SOCIAL GAME EPITOME VERSUS AUTOMATIC VISUAL ANALYSIS

Paper ID ***

ABSTRACT

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 in one's life, it is desirable to have an efficient summarization tool which can help people to get a quick overview of an album containing huge number of photos. In this paper, we 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, at the same time, the user enjoys playing the game. As a benchmark comparison to this game, we performed automatic visual analysis considering several state-of-the-art features.

Index Terms— photo summarization, social game, social networks, Facebook application, visual analysis

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. Beside storing them on computer hard drives, people also 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.

Users usually organize their photos in albums (collections) based on places, events or people. By sharing these albums with others, they want to tell their own stories of some important events in their life, such as birthday party, vacation, wedding, or birth of a baby. It can be very time-consuming to go through all photos in one album, and therefore summarization is an effective way to help getting a quick overview of a set of photos. Album summarization can be 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, a cover for an album, or to be included in a photo book.

In this paper, we analyse an approach for photo album summarization through a novel social game "Epitome" introduced by Ivanov *et al.* in [1], which is extended to a Facebook application. 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 all users in order to produce a summarization for the whole album. There are two games involved in this approach to album summarization: "Select the Best!" and "Split it!". In the first game, a user has to select the most representative (appealing) photo of a reduced set of images. The goal of the second game is to split a 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 sequence of photos which gets the highest number of users' votes represents a summarization sequence of photos. The proof of concept of the proposed method is demonstrated through a set of experiments on several photo collections. Moreover, we compare results obtained by this game with an automatic image selection, making use of visual and time features.

The paper is organized as follows. The social game application and its implementation is presented in Section 3. Experiments 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

Current state-of-the-art techniques are based on automatic summarization which considers time separated events, spatial information using GPS coordinates and content-based image similarities. Susumu *et al.* [2] developed an interface for automatic personal photo structuring based on the time difference between two consecutive photos in order to determine different events. Naaman *et al.* [3] developed a system which does automatic organization of digital photographs considering the geographic location of photo or event based description extracted from user tags. Combination of spatial, temporal and content-based similarity is then used for photo collection clustering. This clustering can be used for photo navigation and search 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 should be considered. Atkins [4] proposed a photo collection page layout generation method based on a 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 [5] 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. An automatic summarization has its limitations. There is a gap between what people think the summary should look like and what we get with an automatic summarization.

Regarding content based image similarity, different visual features have been used in automatic photo album summarization. Bag of Words (BoW) model is based on histogram of local features [6]. Zhang *et al.* [7] presented a comparative study on 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 image. BoW mostly uses local feature descriptors, the Scale Invariant Feature Transform (SIFT) [8], which is based on an approximation of the human visual perception. A faster version of SIFT descriptor with comparable accuracy, called Speeded Up Robust Features (SURF), is proposed in [9]. Another popular feature is the Histogram of Oriented Gradient (HOG) [10]. 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 consider it for general image retrieval. In this paper we use feature called ”tiny”, which is a simple $32x32$ color images, resized from the original image. Motivated by psychophysical results showing the remarkable tolerance of the human visual system to degradations in image resolution, this feature was evaluated in [11].

Ames and Naaman [12] showed that providing incentives to the user in form of entertainment or rewards, e.g. games, can motivate them to tag photos in online and mobile environments. Gaming also provides a new way of motivating people to make the subjective data acquisition interesting and enjoyable. The most famous examples of these kind of games are the ESP Game and Peekaboom, developed for collecting information about image content. In *ESP Game* [13], two players, who are not allowed to communicate with each other, are asked to enter a textual label which describes a shown image. The aim of each user is to enter the same word as his/her partner in the shortest possible time. In *Peekaboom* game [14], 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, such as video tagging, music description and tagging, tag description, object segmentation, visual preference and image similarities. The social game presented in this paper can collect research data and, at the same time, it provides a collage or a cover photo of the user's photo album, while, at the same time, the user enjoys playing a game. In this way, both users and research community can benefit.

3. EPITOME GAME

The goal of this application is to provide an intuitive and enjoyable user interface as a Facebook application, which creates and annotates photo collages for Facebook photo albums. Therefore, the game “Epitome” 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 a user's photo album and provides useful research data.

The scenario of the game is as follows. A Facebook user, in this paper denoted as a player, installs the game and allows access to his/her photo gallery, as shown in Figure 1. Then, the player can select between two games. In both games, 9 consecutive photos are selected from one of the Facebook albums chosen randomly. 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. If the player chooses the photo which is the most frequently selected by other players, then player's score increases. The second game is called “Split it!”, where the player should split images into two parts which have distinct semantic meanings. In this game, the photos are shown in the time order in which they were captured. The time stamp is extracted from EXIF tags associated to each photo. The results of “Select the Best!” and “Split it!” games are combined to form a score and if a user reaches a certain score level, then the photos for the collage of the user's photo album are shown to the owner. Therefore, the player can get a feedback from all other players, regarding his/her Facebook photo albums. The game has appealing look using different visual and audio effects, as shown in Figure 1.

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. 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 randomly chosen photos. A feature vector $BestSmall_n$, $n \in [1, N]$, is calculated for each player, n among N players, as follows:

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

where $\alpha_{n,k} \in \{0, 1\}$, for $k \in [1, K]$, is either 1 or 0 depend-

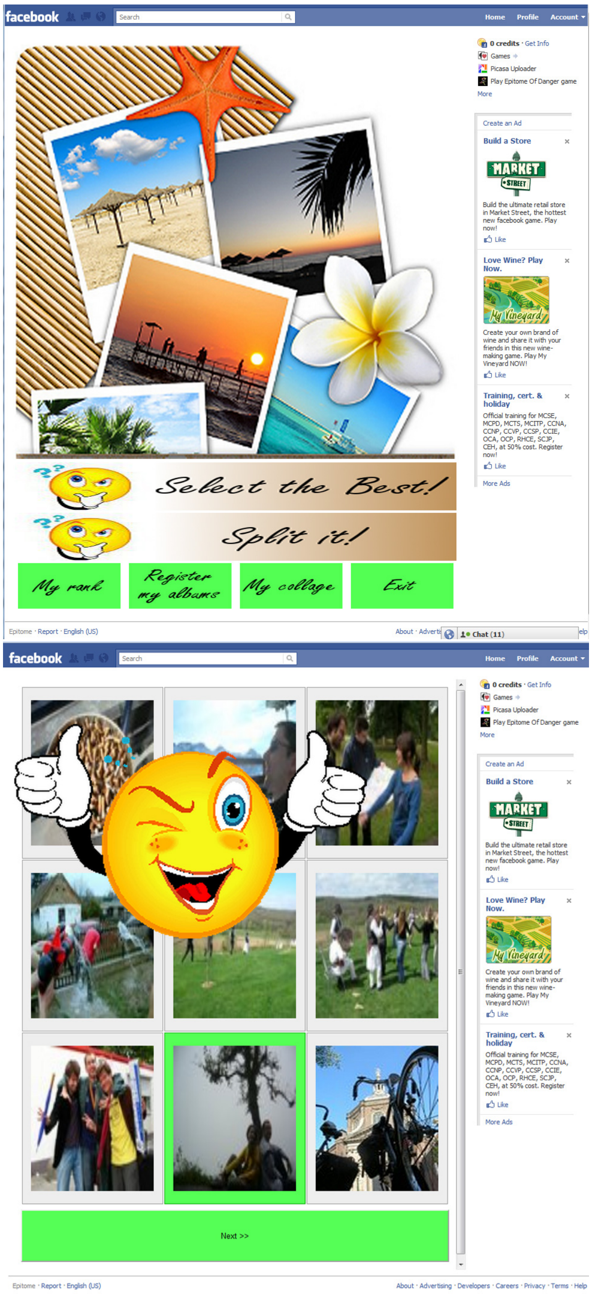

Fig. 1. Main game selection, "Select the Best!" screenshots of the games are shown.

ing on whether the corresponding photo is chosen as the most representative photo. This vector is then expanded to a vector $Best_n$ of dimension M, where M is the size of a particular Facebook album, as follows:

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

where $y \in [1, M-K+1]$ is the index of the first photo shown to the player. A vector $BestFreq$ of dimension M stores the frequency of all photos that appear in the game. An M-dimensional vector $BestCount$ is then calculated as: $BestCount = \sum_n Best_n$, $n \in [1, N]$. At the end, we perform normalization on vector $BestCount$ by element-wise division:

$$Importance = \frac{BestCount}{BestFreq}, \tag{3}$$

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

Segmentation vector is calculated in "Split it!" game for each photo album separately in an analogous way as explained for *Importance* value. It shows the frequency with which each photo in one album is selected as a starting photo in a new segment.

Finally, vectors $Importance$ and $Segmentation$ are used to automatically select $L = 5$ most representative photos within one Facebook photo album. At first, the particular album is segmented into L most probable segments by determining $L-1$ maximum values from the vector $Segmentation$. 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 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. Dataset

The dataset used in our experiments is the official dataset from "HP Challenge 2010: High Impact Visual Communication" at the "Multimedia Grand Challenge 2010" [15]. Some example photos are shown in Figure 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, flow-

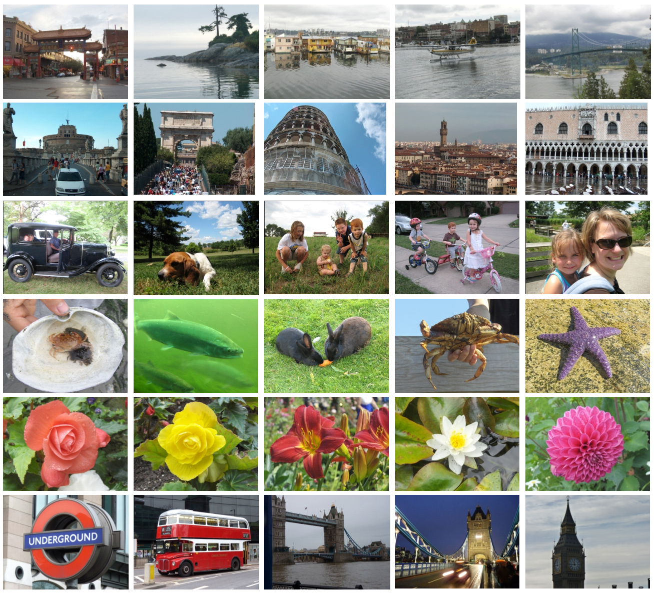

Fig. 2. Some example photos for each of 6 datasets. Photos in each row belong to different datasets. The datasets cover a large variety of objects and scenes usually taken during a vacation.

ers and sea animals. Figure 5 provides example photos of the datasets.

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

Then, participants were asked to play two games “Select the Best!” and “Split it!” with a dataset from Section 4.1. The results obtained from these games are used to assess the performance of the approach by comparing them with the ground truth and results from automatic visual analysis.

Furthermore, we performed automatic photo album summarization considering different visual and temporal features. At first ”Bag of Words” method based on SURF features, ”Histogram of Oriented Gradients”, ”HSV Color histogram” and ”Tiny” features are extracted. Where ”Tiny” feature is used as benchmark representing scaled $32X32$ grayscale tiny images. The dimensions of the features are around 1000. Moreover creation ”Time” stamp is extracted from EXIF for further analysis. We segment the album into 5 parts by extracting the four highest Euclidean distances of the consecutive photos’ features. For each image in the 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. Therefore we calculated the performance of 20 different feature pairs.

4.3. Results and analysis

For simplicity of the explanation on how this 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:

$$Full_n = [\ \delta_{n,1},\quad \delta_{n,2},\quad \delta_{n,3},\quad \ldots,\quad \delta_{n,M}\ ], \tag{4}$$

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:

$$S_{i,j} = Full_i \cdot Full_j^{\mathrm{T}} \tag{5}$$

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:

$$Score_i = \sum_{j=1}^{N} S_{i,j} \tag{6}$$

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 $5XN$, which in our case becomes 315. These results are considered as the ground truth data and compared with the results obtained from the games in order to prove the concept of the approach. All computations are repeated in a similar way for all 6 datasets.

Furthermore, the results obtained in this game are compared with the results of an automatic image selection by making use of visual and time features. We calculated the performance of 20 different feature pairs as shown in Figure 3. The result shows that the best performance is achieved

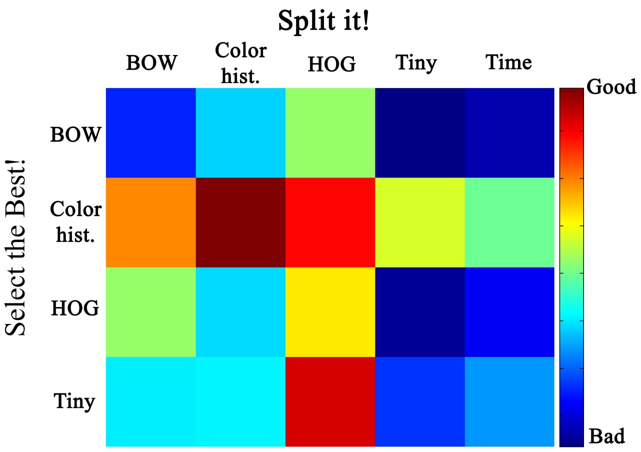

Fig. 3. Comparison between different visual feature. The best result is achieved with ”color histogram” feature for “Split it!” and also for “Select the best!” task.

by ”Color histogram” for both, album segmentation and best photo selection in the segment. Which shows the robustness of the ”Color histogram” features to any kind of images.

Figure 4 shows the distribution of the participants' scores, including the choice of the proposed method and the automatic visual analysis. All scores are sorted in a descending order. These 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, this approach achieves 80% of the best score for each dataset, which proves the concept of the game. It also outperform the automatic visual analysis. 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.

Finally, we discuss user feedback related to the game and automatic summarization as shown in Table 1. The main disadvantage of this game is that the user has to wait maybe several days for the generated album collage, however we showed that it outperforms the automatic analysis. The main advantage of the game is that the user enjoys it and they receives interesting feedback from his friends about his photo albums.

	Epitome game	**Automatic**
Performance	+	-
Processing time	-	+
Fun	++	+
Friends involvement	++	-

Table 1. Comparison between automatic visual summarization and Epitome social game.

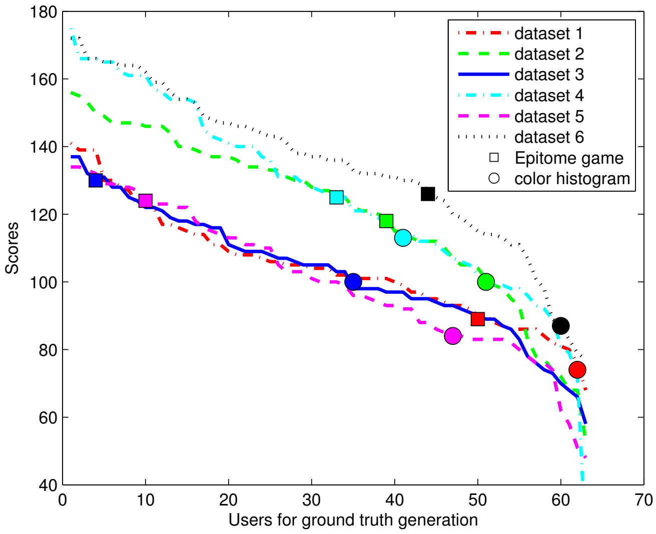

Fig. 4. The distribution of the participants' scores. The results of the proposed method are shown with square markers and the results of automatic visual analysis with circle marker. Different colors of the markers correspond to different datasets. The results are promising and prove the concept of the approach.

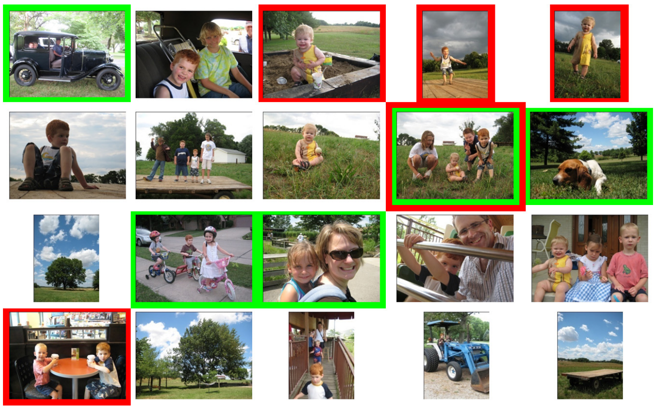

Fig. 5. Photos from the dataset 3. The most representative photos selected by the proposed method are marked with green bounding box, while the red bounding box denotes photos selected by making use of color histogram.

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

In this paper, we analyzed a social games for an album summarization on Facebook. The proof of concept of these games was demonstrated and validated through a set of experiments on several photo collections. The results of our experiments show that the summarization game achieves 80% of the best score of different participants and significantly outperforms automatic visual summarization methods.

As a future study, we will include in this approach more sophisticated visual analysis and make the game more attractive for users.

6. REFERENCES

[1] Ivan Ivanov, Peter Vajda, Jong-Seok Lee, and Touradj Ebrahimi, "Epitome – a social game for photo album summarization," in *Proceedings of the First ACM International Workshop on Connected Multimedia*, 2010.

[2] Susumu Harada, Mor Naaman, Yee Jiun Song, QianYing Wang, and Andreas Paepcke, "Lost in memories: Interacting with large photo collections on PDAs," in *Proc. of the JCDL*, 2004, pp. 325–333.

[3] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina, "Automatic organization for digital photographs with geographic coordinates," in *Proc. of the JCDL*, 2004, pp. 53–62.

[4] B.C. Atkins, "Adaptive photo collection page layout," in *Proc. of the ICIP*, 2004, pp. 2897–2900.

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

[6] *A Bayesian hierarchical model for learning natural scene categories*, vol. 2, Washington, DC, USA, June 2005. IEEE Computer Society.

[7] J. Zhang, M. Marszalek, 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. 213238, 2007.

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

[9] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded-up robust features," in *Proceedings of the 9th European Conference on Computer Vision*, 2006, pp. 404–417.

[10] P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2008.

[11] Antonio Torralba, Rob Fergus, and William 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, pp. 1958–1970, 2008.

[12] Morgan Ames and Mor Naaman, "Why we tag: Motivations for annotation in mobile and online media," in *Proc. of the SIGCHI*, 2007, pp. 971–980.

[13] Luis von Ahn and Laura Dabbish, "Labeling images with a computer game," in *Proc. of the SIGCHI*, 2004, pp. 319–326.

[14] Luis von Ahn, Ruoran Liu, and Manuel Blum, "Peekaboom: a game for locating objects in images," in *Proc. of the SIGCHI*, 2006, pp. 55–64.

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