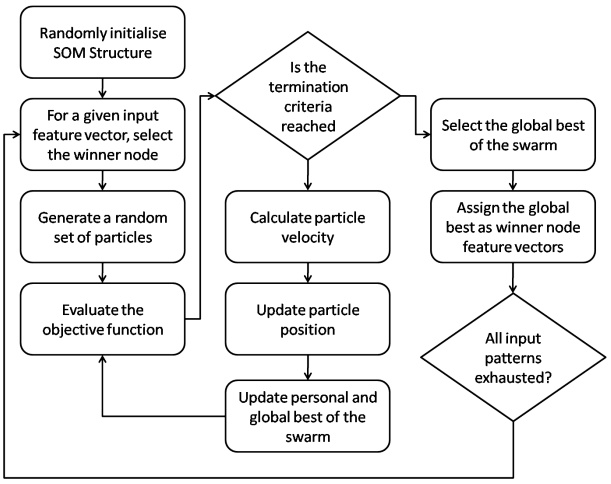

Figure 4: Pseudocode of the clustering algorithm

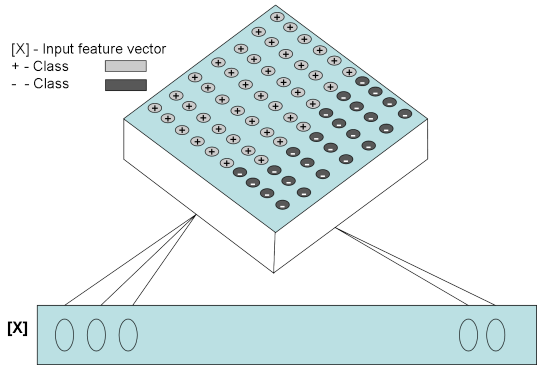

Figure 5: PSO Trained Network Structure of SOM

To further improve the performance of the SOM classifier, the weight of the neurons m_d is optimized with PSO. The pseudo code of the improved algorithm is presented in Fig. 4 followed by the completely trained network in Fig. 5. The optimisation is achieved by evaluating the $L1$ norm between the input feature vector and the feature vector of the winner node. The global best solution obtained after the termination of the PSO algorithm is assigned as the feature vector of the winner node. The training process is repeated until all the input training patterns are exhausted. In the testing phase, the distance between the input feature vector is compared against the trained nodes of the network. The label associated with the node is assigned to the input feature vector.

5. FOAF CLUSTERING AND NAVIGATION

In Facebook an event definition could be created as shown in Fig. 6. For each of such events, a list of attendees could be added in addition to the following multimedia content added as shown in Fig. 7. Ranking this socially generated user-generated content involves fusing multimedia items from different users into a single time line using EXIF and applying visual similarity measures to refine the initial clusters. In Section 7 a brief review of different fusion mechanisms and results obtained is discussed.

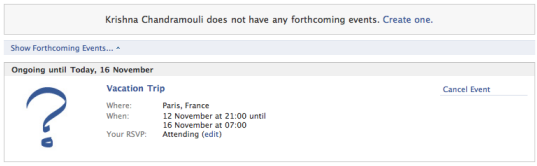

Figure 6: Event definition in Facebook

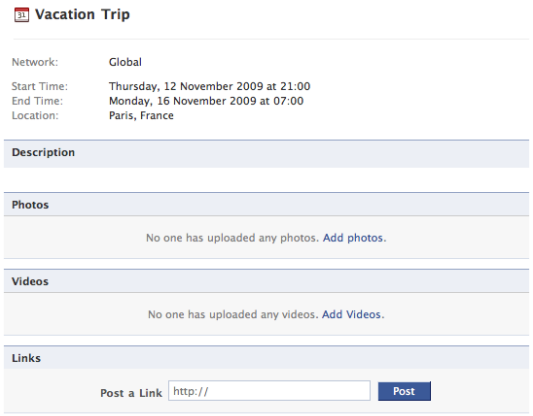

Figure 7: Attributes of event in Facebook

6. RETRIEVAL OF EVENT CHAPTERS

Following the event chapter segmentation of the albums, the next step would be to retrieve relevant event chapters from one's social peer(s). In order to effectively achieve retrieval of chapter events, an image retrieval system with relevance feedback has been implemented. In order to achieve real-time interaction of the retrieval system with the user, the image collections are stored in a separate database as opposed to stored centrally in Facebook. The retrieval model is also based on the PSO-based classifier as previously described. However, instead of clustering visual coherent images to segment an event from the album, the algorithm is used to rank the events extracted from the album. The algorithm considers a set of events which are previously extracted from different albums. The user selects events that are of interest to him/her and a visual search across social peers' albums is carried out to highlight similar or interesting events from social peers albums. The evaluation of this system is discussed in the next section 7.

7. EVALUATION RESULTS

One of the biggest challenges in evaluating algorithms for segmenting and retrieval of event chapters from social networking sites has been the availability of content. As the content is captured by individual users who wish to share among a group of trusted social peers and are often reluctant to share it with non-trusted sources, the dataset could be not be made publicly available. Therefore the comparison of the proposed techniques has been evaluated against other methodologies used in the literature (both on the individual level and on the fusion metrics). In the following, a set of evaluation criteria is defined for evaluating the framework.

The dataset used in the experiments was collected from three users subscribed to Facebook and related to each other through friendship. The collected dataset consists of 4 albums taken over a period of time from 3 different users. Each of these 4 albums contains on average more than 150 images captured over 3 to 14 days. These images were obtained as a part of tourist activities in which different users visited different places in Europe. The ground truth of the dataset for chapter segmetation was manually annotated by content creators. The criterion used for the ground truth is that an event chapter should consist of a temporal correlation between events. The ground truth thus obtained from different users has been stored in a metadata repository and further used for evaluating the proposed techniques. In Table 1, a brief description of 4 albums from which the pictures were acquired is presented. In Fig. 8, an overview of the dataset from different albums is presented.

Table 1: Album description from Facebook

Album	Location	Duration	Members	Images
1	London	14 days	3	80
2	Bologna	4 days	2	120
3	Paris-Brussels	3 days	2	218
4	Geneva	3 days	2	170

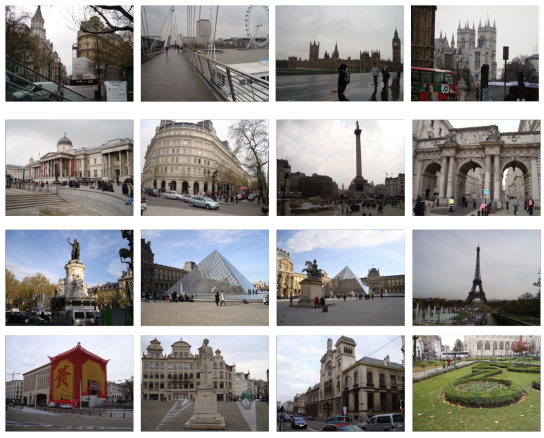

Figure 8: Overivew of Dataset

7.1 EXIF timeline based Clustering

As previously mentioned in the paper 4.1, time is used to cluster images according to events and in Fig. 9 results

of four albums with different temporal segments are presented. According to Tab. 1, the time span of Album 1 is considerably longer than the rest of the Albums (2, 3 and 4) and therefore increasing the threshold leads to an improved event chapter segmentation. However, different event chapters are of different lengths and therefore still contribute towards error. On the other hand, Album 3 generally shows a decreasing performance of chapter segmentation with increasing threshold. This is evidently due to the shorter duration of the event. On the contrary, the results of Album 4 are particularly interesting with respect to the chapter segmentation, the reason being, here the users have spent varying amounts of time at different parts of the location depending on personal interest and also the significance of the monument that was viewed. Another interesting case that could be detected is with Album 2 which was taken mostly in Bologna. One of the contributing factors could be that, being a relatively small city, users have spent most of the time taking pictures of neighbouring places. Therefore, in comparison with the ground truth, the chapter segmentation tends to get better by increasing the threshold value of the temporal segmentation. The feature vectors for the clustering are constructed by considering the temporal creation of the image with respect to the Album reference (the first picture captured in the Album) and the temporal distance between the photos.

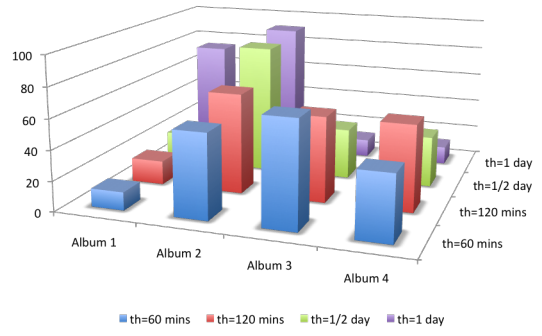

Figure 9: Event Segmentation based on EXIF time features(Accuracy, y-axis and Album, x-axis)

7.2 Visual clustering techniques

The visual clusters are generated using low-level MPEG-7 features, Colour Layout Descriptor (CLD) and Edge Histogram Descriptor (EHD). The clustering algorithm is based on PSO-enhanced SOM network as discussed in Section 4.2. The experimental results are presented in Figure 10. As it can be noted the PSO-based clustering algorithm outperforms the classical K-Means algorithm [22]. In order to account for the random initialisation of the K-Means centroids, the results presented are averaged over five different centroid selections against the the PSO algorithm. The performance of PSO-based clustering algorithm has shown to provide better clustering accuracy for all albums except Album 2. In the next section, the experimental evaluation of the combination of EXIF with visual clustering results is presented in order to account for temporal and visual changes when determining the cluster boundaries.

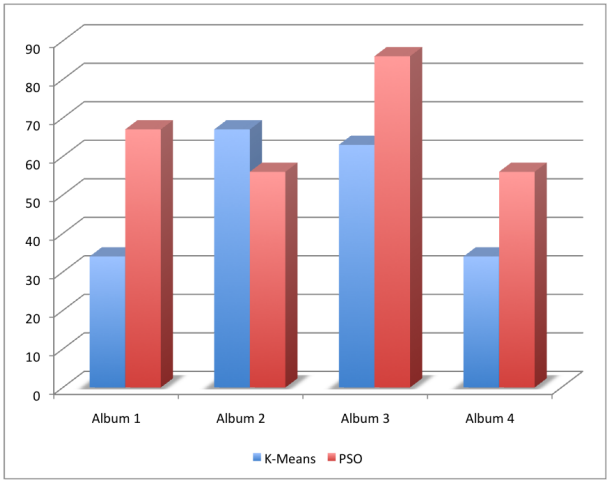

Figure 10: Event Segmentation based on Visual features (Accuracy, y-axis vs Album, x-axis)

7.3 Fusion of clustering techniques

The sources of decreased performance from EXIF and visual clustering techniques could be attributed to either over-segmenting or under-segmenting and disagreement over the cluster assignment for a chapter event. Therefore, in this section the performance of a fusion of EXIF and visual similarities is presented. In order to carry out a fair evaluation, there are many different scenarios that could be considered. As presented in previous evaluation sections, different thresholds for the EXIF clustering technique result in different performance measures and similarly combining all these multiple sets of results will result in a wide range of evaluation results. After a thorough evaluation, the best results are obtained for all Albums when EXIF-based over-segmentation of chapter results is refined with clustering results obtained from visual similarity measures as shown in Figure 11. An average of 20% accuracy increase has been achieved using this schema.

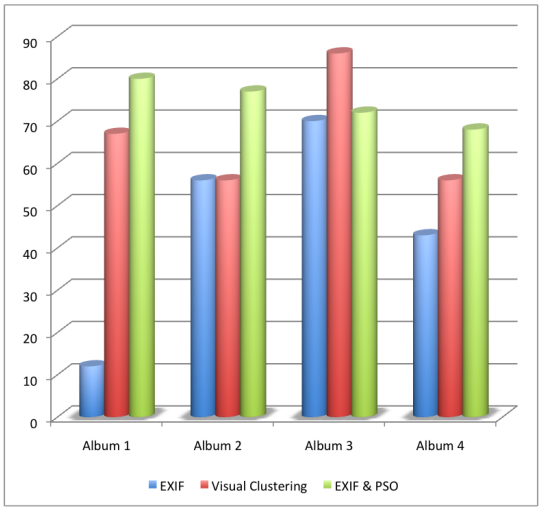

Figure 11: Cluster fusion results (Accuracy, y-axis vs Album, x-axis)

7.4 Retrieval of Event Chapters

In Fig. 12, the evaluation results of the event retrieval system is presented. The evaluation was carried out with three different users selecting an event chapter which is of interest. The retrieval system employs visual similarity measures on the query provided by the user to rank different events from the database. The recall measure is used as the evaluation metric for the system. The results shown in Fig. 12 indicates that the performance of the retrieval system for a set of queries enables users to identify other users with similar interests. As the dataset used in this evaluation consists of three users and four albums, the evaluation of a common interest metric does not reflect true interest sharing among users. However, we believe on a larger dataset collection with more than 50 users the metric evaluation could be suitably justified.

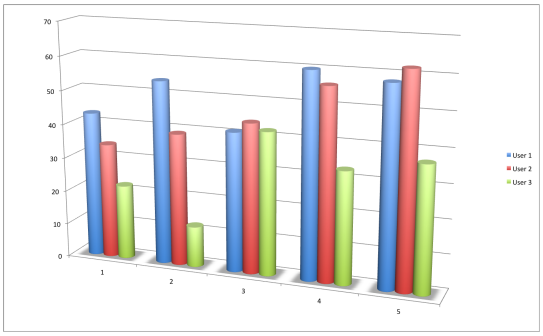

Figure 12: Event Retrieval from the System

8. CONCLUSION AND FUTURE WORK

In this paper, we presented techniques for segmenting event chapters from multimedia collections to facilitate user navigation and retrieval through socially networked websites. The proposed approach is based on merging EXIF metadata with visual similarity measures in order to obtain a meaningful clustering of user-generated content such as is commonly a set of four albums is published on social networking sites. The evaluation of the proposed techniques has been performed on a set of 4 albums created from different users while travelling through Europe for vacation. As a common platform for evaluating social networking techniques is not available, the proposed techniques have been evaluated against conventional multimedia techniques based on accuracy.

Future work will focus on studying FOAF-based navigation and merging content from different users in order to create a personalised remembrance of a specific event. Such events could range from a simple personal event to a collaborative event with participation from many different users. In addition, the analysis of location information for improving the performance of the event event chapter segmentation will be further studied.

9. ACKNOWLEDGMENTS

The research was partially supported by the European Commission under contract FP7-216444 PetaMedia.

10. REFERENCES

[1] K. Chandramouli. *Image Classification using Particle Swarm Optimisation*. PhD thesis, Queen Mary, University of London, January 2009.

[2] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox. Temporal event clustering for digital photo collections. In *Proc. of ACM Multimedia*, pages 364–373, 2003.

[3] R. Eberhart and Y. Shi. Tracking and optimizing dynamic systems with particle swarms. *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, 1, 2001.

[4] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as the essence for photo browsing through personal digital libraries. In *Proc. of ACM/IEEE Conference on Digital Libraries*, pages 326–335, 2002.

[5] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.*, 79:2554–2558, 1982.

[6] M. Inoue. Image retrieval: Research and use in the information explosion. *Progress in Informations*, 6:3–14, 2009.

[7] J. Kennedy and R. C. Eberhart. *Swarm intelligence*. Morgan Kaufmann, 2001.

[8] T. Kliegr, K. Chandramouli, J. Nemrava, V. Svátek, and E. Izquierdo. Combining captions and visual analysis for image concept classification. In *MDM/KDD'08: Proceedings of the 9h International Workshop on Multimedia Data Mining*. ACM, 2008.

[9] T. Kohonen. The self organizing map. *Proceedings of IEEE*, 78(4):1464–1480, September 1990.

[10] A. Loui and A. Savakis. Automatic image event segmentation and quality screening for albuming applications. In *Proc. of IEEE Conference on Multimedia and Expo*, pages 1125–1128, 2000.

[11] T. Mei, B. Wang, X. S. Hua, H. Q. Zhou, and S. Li. Probabilistic multimodality fusion for event based home photo clustering. *IEEE International Conference on Multimedia and Expo*, pages 1757–1760, 2006.

[12] M. Mitchell. *An introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, 1996.

[13] M. Naaman, Y. J. Song, A. Paepcke, and H. G. Molina. Automatic organisation for digital photographs with geographic coordinates. In *Proc. of ACM/IEEE Conference on Digital Libraries*, pages 141–150, 2004.

[14] Picasa. http://picasa.google.com.

[15] A. Pigeau and M. Gelgon. Building and tracking hierarchical geographicsl and temporal partitions for image collection management on mobile devices. In *Proc. of ACM Multimedia*, pages 141–150, 2005.

[16] J. Platt. Autoalbum: Clustering digital photographs using probabilistic model merging. In *Proc. of IEEE Workshop on content-based Access of Image and Video*, pages 96–100, 2000.

[17] J. C. Platt, M. Czerwinski, and B. Field. Phototoc: Automatic clustering for browsing personal photographs. In *Technical Report*, 2003.

[18] A. Rae. Social context aiding image search. In *Poster*, 2008.

[19] C. Reynolds. Flocks, herds and schools: a distributed

behavioural model. In *Computer Graphics*, pages 25–34, 1987.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1:318–362, 1986.

[21] B. Shu and B. Bederson. Semi-automatic image annotation using event and torso identification. *Technical report, Computer Science Department, University of Maryland, College Park MD*, 2004.

[22] R. Xu and D. W. II. Survey of clustering algorithms. *IEEE Trans. Neural Network*, 6(3):645–678.

[23] M. Zhao, Y. Teo, T. C. S. Liu, and R. Jain. Automatic person annotation of family photo album. In *International Conference on Image and Video Retrieval*, pages 163–172, 2006.