Hindawi Publishing Corporation
International Journal of Computer Games Technology
Volume 2011, Article ID 706893, 11 pages
doi:10.1155/2011/706893

Research Article

Epitomize Your Photos

Peter Vajda,[1] Ivan Ivanov,[1] Jong-Seok Lee,[2] and Touradj Ebrahimi[1]

[1] *Multimedia Signal Processing Group (MMSPG), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland*
[2] *School of Integrated Technology, Yonsei University, Incheon 406-840, Republic of Korea*

Correspondence should be addressed to Peter Vajda, peter.vajda@epfl.ch

Received 1 September 2011; Accepted 31 December 2011

Academic Editor: Mark Green

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 into albums according to events or places. To tell the story of some important events in one's life, it is desirable to have an efficient summarization tool which can help people to receive a quick overview of an album containing large number of photos. In this paper, we present and analyze an approach for photo album summarization through a novel social game "Epitome" as a Facebook application. This social game can collect research data, and, at the same time, it provides a collage or a cover photo of the user's photo album, while the user enjoys playing the game. The proof of concept of the proposed method is demonstrated through a set of experiments on several photo albums. As a benchmark comparison to this game, we perform automatic visual analysis considering several state-of-the-art features. We also evaluate the usability of the game by making use of a questionnaire on several subjects who played the "Epitome" game. Furthermore, we address privacy issues concerning shared photos in Facebook applications.

1. Introduction

Rapid growth of digital photography in recent years has increased the size of personal photo collections. People use their digital cameras or mobile phones equipped with cameras to take photos. Besides storing them on computer hard drives, they often share their digital photos with friends, family, and colleagues through social networks. Facebook (http://www.facebook.com/), Flickr (http://www.flickr.com/), and Picasa (http://picasa.google.com/) are examples of such photo sharing web sites. Some people also print their photos on post cards, calendars, or photo books, often to give them as presents or to create physical souvenirs.

Photos are often organized into albums (collections) based on places, events or dates, and people. Consumers tend to take several photos from one scene, hoping that one of them will be outstanding, and this leads to large number of similar photos. Therefore, it can be very time-consuming to go through all photos in one of these albums. Summarization is an effective way to provide a quick overview of a set of photos. In this paper, album summarization is defined as selecting a set of photos from a larger collection which best represents the visual information of the entire collection. Selected photos can be used to create a collage of a given album or a cover for an album or to be included in a photo book. However, as already mentioned, manual photo album summarization can be very time-consuming.

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 (http://comminfo.rutgers.edu/conferences/mmchallenge/2010/02/10/hp-challenge-2010/). 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. There is a gap between what people think the summary should look like and what we get with an automatic summarization. For example, funny photos are usually chosen within summarized photos, and they are not easy to detect using computer vision techniques. Therefore, including photos containing humans, such as one's family or friends, in the process of album summarization is needed.

Besides spending a lot of time sharing and consuming content in online social networks, people also use online

applications, especially social games. Players pour huge amounts of time and efforts into games. For example, a recent survey [1] revealed that most players (95%) play social games several times a week, with 64% playing daily. The average game session lasts more than half an hour (i.e., how long 61% play), while 10% may play more than three hours at a time. Work by Von Ahn and Dabbish [2] showed the tremendous power that networks of people possess to solve problems while playing social games. Therefore, the time and effort in playing a game can be utilized to address some issues in image processing community, that is, users entertain themselves while playing an enjoyable game, with the added side effect that they are doing useful work in the process, for example, summarizing one's photo album. This is one of our motivations to develop a novel approach for photo album summarization through gaming.

In this paper, we present and evaluate an approach for photo album summarization through a novel social game "Epitome," which was previously introduced in [3]. It has been implemented as a Facebook application and as an application for mobile phones on the Android OS platform. The main idea of this approach is to show a reduced set of photos from a Facebook album, ask users to play the game, and then integrate results of several users in order to produce a summarization for the whole album. There are two games involved in this approach: "Select the Best!" and "Split it!." In the first game, a user has to select the better of two photos randomly selected from one Facebook album. The goal of the second game is to mimic separation of one album into different events or scenes, by selecting a pair of photos that are more different. 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 sequence of photos which gets the largest number of users' votes represents a summarization sequence of the album. The proof of concept of the proposed method is demonstrated through a set of experiments on several photo albums. We compare results obtained by this game with an automatic image selection, by making use of visual and temporal features. Furthermore, the usability of the game is evaluated by making use of a questionnaire (a user study) on several subjects who played the "Epitome" game. We also address privacy issues concerning shared photos in Facebook applications.

The paper is organized as follows. We introduce related work in Section 2. The proposed social game application is presented in Section 3. Evaluation methodologies and results are discussed in Section 4. Finally, Section 5 concludes the paper with a summary and some perspectives for future study.

2. Related Work

The proposed game is related to different research fields including visual analysis, automatic photo album summarization, and gaming. Therefore, the goal of this section is to review the most relevant work in these fields.

2.1. Automatic Photo Album Summarization. State-of-the-art techniques for automatic photo album summarization are based on time-separated events, spatial information using GPS coordinates, and content-based image similarities. Harada et al. [4] developed an interface for automatic personal photo structuring, considering the time difference between two consecutive photos in order to determine different events. Naaman et al. [5] developed a system which automatically organizes digital photographs considering their geographic location or event-based description extracted from user tags. For photo collection clustering, combination of spatial, temporal, and content-based similarity is then used. This clustering can be used for photo navigation for different categories, such as elevation, season, time of the day, location, weather status, temperature, and time zone. Once photos are clustered, different page layouts are shown. Atkins [6] proposed a photo collection page layout generation method, considering hierarchical partition of the page, which provides explicit control over the aspect ratios and relative areas of photos. This approach attempts to maximize page coverage without having overlapping photos. Geigel and Loui [7] emphasized the 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. In general, however, automatic summarization has its limitations. There is usually a gap between what people think the summary should look like and what automatic summarization produces. A promising solution to narrow the gap is to incorporate human knowledge and preference into the summarization process.

Regarding content-based image similarity, various visual features have been used in automatic photo album summarization. Bag of Words (BoW) model is based on the histogram of local features [8]. Zhang et al. [9] presented a comparative study on the performance of different local features on texture and object recognition tasks based on global histogram of features. BoW method gives a robust, simple, and efficient solution for measuring image similarity without considering the spatial information in images. The BoW mostly uses local feature descriptors, and the Scale Invariant Feature Transform (SIFT) [10] is based on an approximation of the human visual perception. A faster version of the SIFT descriptor with comparable accuracy, called Speeded Up Robust Features (SURF), is proposed in [11]. Another popular feature is the Histogram of Oriented Gradient (HOG) [12]. It is a grid-based histogram on gradient information of the image. This feature was first proposed for human detection, while the recent literature also considers it for general image retrieval. In this paper we use the feature called "tiny", which is a simple 32×32 color image, resized from the original image [13]. It was motivated by psychophysical results showing the remarkable tolerance of the human visual system to degradations in image resolution.

2.2. Crowdsourcing through Games. Ames and Naaman [14] showed that providing incentives to users in form of entertainment or rewards, for example, games, can motivate them to tag photos in online and mobile environments.

Gaming also provides a new way of motivating people by making the subjective data acquisition interesting and enjoyable. The most famous examples of these kind of games are the ESP game and the Peekaboom, developed for collecting information about image content. In the *ESP game* [2], two players, who are not allowed to communicate with each other, are asked to enter a textual label which describes a shown image. The task of each user is to enter the same word as his/her partner in the shortest possible time. In the *Peekaboom* game [15], one player is given a word related to the shown image, and the aim is to communicate that word to the other player by revealing portions of the image, while the second player sees an empty black space in the beginning. This idea served as a basis for several other games [16], such as video tagging, music description and tagging, tag description, object segmentation, visual preference, and image similarities. Foldit [17] is a game that presents simplified three-dimensional protein chains to players and provides a score according to the predicted quality of the folding done by the player. All actions by the player are performed in a three-dimensional virtual world. It requires training to solve complex open protein puzzles which in turn requires a lot of commitment by the players.

Following the presented state-of-the-art techniques, a game-based approach for photo album summarization called "Epitome" was developed first in [3], which provides a collage or a cover photo of the user's photo album, while, at the same time, the user enjoys playing a game. It can also collect research data. In this way, both users and research community can benefit. In this paper, we present an improved version of the game and evaluate it. We compare results obtained by this game with an automatic image selection, by making use of visual and temporal features. Furthermore, the usability of the game is evaluated by making use of a questionnaire (a user study) on several subjects who played the "Epitome" game. We also address privacy issues concerning shared photos in Facebook applications.

3. Algorithms

In this section two algorithms for photo album summarization are described. First, the proposed "Epitome" game is described which takes advantage of many casual gamers to solve the complex problem of album summarization. Then, an automatic visual algorithm is presented as a comparison benchmark to the task.

3.1. Social Game "Epitome". A social game, "Epitome," provides an intuitive and enjoyable user interface as a Facebook application, as shown in Figure 1. The main purpose of the game is to create photo collages for Facebook photo albums considering the feedback of the owners' Facebook friends.

The scenario of the game is as follows. A Facebook user, denoted as a player in this paper, installs the game in his/her Facebook applications page and allows access to his/her photo gallery, as shown in Figure 9(a). Then, the player can select between two games. In the first game, called "Select the Best!," two random photos are shown to the player from one of his/her friends' photo albums chosen randomly and he/she has to choose a better photo, which the player likes more. If the player chooses the photo which is the most frequently selected, then the player's score increases. The second game is called "Split it!." In this game, two pairs of consecutive photos are shown, where the player should select a photo pair which is more different. The results of the two games by many players are combined to produce the summarization of a photo album. In this way, the summarization is conducted based on the feedback of the album owner's friends. The game has appealing look using different visual and audio effects, as shown in Figure 1.

In order to perform summarization using players' inputs, the application calculates three different values: *Importance*, *Segmentation,* and *UserScore*.

Importance value is determined in the "Select the Best!" game for each photo album separately. Two randomly chosen photos are shown to the user and he/she selects the better one in his/her opinion. A feature vector $Selected_n^{\text{best}}$, $n \in [1,\ldots,N]$, is calculated for each player, n among N players, as follows:

$$\begin{gathered} Selected_n^{\text{best}}[i] = \delta_{i,s}, \\ Appeared_n^{\text{best}}[i] = \delta_{i,j} + \delta_{i,s} \\ \delta_{i,j} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \end{gathered}, \tag{1}$$

where $i, j, s \in [1,\ldots,M]$, M is the size of a particular Facebook album, j, s are indices of the two photos shown to the player, and s is index of the selected photo. The vector $Appeared_n^{\text{best}}$ of dimension M stores the frequency of all photos that appear in the game. At the end, we perform normalization on vector $Selected_n^{\text{best}}$ by element-wise division in order to obtain *Importance*:

$$Importance[i] = \frac{\sum_n Selected_n^{\text{best}}[i]}{\sum_n Appeared_n^{\text{best}}[i]}, \tag{2}$$

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

Segmentation vector is calculated in the "Split it!" game for each photo album separately in an analogous way to that for *Importance*. Two pairs of consecutive photos are chosen randomly from the album and the player selects the more different pair of photos. A feature vector $Selected_n^{\text{segm}}$, $n \in [1,\ldots,N]$, is calculated for each player, n among N players, as follows:

$$\begin{gathered} Selected_n^{\text{segm}}[i] = \delta_{i,s}, \\ Appeared_n^{\text{segm}}[i] = \delta_{i,j} + \delta_{i,s}, \end{gathered} \tag{3}$$

where $i, j, s \in [1,\ldots,M-1]$, M is the size of a particular Facebook album, $j, j+1$ and $s, s+1$ are indices of the the photo pairs, shown to the player, and $s, s+1$ are indices of selected photo pair. The vector $Appeared_n^{\text{segm}}$ of dimension M stores the frequency of all photos that appear in the game. At

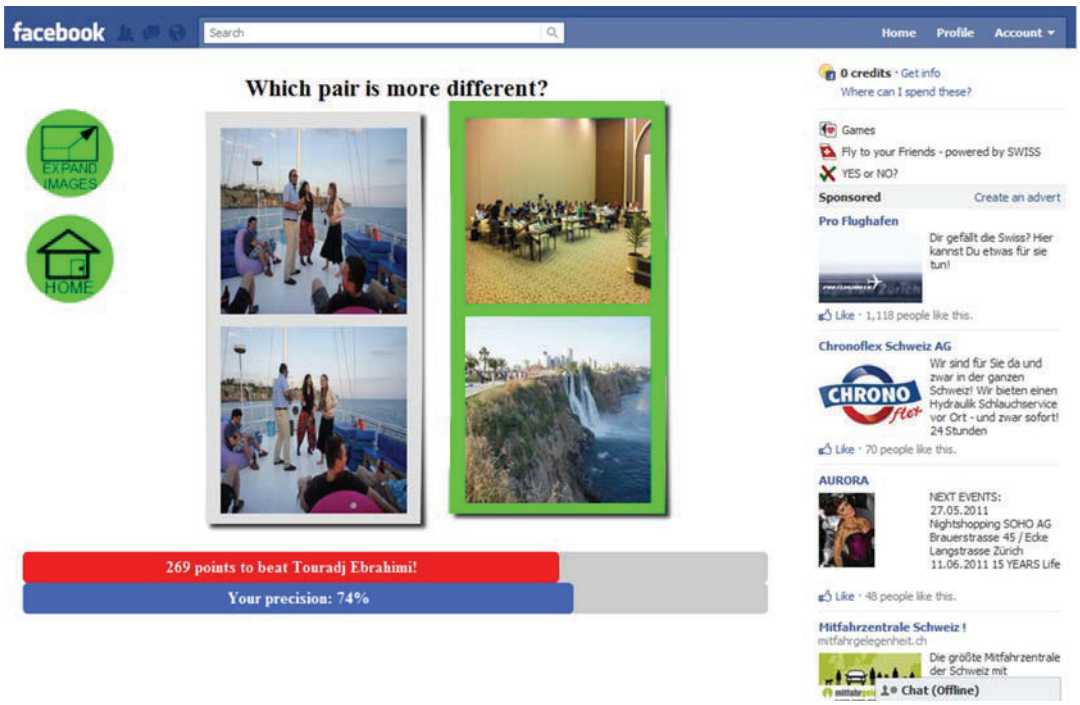

Figure 1: Screenshot from the Epitome game.

the end, we perform normalization on vector $Selected_n^{\text{segm}}$ by element-wise division in order to obtain *Segmentation*:

$$Segmentation[i] = \frac{\sum_n Selected_n^{\text{segm}}[i]}{\sum_n Appeared_n^{\text{segm}}[i]}, \tag{4}$$

which is an M-dimensional vector showing the frequency with which each photo in one Facebook album is selected as a starting photo in a new segment.

Finally, vectors *Importance* and *Segmentation* are used to automatically select L most representative photos within one Facebook photo album, as shown in Figure 2. In this game L was arbitrarily set to five. First, $L - 1$ maximum values from the vector *Segmentation* of the album are determined in order to segment the album into L most probable segments. For each of these segments, a photo with the highest score in the vector *Importance* is chosen. These L photos represent a collage of the album, which is shown to the owner of the album if he/she reaches a certain level of *UserScore*.

UserScore value is defined to motivate players to play this game frequently. For example, in the "Select the Best!" game, the player increases his/her own *UserScore* if he/she selects the photo which has the higher or equal *Importance* value among two photos. The same approach is used in the "Split it!" game, where the player increases his/her *UserScore* if he/she selects the separation place where *Segmentation* value is the highest among two separation places. Initially *UserScore* is set to zero. *UserScore* values for all players are sorted to show ranking of players in the "Epitome" game.

3.2. Automatic Photo Album Summarization. Automatic photo album summarization is performed considering different visual and temporal features. After extracting these

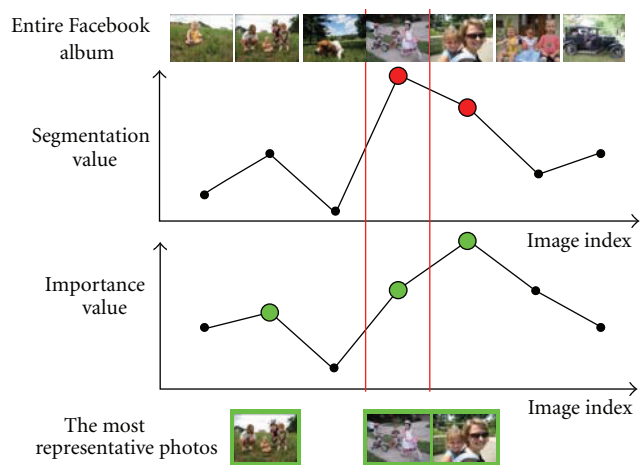

Figure 2: An example of selecting the three most representative photos within one Facebook album through the "Epitome" game.

features, the album is segmented into five parts by calculating the four highest Euclidean distances of the consecutive photos features. For each image in a particular segment, we calculate the sum of the Euclidean distances between that feature of the photo and the rest of the image features in the segment. The image with the lowest sum is then selected as the most representative photo in that segment.

Different features can be used for segmentation and to select the most representative photo in the segments. We considered the following features: Bag of Words (BoW) method based on Speeded Up Robust Features (SURF), Histogram of Oriented Gradient (HOG), HSV (Hue, Saturation,

Value) color histogram, "tiny" features, and time stamp, as described below.

Bag of Words model in computer vision was derived from BoW model in natural language processing (NLP) [8]. A similar method in computer vision documents represents images or objects, and visual clusters of local features are considered as a word. In our case, SURF features were used as local features [11]. BoW is a vector which represents the histogram of visual features. Therefore, this method does not consider spatial information or order of visual features. 1000 feature clusters were calculated by a hierarchical k-means algorithm. Each image is represented by 1000 normalized values.

Histogram of Oriented Gradients [18] calculates the histogram of gradients in the region around one keypoint. It is evaluated on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. Using gradient information for feature description is very robust to different illumination conditions. The dimensions of the HOG features are around 1000.

Color histogram descriptor is extracted from photos in the HSV domain. Color descriptors often fail in image retrieval in different lighting conditions; however, in our case photos from one Facebook album are compared and assumed to have similar lighting conditions. The dimensions of the HSV color features are around 1000.

Tiny feature is used as baseline representing scaled 32×32 grayscale tiny images.

Time stamp is extracted from EXIF for further analysis. It corresponds to the time order by which photos were uploaded to Facebook.

4. Evaluation and Results

Evaluation of the "Epitome" game can be performed in two ways: performance and usability of the game.

4.1. Performance Evaluation. The performance of summarizing albums with the "Epitome" game is evaluated with respect to the ground truth given by humans.

The dataset of photos used for performance evaluation is the official dataset from "HP Challenge 2010: High Impact Visual Communication" at the "Multimedia Grand Challenge 2010" [19]. Some example photos are shown in Figure 3. It consists of six albums, each with 20 photos. These albums 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 7 provides example photos within one of the albums.

We first constructed a ground truth by asking different people to subjectively perform summarization 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 (average age was 31), with different backgrounds and cultural differences. In the collection of the ground truth data, participants were shown 20 photos belonging to the same album. The task of the participants was to select the five most representative photos of the whole album, while looking at all photos of that album.

For simplicity of the explanation on how the designed photo selection tool (social game) was evaluated, let us consider only one album with $M = 20$ photos. First, ground truth data is collected. Every user n among N users is asked to select the five most representative photos. After his/her participation in collecting the ground truth data, the corresponding feature vector $Selected_n$, $n \in [1, \ldots, N]$, is formed as follows:

$$Selected_n[i] = \sum_{k \in [1,\ldots,5]} \delta_{i,s_k}, \tag{5}$$

where $i, s_k \in [1, \ldots, M]$ and for all $k, l \in [1, \ldots, 5] : s_k \neq s_l$. s_k for $k \in [1, \ldots, 5]$ are the five indexes of the photos which were chosen as the representative ones. The selected indexes are distinctive. Feature vectors of the users n and m, $n, m \in [1, \ldots, N]$, are then compared to each other, and the score of their matching $S_{n,m}$ is calculated as

$$S_{n,m} = Selected_n \cdot Selected_m^T. \tag{6}$$

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

$$Performance_n = \sum_{i=1}^{N} S_{n,i}. \tag{7}$$

The maximum value in the vector *Performance* 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 performance is $5 \times N$, which in our case becomes 315. These results are considered as the ground truth data and compared with the results obtained from the games and from the automatic album summarization algorithms, in order to prove the concept of the approach. All computations are repeated in a same way for all albums.

Then, the participants are asked to play our game with the selected dataset. The vectors *Importance* and *Segmentation* of dimension M are determined for each album, which are described in Section 3.1. These values are used to automatically select the $L = 5$ most representative photos within each album in 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 (6) and (7).

Furthermore, the performance of this game is compared to the performance of an automatic image selection which considers different visual and time features described in Section 3.2. We calculated the performance of 20 different feature pairs, considering five features for segmentation and four features for choosing the representative images, as shown in Figure 4. The result shows that the best

Figure 3: Some example photos for each of six albums. Photos in each row belong to the same album. The albums cover a large variety of objects and scenes usually taken during a vacation.

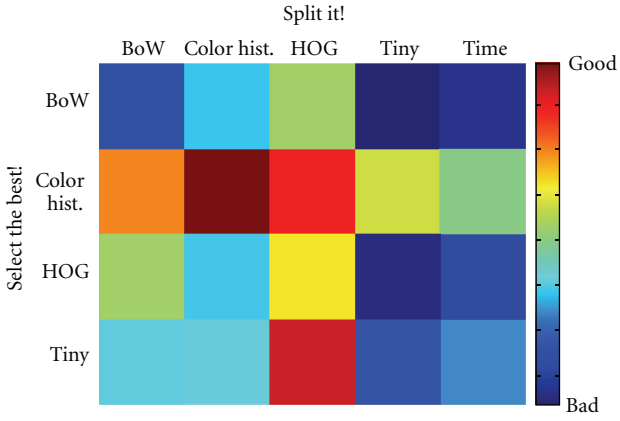

Figure 4: Comparison between different visual and time features. The best performance is achieved with "color histogram" feature for both "Split it!" and "Select the best!" tasks. Dark red color indicates the best (*Performance* ≈ 100) and dark blue color indicates the worst performing algorithm (*Performance* ≈ 70). For example, using "time" feature for segmenting an album and "BoW" feature for selecting the most representative images gives poor results on evaluation.

performance (around 100) is achieved by the pair of "color histogram" features for album segmentation and best photo selection in the segment. In the following the performance of automatic visual analysis, represented by color histogram, is compared with the "Epitome" game.

Figure 5 shows the distribution of the participants' performance, including the choice of the proposed method and the automatic visual analysis. All performances are sorted in a descending order. As one can see, the performance of the proposed method is better than the automatic visual analysis since it is closer to the best performance of users for ground truth generation. On average, this approach achieves 80% of the performance of the best user for each album, which proves the concept of the game. It also outperforms the automatic visual analysis, which can achieve performance of 64%. For albums three and five, this value is even higher, that is, about 95%. The most representative photos for one of the albums selected by the proposed method are shown in Figure 7. Figure 6 shows the comparison of performance in summarizing photo albums performed by the "Epitome" game, automatic photo selection using color histogram, and users who participated in creating the ground truth data.

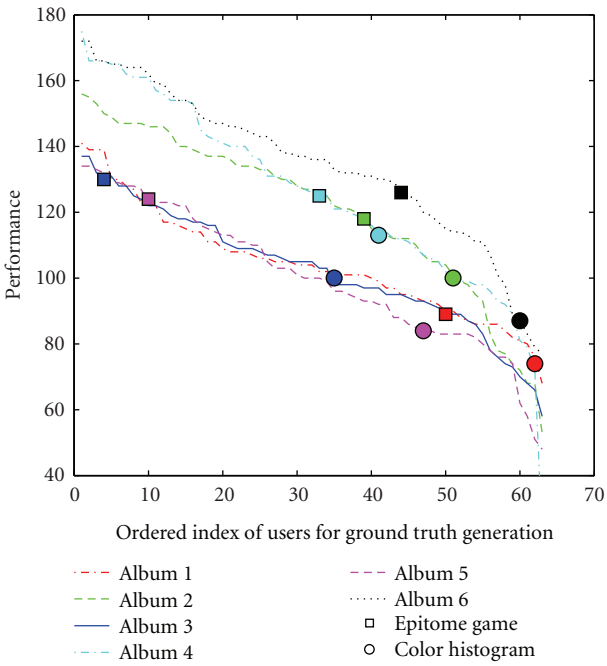

Figure 5: The distribution of the participants' performance. The results of the "Epitome" game are shown with square markers and the results of automatic visual analysis with circle marker. Different colors of the markers correspond to different albums.

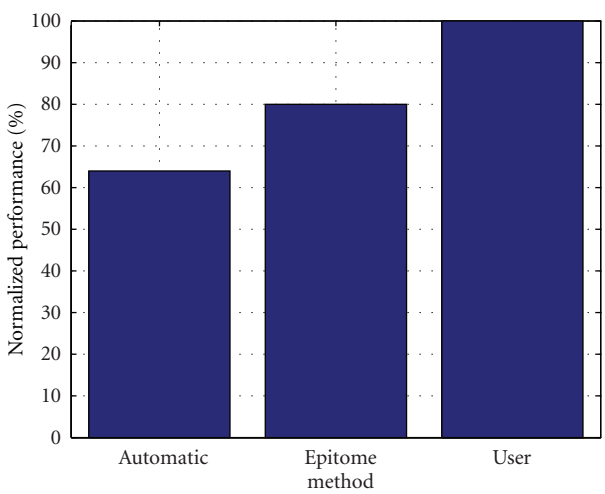

Figure 6: The comparison of normalized performance in summarizing photo albums performed by the "Epitome" game, automatic photo selection using color histogram, and users who participated in creating the ground truth data.

4.2. Usability Evaluation. The usability of the "Epitome" game is evaluated through a user study. We asked participants (users) to play the game with different Facebook photo albums and to provide us with their feedback on the game in the form of a questionnaire.

We recruited 40 participants, aged 23–46 (average age was 28), with different cultural backgrounds. First, all participants were introduced to the "Epitome" game by showing them basic rules on how to play the game. Then, all participants spent sufficient time to play the game. After a participant played with the "Epitome" game with different Facebook photo albums, a questionnaire was used to obtain the feedback from the participant. The questionnaire consists of three groups of questions:

(i) general questions about motivation to play the game and enjoyment;

(ii) questions to assess different platforms for playing the game (mobile, Facebook, or simple web page), for example, satisfaction with visual presentation for each of them;

(iii) questions about privacy issues regarding showing one's photos to his/her friends, friends of friends, and everybody or nobody.

In this study, we used discrete rating scales with adjective description of each level. Depending on the question, participants had to choose one of the answers or to rank answers according to their preferences. For each of the questions we calculated mean of the participants responses.

In this paper, we do not describe the whole questionnaire and results, but we rather discuss some of the interesting outcomes from our study. All questions are listed in the appendix. The questionnaire with choices is public ly available (http://mmspg.epfl.ch/files/content/sites/mmspl/files/shared/questionnaire_epitome.pdf).

4.2.1. Motivation to Play the Game. Questions (1)–(6), (9)–(14), and (21) listed in the appendix belong to questions about motivation to play the game. Results showed that 70% of the players are very satisfied with the game. We further asked players of the "Epitome" game what *motivated* them most to play the game. Results are shown in Figure 8. Players enjoyed the most to watch their Facebook friends' photos, which was even more preferred than the original goal of the game, that is, getting their own albums summarized. We observed another interesting value of the game that people like the idea of watching (browsing) friends' photos through the "Epitome" game. Players were not motivated to play "Epitome" by the fact that they participate in collecting research data. This shows that fun and enjoyment are important aspects of the game that should be considered. In another question about motivation, players prefer more to see their friends' photos compared to photos of some unknown people. This promotes the importance of the social part of the game.

One of the questions was about preferred patterns of playing the game. Like other casual games, players would like to play our game several times a month, and around five minutes every time.

4.2.2. Platform. An important question we discuss here is about *different platforms* for playing the "Epitome" game (questions (7) and (8) listed in the appendix), such as a simple web page, a mobile phone, and a Facebook application. Average ranks for these platforms are 2.3, 2.2, and 1.5, respectively, which shows that players prefer Facebook

Figure 7: Photos from album 3. The most representative photos selected by the proposed method are marked with green bounding box, while the red bounding box denotes photos selected by making use of color histogram.

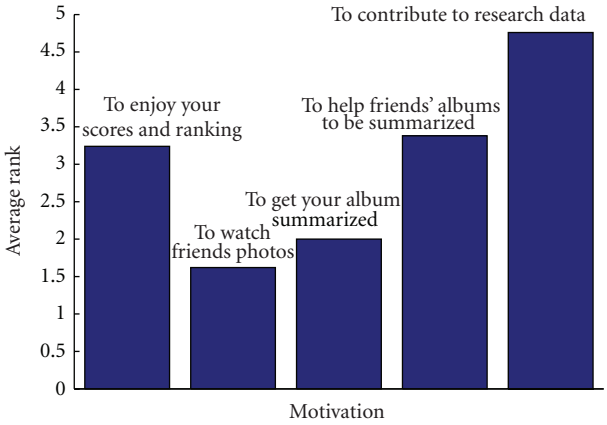

Figure 8: Average rank of different motivations to play the "Epitome" game. Lower average ranks are better.

the most. Surprisingly, players have similar preferences for mobile phone application and simple web page. One of the reasons for this could be that the mobile phone had limited bandwidth in wireless connection and the game was faltering while loading some images from Facebook.

4.2.3. Privacy Issues. We also addressed *Facebook permission issues*. To play the third party applications, like the "Epitome," users should accept an agreement with an application on accessing users' data stored in Facebook. But, we note that users' privacy settings in Facebook are different from what the third party application actually access. If the user allows the third party applications to access his/her photos, they get right to distribute and modify (resize, rotate, change in color perception, etc.) photos of the user. Before using this application for the first time, Facebook shows to the player a permission page informing him/her what kind of data will be retrieved from his/her Facebook account if he/she allows access to the application, as shown in Figure 9(a). People are usually not concerned about this issue and easily allow access to data by the application. In order to address the privacy issue regarding this process of allowing access to the data, we created new visually intuitive permission pages, as shown in Figure 9(b)–9(d).

In our experiments, users were asked whether they allow access to their data using the default Facebook permission page either on the mobile phone or in Facebook, and were separately asked if they would allow access to any of the three new permission pages. We measured how many players allow the "Epitome" game to access photos they have uploaded, photos they have been tagged in, and photos their friends have uploaded. Results are depicted in Figure 10. Clearly, the players understand the risk better by viewing our illustrative permission pages than the default Facebook permission page, and more than 90% of the players did not allow the application to retrieve their photos that can be further modified or distributed within "Epitome." This shows that the default Facebook permission page is neither sufficiently intuitive nor informative.

The users do not have sufficient control over details about permission in Facebook applications. From the questions related to permission settings of shared photos (questions (15)–(20) listed in the appendix), we conclude that players would not like to give more permissions to the application compared to the permission they already set for their photos in Facebook. For example, 86% of the users would like to share their private photos through this application only with their friends.

4.3. Statistics of the Game. The "Epitome" game was published on Facebook in June 2011, and during two months, 49 users played it 5870 times on a dataset of 21780 photos. Distribution of players' score is shown in Figure 11. A few

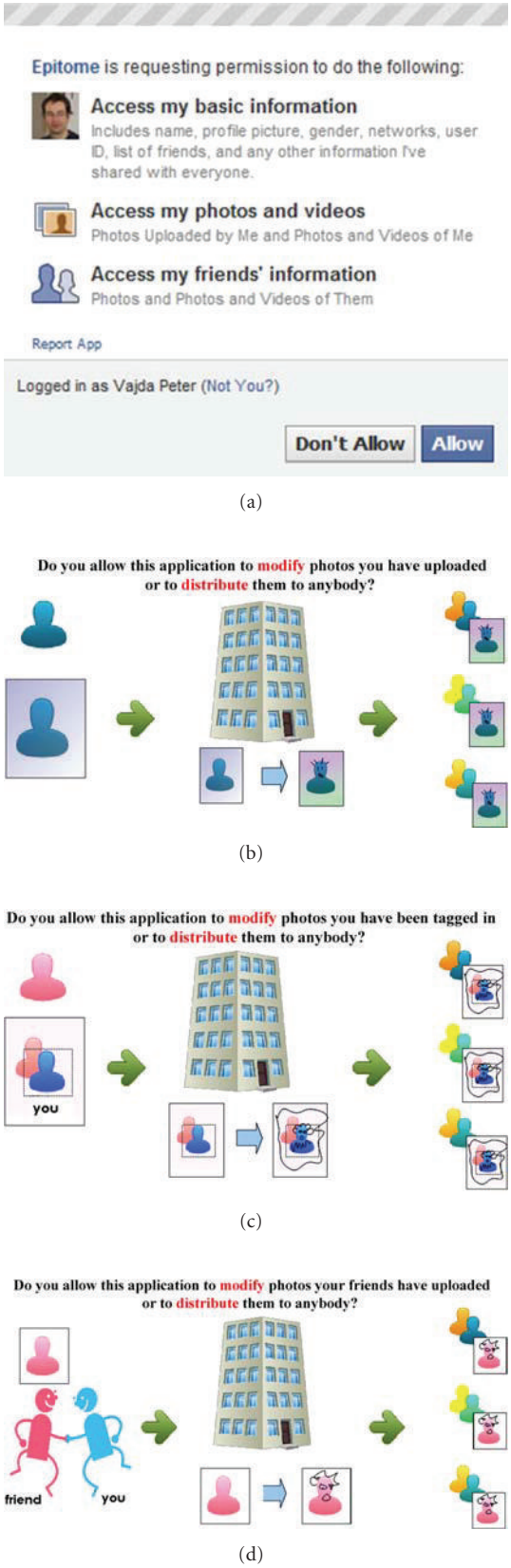

Figure 9: Different permission pages used in our study: (a) default Facebook permission page, (b) *user_photos* permission page, (c) *user_photo_video_tags* permission page, and (d) *friends_photos* permission page.

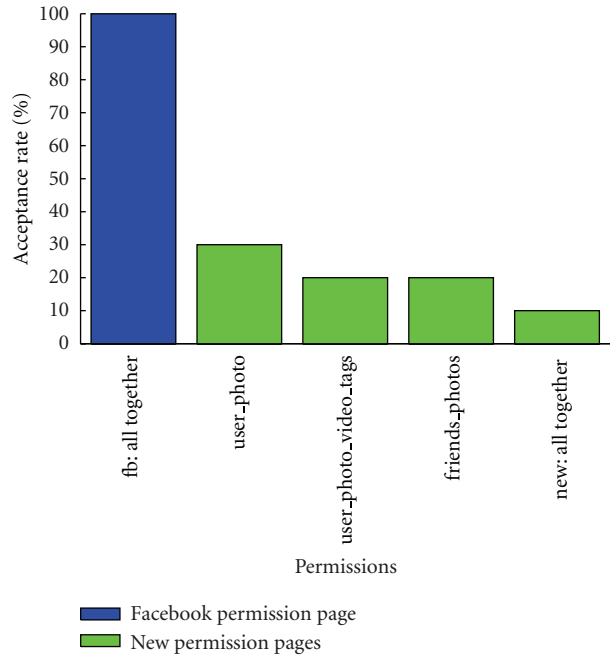

Figure 10: Acceptance rate for the default Facebook and permission pages used in our study.

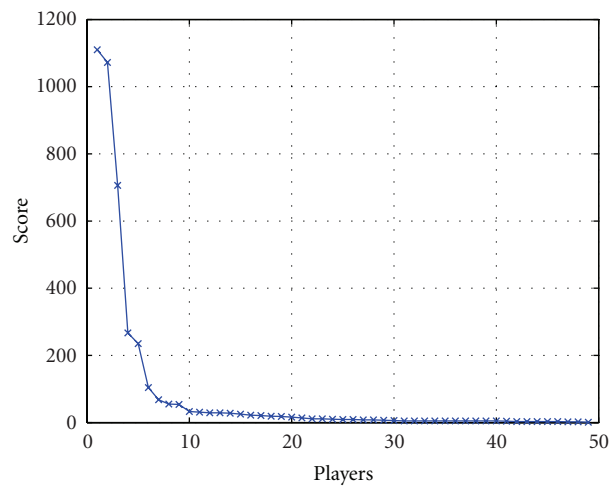

Figure 11: The distribution of players' score in the "Epitome" game. Scores are sorted in descending order.

players played the game frequently and thus had higher scores than the others. Many new users started recently playing this game and therefore they still have low scores. Figure 12 shows the distribution of the photos' score, that is, the number of votes per appearance of each photo. Again, since the "Epitome" game was recently published online, there are much more photos available, especially from new users, than those photos users played with in the game, and therefore many photos are not shown yet to users. This is the reason for many extreme values (score of zero and one) in Figure 12. Figure 13 shows the number of pictures changed in collages over time. It can be concluded from this figure that it converges.

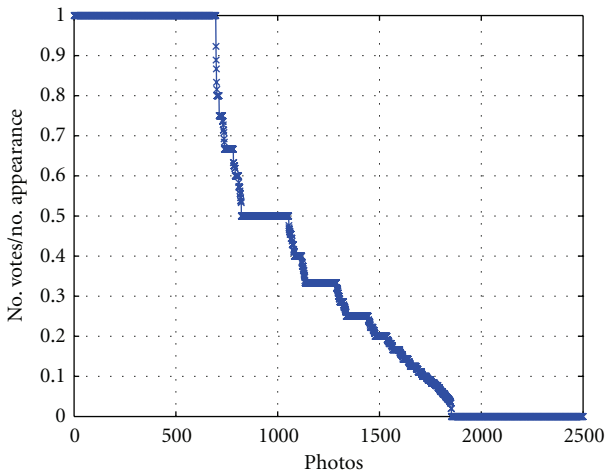

Figure 12: The distribution of photos' scores in the "Epitome" game, that is, the number of votes per appearance of each photo. Scores are sorted in descending order. The rest of the photos did not yet appeared or nobody voted for them.

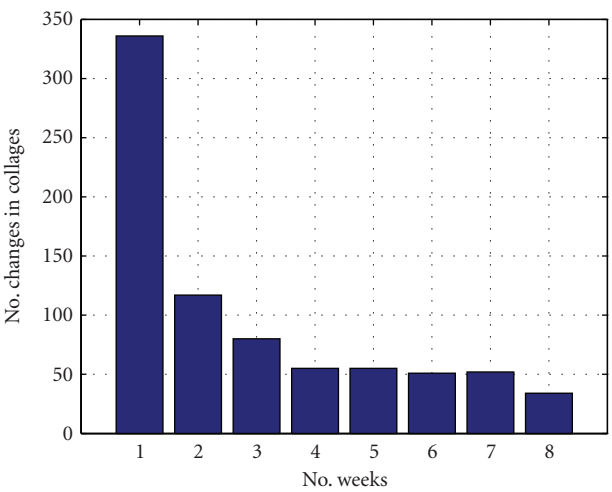

Figure 13: The number of photos changed in collages over time since the "Epitome" game was launched.

4.4. Advantages and Disadvantages. In summary, the "Epitome" game has the following advantages.

(1) Performance of the game-based album summarization is better than using only computer vision approaches, which was shown in [20].

(2) People like to watch their friends' photos through this game, which also encourages social interaction between them.

(3) The game itself is interesting and people can have fun through the game.

However, a disadvantage is the processing time for generating fine album summarization, as shown in [20].

5. Conclusion

In this paper, we described and analyzed a social game, "Epitome," for photo album summarization on Facebook. The game is a social application to enjoy photos of one's Facebook friends, while contributing to summarization of their photo albums and collecting research data. The proof of concept of the game was demonstrated and validated through a set of experiments on several photo albums. The results of the experiments showed that the summarization game achieves 80% of the best performance of different participants and significantly outperforms automatic visual summarization methods (64%). The usability of this game was validated by making use of a questionnaire. The results of our user study showed that the main motivation for a player of the game is to watch his/her friends' photos and obtain his/her album summarization. Finally, a default Facebook permission page was analyzed and considered as not sufficiently intuitive nor informative.

As a future study, we will make the game more attractive for users and also consider to include in this approach more sophisticated visual analysis. We also plan to improve the game by reducing the bandwidth which is necessary to load all images.

Appendix

The usability of the "Epitome" game is evaluated by making use of a questionnaire (a user study) on several subjects who played the game. The questionnaire consists of three groups of questions: (1) general questions about motivation to play the game and enjoyment, (2) questions to assess different platforms for playing the game, and (3) questions about privacy issues regarding showing one's photos to his/her friends, friends of friends, and everybody or nobody. Questions are listed in the following.

(1) Are you satisfied with the Epitome game? If not completely satisfied, what is the main reason for that?

(2) Please rank the motivations to play the Epitome game according to your preferences in order to make it more enjoyable?

(3) Please rank the improvements of the Epitome game according to your preferences in order to make it more enjoyable?

(4) How often would you play the Epitome game?

(5) How long would you play the Epitome game at once?

(6) Would you prefer to play only one integrated game?

(7) Please rank three platforms for playing the Epitome game according to your preferences?

(8) How do you like the Mobile interface? How do you like the Facebook interface?

(9) Would you enjoy the Epitome game more if you play with less than nine images? If yes, how many images should be displayed?

(10) How much do you prefer to watch your friends' photos compared to the photos of unknown people?

(11) Is it good to show your rank and compare it with your friends' ranks for the enjoyment of Epitome?

(12) Is it good to have your summarization sequence as a result of the Epitome game?

(13) How many images in album summarization sequence of photos would you prefer?

(14) There are two statements: 1st statement—to have perfectly summarized Facebook album, but waiting for it long time period; 2nd statement—to have preliminarily summarized Facebook album after a short time. Which of these statements is more important for you?

(15) To whom would you allow Epitome to show your private photos which are not shared even with your friends in order to receive a good summarization of your Facebook albums?

(16) To whom would you allow Epitome to show your private photos which are shared just with your friends in order to receive a good summarization of your Facebook albums?

(17) To whom would you allow Epitome to show your private photos which were shared with friends of friends in order to receive a good summarization of your Facebook albums?

(18) To whom would you allow Epitome to show photos in which you were tagged in order to receive a good summarization of your Facebook albums?

(19) To whom would you allow Epitome to show photos of your friends in order to receive a good summarization of your Facebook albums?

(20) Do you want to play with photos of your friends even if they do not play Epitome?

(21) Any suggestions to improve the game?

Acknowledgments

This work was supported by the Swiss National Foundation for Scientific Research in the framework of NCCR Interactive Multimodal Information Management (IM2) and the Swiss National Science Foundation Grant "Multimedia Security" (no. 200020-113709), partially supported by the European Network of Excellence PetaMedia (FP7/2007-2011).

References

[1] "Social games: news and survey findings," http://content.usatoday.com/communities/gamehunters/post/2010/02/social-games-news-and-survey-findings/1.

[2] L. Von Ahn and L. Dabbish, "Labeling images with a computer game," in *the Conference on Human Factors in Computing Systems (CHI '04)*, pp. 319–326, April 2004.

[3] I. Ivanov, P. Vajda, J. S. Lee, and T. Ebrahimi, "Epitome—a social game for photo album summarization," in *the 1st ACM Workshop on Connected Multimedia (CMM '10)*, pp. 33–38, October 2010.

[4] S. Harada, M. Naaman, Y. J. Song, Q. Wang, and A. Paepcke, "Lost in memories: interacting with photo collections on PDAs," in *Proceedings of the 4th ACM/IEEE Joint Conference on Digital Libraries (JCDL '04)*, pp. 325–333, June 2004.

[5] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina, "Automatic organization for digital photographs with geographic coordinates," in *Proceedings of the 4th ACM/IEEE Joint Conference on Digital Libraries (JCDL '04)*, pp. 53–62, June 2004.

[6] C. B. Atkins, "Adaptive photo collection page layout," in *the International Conference on Image Processing (ICIP '04)*, pp. 2897–2900, October 2004.

[7] J. Geigel and A. Loui, "Using genetic algorithms for album page layouts," *IEEE Multimedia*, vol. 10, no. 4, pp. 16–27, 2003.

[8] F. F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05),* pp. 524–531, June 2005.

[9] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[11] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speededup robust features," in *Proceedings of the 9th European Conference on Computer Vision*, pp. 404–417, 2006.

[12] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.

[13] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: a large data set for nonparametric object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1958–1970, 2008.

[14] M. Ames and M. Naaman, "Why we tag: motivations for annotation in mobile and online media," in *the 25th SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, pp. 971–980, May 2007.

[15] L. Von Ahn, R. Liu, and M. Blum, "Peekaboom: a game for locating objects in linages," in *the Conference on Human Factors in Computing Systems (CHI '06)*, pp. 55–64, April 2006.

[16] E. Law and L. von Ahn, "Input-agreement: a new mechanism for collecting data using human computation games," in *the Conference on Human Factors in Computing Systems (CHI '09)*, pp. 1197–1206, 2009.

[17] F. Khatib, F. Dimaio, S. Cooper et al., "Crystal structure of a monomeric retroviral protease solved by protein folding game players," *Nature Structural and Molecular Biology*, vol. 18, no. 10, pp. 1175–1177, 2010.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 886–893, June 2005.

[19] "HP Challenge 2010 Dataset: High Impact Visual Communication," http://comminfo.rutgers.edu/conferences/mmchallenge/2010/02/10/hp-challenge-2010.

[20] P. Vajda, I. Ivanov, L. Goldmann, and T. Ebrahimi, "Social game epitome versus automatic visual analysis," in *Proceedings of IEEE International Conference on Multimedia and Expo*, July 2011.