Epitome – A Social Game for Photo Album Summarization

Ivan Ivanov, Peter Vajda, Jong-Seok Lee, Touradj Ebrahimi
Multimedia Signal Processing Group – MMSPG
Institute of Electrical Engineering – IEL
Ecole Polytechnique Fédérale de Lausanne – EPFL
CH-1015 Lausanne, Switzerland
{ivan.ivanov, peter.vajda, jong-seok.lee, touradj.ebrahimi}@epfl.ch

ABSTRACT

In this paper, we propose an approach for photo album summarization through a novel social game "Epitome" for mobile phones. Our approach to album summarization consists of two games: "Select the Best!" and "Split it!". The goal of the first game is to allow a user to select the most representative photo of a reduced set of images, while in the second game, the user has to split the reduced set into two distinct parts. As it could be time-consuming to look at a huge collection of photos on a mobile phone, it is more enjoyable and pleasant to show only a limited number of images which can be fit into one mobile screen. The results obtained in these games are combined to produce a summarization and are then compared with the results of other users. As a final result, a unique summarization sequence of photos is determined. The determined sequence of photos can be used to create a collage of one album or a cover for an album. The proof of concept of the proposed method is demonstrated through a set of experiments on several photo collections.

Categories and Subject Descriptors

I.4.9 [**Image Processing and Computer Vision**]: Applications; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*Web-based interaction*

General Terms

Design, Experimentation, Performance

Keywords

social game, social networks, photo summarization, collage, mobile game

1. INTRODUCTION

The popularity of mobile phones equipped with digital cameras in recent years has much increased the number of photos taken in our daily lives. In order to make these photos easily accessible, people usually organize their photos in albums (collections) based on events, places, or people. Besides storing them on computer hard drives, people also transfer their digital photos to social networks where their friends, family and colleagues can also access them. Some do not stop there, and print their photos on post cards, calendars or photo books, often to give them as presents or to create physical souvenirs. According to a recent study [3], 40% of adults who use Internet, upload photos to websites to share with others online, while 34% of adults print their photos as photo books. Photo sharing web sites contain a huge volume of publicly available photos. For example, Flickr[1] contains over 4 billion photos [4] and more than 2.5 billion photos are uploaded to Facebook[2] each month [1].

There is a saying: "A picture is worth a thousand words." Therefore, people like to use their photos to tell their own stories of some important events in their life. One's wedding, birth of a baby, vacation, birthday party or even a long lasting period - from the date of one's birth till celebration of 18th birthday, are only a few examples of such events. One of the reasons why people share photos is to ask their friends to comment and tag photos. Summarization is an effective way to help getting a quick overview of a set of photos. These photos can be used to create a collage of one album, a cover for an album, or to be included in a photo book.

Which photos are the most suitable to summarize a photo album? Creation of a photo summary is a very subjective task. There are different criteria upon which a human user would rate digital photos. The color, composition, content, lighting and sharpness of a photo, all contribute to viewer's response to that photo. These characteristics are used extensively by professionals on web sites, magazine covers and printed advertisements to draw attention, communicate a message and leave a lasting emotional impression. Summarization is currently often done automatically, with obvious limitations. There is a gap between what people think the summary should look like and what we get with an automatic summarization. While watching shared photos, it would be desirable if your friends can also select the most representative photos which summarize an album.

In this paper, we propose an approach for photo album summarization through a novel social game "Epitome". The name "Epitome" comes from the Latin word for summarization. We formulate the problem of album summarization as

CMM'10, October 29, 2010, Firenze, Italy.

[1] http://www.flickr.com
[2] http://www.facebook.com

selecting a set of photos from a larger collection which best preserves the visual information of the entire collection. The ideal summarization presents the most interesting and important aspects of the scenes in the album with minimal redundancy. The main idea of our approach is to show a reduced set of photos on a mobile phone, ask users to play the game and then integrate results of all users in order to produce a summarization for the whole dataset. Our approach to album summarization involves two games denoted in this paper as "Select the Best!" and "Split it!". Each of these social games starts with a collection of photos, for example, taken during a vacation. The goal of the first game is to allow a user to select the most representative photo of a reduced set of images. In the second game, the user has to split this reduced set into two distinct parts in order to mimic separation of one album into different events. The results achieved in the two games are compared with those of other users, and every user receives a score based on his/her performance. A unique summarization sequence of photos is determined as a sequence of photos which gets the highest number of users' votes. The determined sequence of photos is used as the most appealing photos which best represent the original collection. The proof of concept of the proposed method is demonstrated through a set of experiments on several photo collections.

The remaining sections of this paper are organized as follows. We introduce related work in Section 2. Section 3 describes our social game application and its implementation on a mobile phone. Experiments and results are shown in Section 4. Finally, Section 5 concludes the paper with a summary and some perspectives for future work.

2. RELATED WORK

Most mobile game applications are stand alone client based games. However, the network is becoming cheaper and faster, which facilitates fast spread of online game applications. Our online game is designed for this new segment of mobile game applications.

There are already several kinds of photo collection browsers on the market. However, difficulties appear for large personal photo collections. State-of-the-art navigation hierarchy considers time separated events, spatial information using GPS coordinates, and content-based image similarities. Susumu *et al.* [9] developed an interface for photo navigation that is based on a zoomable timeline. The basic idea is to cluster images using the time difference between two consecutive photos. Initial clusters are created by detecting gaps of more than 24 hours, which can determine different events. Further clustering is applied recursively on each cluster regarding to outliers compared to the mean time gap distance between consecutive images. The result is an automatic personal photo structuring and navigation interface for PDAs. Cooper *et al.* [7] explored the temporal photo collection clustering, extending with content-based image similarity measurement. Naaman *et al.* [10] developed a system called PhotoCompas. This system does automatic organization of digital photographs, which additionally considers the geographic location of photo or event based description extracted from user tags. Spatial and temporal clustering therefore can be applied for photo search and navigation. Furthermore, search categories can be applied for navigation, such as elevation, season, time of the day, weather status, temperature and time zone. Finally, combination of temporal similarity, content-based similarity, and location-based similarity is used for photo collection clustering. Hewlett-Packard is highly involved in commercial side of image processing for photo clustering. One of their research areas is automatic generation of photo collection page layout. In [6], they presented a photo collection page layout generation method that attempts to maximize page coverage without having overlapping photos. Layout is based on a hierarchical partition of the page, which provides explicit control over the aspect ratios and relative areas of photos. Geigel and Loui [8] emphasized aesthetic side of a page layout for image collections. They used a genetic algorithm to optimize aspects such as balance and symmetry for a good placement of images in the personalized album pages.

Gaming provides a new way of motivating people to make the subjective data acquisition interesting and enjoyable. Ames and Naaman [5] have explored different factors that motivate people to tag photos in mobile and online environments. One way is to decrease the complexity of the tagging process through tag recommendation which derives a set of possible tags from which the user can select suitable ones. Another approach is to provide incentives to the user in form of entertainment or rewards, e.g. games. The most famous examples of games which are based on the latter approach are the ESP Game and Peekaboom, developed for collecting information about image content. The *ESP Game* [11] randomly matches two players who are not allowed to communicate with each other. They are shown the same image and asked to enter a textual label that describes it. The aim of each user is to enter the same word as his/her partner in the shortest possible time. *Peekaboom* [12] extends the ESP Game. Unlike the ESP Game, it's asymmetrical. To start, one player is shown an image and the other sees an empty black space. The first user is given a word related to the image, and the aim is to communicate that word to the other player by revealing portions of the image. Peekaboom improves the data information, collected by the ESP Game, by asking for the object location in the image. Several other games have been created based on this idea, such as video tagging, music description and tagging, tag description, object segmentation, visual preference and image similarities.

Our social game can collect research data and, at the same time, it provides a collage or a cover photo of the photo albums, while, at the same time, the user enjoys playing the game. In this way, both users and research community can benefit. Therefore, this concept is novel compared to previously mentioned games.

3. APPLICATION

The goal of our application is to provide an intuitive and enjoyable user interface for mobile phones, which creates and annotates photo collages for Facebook photo albums. Therefore, a game is created, which can provide its potential users with many pleasant hours while playing it, and enjoying photos. At the same time, it determines the most representative photos of the user's photo album and provides useful research data.

3.1 Epitome game

The scenario of the game is as follows. A player logs in to the game with his/her Facebook account and allows access to his/her photo gallery in order to force all players to also

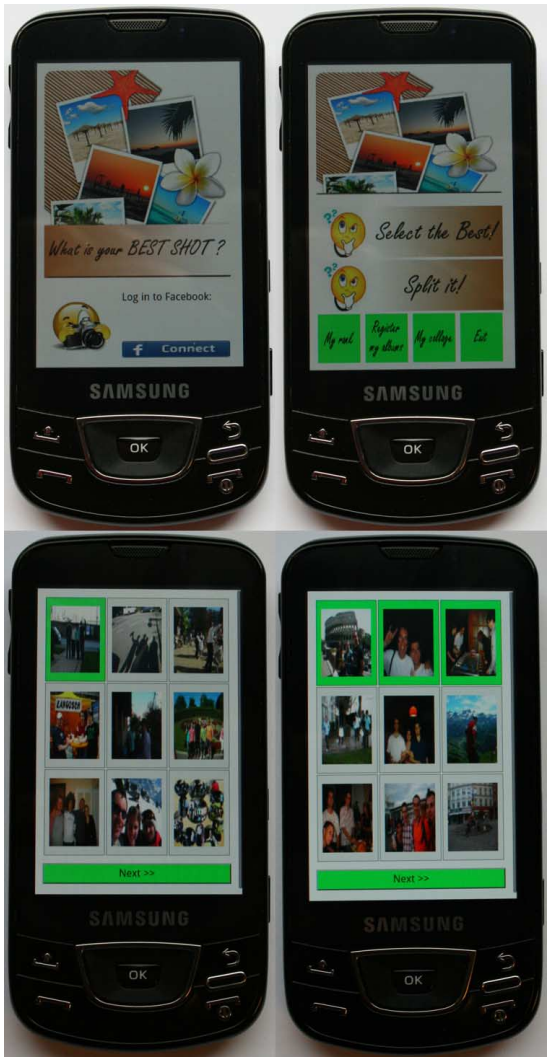

Figure 1: Login, Main game selection, "Select the Best!" and "Split it!" screenshots of the games are shown.

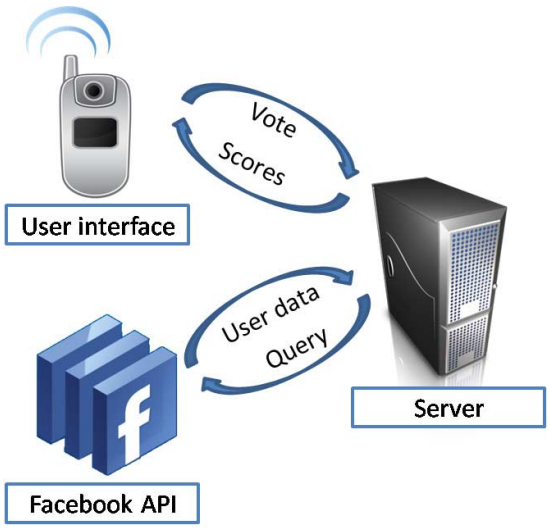

Figure 2: System architecture of the "Epitome" mobile game. It consists of three parts: the *mobile side* which deals with the user interface, the *server side* which performs the analysis, and the *Facebook* which provides necessary data and authentication.

populate the game with their own photos, as shown in Figure 1. For new users who log in for the first time, the server registers the player's photo albums in the database. At this point, the player can select between two games. In both games, 9 consecutive photos are randomly selected from one of the Facebook albums. This feature of showing only the partial album is commonly used, for example in Facebook, where one album is split into several pages, and photos on the same page are placed in a grid. In the first game, called "Select the Best!", 9 images are shown to the player and he/she has to choose the best representative photo. By clicking on one of these 9 images, the player can see that image in a resolution that fits into the entire mobile screen. If the player chooses a photo which is the most frequently selected by other players, then the player's score increases and an acoustic tone is heard. Since albums are chosen at random, it may happen that a new album without any subjective data appears in the game. In that case, the player automatically wins and his/her score increases. The second game is called "Split it!", where the player should split images into two parts.

If a player reaches a certain score level, then the server shows photos for the collage of his/her photo album from the results of "Select the Best!" and "Split it!" games. Therefore, the player can get a feedback from all other players, regarding his/her Facebook photo albums.

The application (Figure 2) consists of three parts, the *mobile side* which deals with the user interface, the *server side* which performs the analysis and the *Facebook* which provides user data, images and authentication. These three parts communicate over an HTTP network using JSON data structure. The mobile side application is a weak interface for server information, where most of the processing is done on the server, therefore it makes our game easily portable to different client platforms. Authentication and photo albums are handled by Facebook through Facebook API. Finally, the scores, the information about photo albums and the results are stored on the external server side. The game was tested with Samsung Galaxy *i*7500 mobile phone on Android OS platform.

3.2 Scoring

The application calculates three different values: *Importance*, *Segmentation* and $UserScore$.

Importance value is calculated in the "Select the Best!" game for each photo album separately. The goal of this game is to select the most representative photo of the particular Facebook album of $K = 9$ photos given the fact that the players can select only one representative photo among K photos. These K consecutive photos are chosen randomly from the album every time a user plays the game. A feature vector $BestSmall_n$, $n \in [1, N]$, is calculated for each user, n among N users, as follows:

$$BestSmall_n = [\ \alpha_{n,1},\quad \alpha_{n,2},\quad \alpha_{n,3},\quad \ldots,\quad \alpha_{n,K}\ ], \quad (1)$$

where $\alpha_{n,k} \in \{0,1\}$, for $k \in [1,K]$, takes either 1 or 0 depending on whether the corresponding photo is chosen as the most representative photo. This vector is then extended to a vector $Best_n$ of dimension M, where M is size of the particular Facebook album, as follows:

$$Best_n = [\ \underbrace{0, \ \ldots, \ 0,}_{y-1}\ \underbrace{BestSmall_n,}_{K}\ \underbrace{0, \ \ldots, \ 0,}_{M-K-y+1}\ ], \quad (2)$$

where $y \in [1, M-K+1]$ is the index of the first photo shown to the player in the corresponding album. The frequency of all photos that appear in the game is stored as a vector $BestFreq$ of dimension M, because photos are shown in random order and thus have different impact on the final results. An M-dimensional vector $BestCount$ is then calculated as: $BestCount = \sum_n Best_n$, $n \in [1,N]$. At the end, a normalized vector $Importance$ is formed by element-wise division:

$$Importance = \frac{BestCount}{BestFreq}, \quad (3)$$

which is an M-dimensional vector showing the distribution of the most representative photos within one Facebook album.

Segmentation value is calculated in "Split it!" game for each photo album separately and updated every time a user plays it. It shows the frequency with which each photo in one album is selected as a starting photo in a new segment. In this game, players are asked to perform partitioning of $K = 9$ consecutive photos from a Facebook album into two distinct parts. These K consecutive photos are chosen randomly from the album every time a user plays the game. For each player n, a feature vector $SegmSmall_n$, $n \in [1,N]$, is formed as follows:

$$SegmSmall_n = [\ \beta_{n,1}, \ \ \beta_{n,2}, \ \ \beta_{n,3}, \ \ \ldots, \ \ \beta_{n,K}\ ], \quad (4)$$

where $\beta_{n,k} \in \{0,1\}$, for $k \in [1,K]$, takes either 1 or 0 depending on whether the corresponding photo is chosen as the beginning of the new segment. Note that the vector always has only one value 1. This vector $SegmSmall_n$ is then used to form an extended vector $Segm_n$ of dimension M equal to the size of the particular Facebook album, as follows:

$$Segm_n = [\ \underbrace{0, \ \ldots, \ 0,}_{x-1}\ \underbrace{SegmSmall_n,}_{K}\ \underbrace{0, \ \ldots, \ 0,}_{M-K-x+1}\ ], \quad (5)$$

where $x \in [1, M-K+1]$ is the index of the first photo shown to the player in the corresponding album. Every time the photos are shown to the player, they are counted and M-dimensional frequency vector $SegmFreq$ is calculated for the whole album which stores information about how many times each photo appeared in the game. All vectors $Segm_n$, $n \in [1,N]$, are then summed up to form an M-dimensional vector $SegmCount = \sum_n Segm_n$. As a final result, $SegmCount$ is normalized with $SegmFreq$ to take into account the frequency with which each of the photos appears in the game, as the photos in the middle of the album appear more frequently. The final result used in the further evaluation is the following ratio in which element-wise division is performed:

$$Segmentation = \frac{SegmCount}{SegmFreq}. \quad (6)$$

This is an M-dimensional vector which shows the distribution of the start indexes of the segments within a photo album.

Finally, the results obtained in these games are combined to produce a summarization. Vectors $Importance$ and $Segmentation$ are used to automatically select $L = 5$ most representative photos within the dataset. At first, $L-1$ maximum values are selected from the vector $Segmentation$. In this way, the particular album is segmented into L most probable segments chosen by N users. For each of these segments, a photo with the highest score in the vector $Importance$ is chosen. If there are multiple photos having the same score, one among them is randomly chosen. These L photos represent a collage of the album, which is shown to the owner of that album, if he/she reaches a certain level of $UserScore$.

UserScore value is defined to motivate players to play this game frequently. In the "Select the Best!" game, the player increases his/her own $UserScore$ if he/she selects the photo which has the highest $Importance$ value among 9 photos. The same approach is used in "Split it!" game, where the player increases his/her $UserScore$ if he/she separates 9 photos at the place where $Segmentation$ value is the highest among 9 photos. Initial $UserScore$ is set to 0.

4. EVALUATION

Creation of a photo summary is always a very subjective task, and thus the evaluation of a summary is difficult. We asked participants (users) to create a ground truth for 6 photo collections. The ground truth contains the most representative photos for the whole dataset (6 collections). In this section, the dataset used and experiments are described.

4.1 Datasets

The dataset used in our experiments is the official dataset from "HP Challenge 2010: High Impact Visual Communication" at the "Multimedia Grand Challenge 2010" [2]. It consists of 6 datasets, each with 20 photos. These datasets cover photos that are usually taken during a vacation, describing a variety of topics: photos depicting different landmarks and famous sightseeing places, photos with parents and kids, and photos of cars, flowers and sea animals. Figure 3 provides example photos of the datasets. Even though photos of each dataset are pinned under the same topic, we can see that their content is rather heterogeneous, presenting different objects or scenes (mainly outdoor scenes) with large variances in color representation, presence of people, etc.

4.2 Experiments

To collect the ground truth data and to evaluate the designed photo selection tool (social game), we conducted two experiments. Since there are different criteria upon which a human user would rate digital photos, we first constructed a ground truth by asking different people for their subjective opinion about photos and then tested our algorithm against the ground truth data. We recruited 63 participants, among whom 61% were males and 39% were females, aged $18-65$, with different backgrounds and cultural differences.

In the collection of the ground truth data, participants were shown 20 photos which belong to the same dataset (collection or album). The task of the participants was to

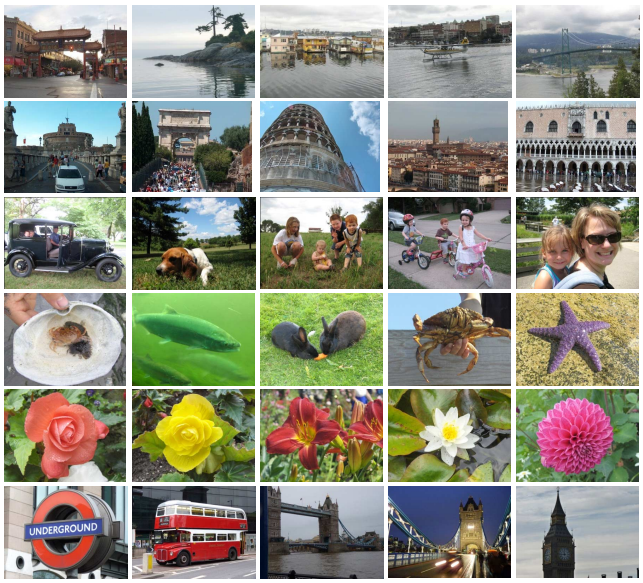

Figure 3: Some example photos for each of 6 datasets. Photos in each row belong to different datasets. The datasets cover a large variety of objects and scenes usually taken during a vacation.

select the 5 most representative photos of the whole album, while looking at all photos of that album.

The goal of experiment 1 was to perform a segmentation of the image collection by splitting it into two parts which have distinct semantic meanings, which is analogous to the game "Split it!". For example, the first part of the set can be photos taken in a down town, while the second part of the set can be represented by photos taken at a lake (e.g., the first row in Figure 3). Or, the first part of photos depicts red flowers and the second yellow flowers. In this experiment, the participants were shown only a reduced set of photos (9 out of 20 photos) which belong to the same collection. The photos were shown in the time order in which they were captured. The time stamp was extracted from EXIF tags associated to each photo. The subset of 9 photos in this experiment was extracted as follows: 20 photos were sorted in the time order as already explained, and then 9 consecutive photos were selected with a random starting position which was different for each user.

In experiment 2, the participants were shown a reduced set of photos (9 out of 20 photos) which belong to the same collection. These 9 photos were selected in the same way as already explained in experiment 1. The participants were asked to select only one, the most representative photo of that collection, which is analogous to the game "Select the Best!".

The results obtained in experiments 1 and 2 are used to assess the performance of our approach by comparing them with the ground truth.

4.3 Results and analysis

For simplicity of the explanation on how our approach was evaluated, let us consider only one dataset with $M = 20$ photos. First, a ground truth data is collected. Every user n among $N = 63$ users is asked to select the 5 most representative photos. After his/her participation in collecting the ground truth data, the corresponding feature vector $Full_n$, $n \in [1, N]$, is formed as follows:

$$Full_n = [\ \delta_{n,1},\quad \delta_{n,2},\quad \delta_{n,3},\quad \ldots,\quad \delta_{n,M}\ ], \tag{7}$$

where $\delta_{n,m} \in \{0, 1\}$, for $m \in [1, M]$, takes either 1 or 0 depending on whether the corresponding photo is chosen as one of the representative photos or not. Feature vectors of the users i and j, $i, j \in [1, N]$, are then compared to each other and the score of their matching $S_{i,j}$ is calculated as:

$$S_{i,j} = Full_i \cdot Full_j^{\mathrm{T}}. \tag{8}$$

In other words, the higher the number of identical photos that are chosen by two users, the better will be the score of the match between them. Note that the maximum score of the match is 5. Finally, to each user i, $i \in [1, N]$, a value $Score_i$ is assigned as:

$$Score_i = \sum_{j=1}^{N} S_{i,j}. \tag{9}$$

The maximum value in the vector $Score_i$ shows the best performing participant who has the highest number of selected photos which are matched with all other users. The maximum possible value of the score is $5 \cdot N$, which in our case becomes 315. These results are considered as the ground truth data and compared with the results from two other experiments in order to prove the concept of our approach.

In experiments 1 and 2, participants are asked to perform partitioning of the reduced dataset and to select the most representative photo of the dataset. The vectors $Importance$ and $Segmentation$ of dimension M, which are described in Section 3.2, are used to automatically select $L = 5$ the most representative photos within the dataset. These L photos are then represented as a choice of the proposed method. Then, the complete procedure of measuring similarity between the choice of the proposed method and all other users is repeated and the final scores are computed according to Equations 8 and 9.

All computations are repeated in a similar way for all 6 datasets.

The results obtained in our experiments are shown in Figure 4. This figure shows the distribution of the participants' scores, including the choice obtained by the proposed method. All scores are sorted in a descending order. The results of proposed method which is the integration of the game results show how good we are in selection of the most representative photos of one dataset by showing only a reduced set of images in the proposed game. These initial results look promising. As we can see, the scores of the proposed method have a small relative distance from the best ground truth scores achieved in our experiments. In average, our approach achieves 80% of the best score for each dataset, which proves the concept of our game. For datasets 3 and 5, this value is even higher, i.e. about 95%. The most representative photos for one of the datasets selected by the proposed method are shown in Figure 5.

5. CONCLUSION

With the rapid growth of digital photography, sharing of photos with friends and family has become very popular. When people share their photos, they usually organize them in albums according to events or places. To tell the story of some important events, it is desirable to have an efficient

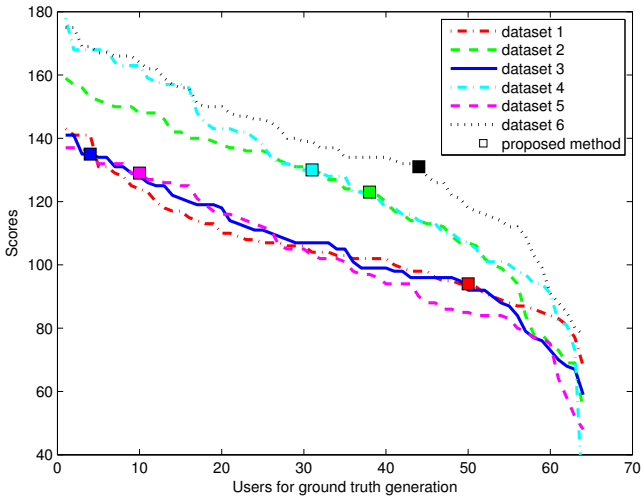

Figure 4: The distribution of the participants' scores. All scores are sorted in descending order. The results of the proposed method are shown with square markers. Different colors of the square markers correspond to different datasets. The results are promising and prove the concept of our approach.

summarization tool which can help people to get a quick overview of an album containing a huge number of photos.

In this paper, we proposed two social games for an album summarization on mobile phones: "Select the Best!" and "Split it!". In order to make them suitable for mobile phones, these games have to concentrate on small tasks: selection of the most representative photo of a reduced set of images and partitioning of the reduced set into two distinct parts. The proof of concept of these games was demonstrated through a set of experiments on several photo collections. The results of our experiments show that concept of the game is validated. In average, our summarization game achieves 80% of the best score of different participants.

As a future study, we will include in our approach different visual features and make the game more attractive for users.

6. ACKNOWLEDGMENTS

This work was supported by the Swiss National Foundation for Scientific Research in the framework of NCCR Interactive Multimodal Information Management (IM2), the Swiss National Science Foundation Grant "Multimedia Security" (number 200020-113709), and partially supported by the European Network of Excellence PetaMedia (FP7/2007-2011).

References

[1] FaceBook Statistics. Available at: `http://www.facebook.com/press/info.php?statistics`.

[2] HP Challenge 2010 Dataset: High Impact Visual Communication. Available at: `http://comminfo.rutgers.edu/conferences/mmchallenge/2010/02/10/hp-challenge-2010`.

[3] Statistics on Sharing of Photos. Available at: `http://iphonephotovideo.com/2009/05/some-interesting-mobile-phone-and-mobile-imaging-statistics`.

[4] Wikipedia - Flickr. Available at: `http://en.wikipedia.org/wiki/Flickr`.

[5] M. Ames and M. Naaman. Why we tag: Motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007)*, pages 971–980, 2007.

[6] B. Atkins. Adaptive photo collection page layout. In *Proceedings of the International Conference on Image Processing (ICIP 2004)*, pages 2897–2900, 2004.

[7] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox. Temporal event clustering for digital photo collections. *ACM Transactions on Multimedia Computing, Communications and Applications*, 1(3):269–288, 2005.

[8] J. Geigel and A. Loui. Using genetic algorithms for album page layouts. *IEEE Multimedia*, 10(4):16–27, 2003.

[9] S. Harada, M. Naaman, Y. J. Song, Q. Wang, and A. Paepcke. Lost in memories: Interacting with large photo collections on PDAs. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2004)*, pages 325–333, 2004.

[10] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2004)*, pages 53–62, 2004.

[11] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*, pages 319–326, 2004.

[12] L. von Ahn, R. Liu, and M. Blum. Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*, pages 55–64, 2006.

Figure 5: Photos from the dataset 3. The most representative photos selected by the proposed method are marked with green bounding box.