Supporting Personal Photo Storytelling for Social Albums

Pere Obrador, Rodrigo de Oliveira, Nuria Oliver
Telefonica Research
Via Augusta 177, 08021, Barcelona – Spain
{pere, oliveira, nuriao}@tid.es

ABSTRACT

Information overload is one of today's major concerns. As high-resolution digital cameras become increasingly pervasive, unprecedented amounts of social media are being uploaded to online social networks on a daily basis. In order to support users on selecting the best photos to create an online photo album, attention has been devoted to the development of automatic approaches for photo storytelling. In this paper, we present a novel photo collection summarization system that learns some of the users' social context by analyzing their online photo albums, and includes storytelling principles and face and image aesthetic ranking in order to assist users in creating new photo albums to be shared online. In an in-depth user study conducted with 12 subjects, the proposed system was validated as a first step in the photo album creation process, helping users reduce workload to accomplish such a task. Our findings suggest that a human audio/video professional with cinematographic skills does not perform better than our proposed system.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human factors; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Indexing methods*

General Terms

Algorithms, Human factors

Keywords

Automatic photo storytelling, time clustering, face clustering, image aesthetics, social networks, user study.

1. INTRODUCTION

In recent years, and mainly due to the pervasiveness of digital cameras and camera-phones, there has been an exponential increase in the overall number of photos taken by users. This dramatic growth in the amount of digital personal media has led to increasingly large media libraries in local hard drives and/or online repositories, such as Flickr!, Picasa Web Album or Facebook. Unfortunately, large photo collections turn the manual task of selecting images into a tedious and time consuming process [9, 34]. In addition, the familiarity that users have with the photos belonging to a specific event will decay over time [34], turning the photo selection task more difficult with time.

On the other hand, the social narrative use of photos – *i.e.*, social photo storytelling – plays an important role in people's lives as it serves to structure and share personal and interpersonal experiences and to express personal and group identities [12]. Hence, it does not come as a surprise that automatic approaches to personal photo collection summarization and event detection have recently been of interest in the research community [7, 10, 19, 21, 24]. Unfortunately, none of these approaches addresses social aspects of these photo stories, such as its target audience.

Fully automatic personal photo collection summarization for storytelling purposes is a very hard problem, since each end-user may have very different interests, tastes, photo skills, etc. In addition, meaningful and relevant photo stories require some knowledge of the social context surrounding the photos [19], such as who the user and the target audience are. Hence, we believe that automatic summarization algorithms should incorporate this information. There has been some work in related areas that take advantage of the user's social context, for instance, in [3] an algorithm is presented that improves a multimedia browser based on social metadata – *i.e.*, places the users spend time at, and people they meet with – obtained via GPS traces of daily life routines. More recently, Loui *et al.* [18], have presented an image value assessment algorithm that takes into account social relationships between detected people in the photographs, where a higher weight is given to photos of close relatives and lower weight to the photos of, for instance, neighbors. Unfortunately the social relationships need to be entered manually by the user.

With the advent of photo and video capabilities in online social networking sites (OSN), an increasing portion of the users' social photo storytelling activities are migrating to these sites, where friends and family members update each other on their daily lives, recent events, trips or vacations. For instance, FaceBook is the largest online repository of personal photos in the world with more than 3 billion photos being uploaded monthly[1]. Hence, there are opportunities to mine existing photo albums in OSN in order to automatically create *relevant* and *meaningful* photo stories for users to share online.

In the same spirit as the work in [3, 18], we propose in this paper a photo storytelling system that leverages infor-

MM'10, October 25–29, 2010, Firenze, Italy.

[1] http://www.facebook.com/press/info.php?statistics

mation from the user's OSN photo albums in order to create a *personalized* photo story, *i.e.*, a photo story adapted to the user's style and target audience.

As shown in Section 5 and in previous research [16], users typically enjoy the creative process involved in photo story creation and they rely heavily on emotional and contextual information in order to select images [19]. Therefore, we hypothesize that the proposed system should be seen as the first component of an iterative, incremental loop based on a construct, examine and improve cycle [11], which leads to the final story to be shared. In other words, by starting from a *half baked* story or *draft*, the user would be *satisficing*[2] rather than *optimizing* the full story creation process from scratch [4]. In our user study described in Section 4, we corroborate this hypothesis. We also hypothesize that a human audio/video (A/V) professional with storytelling skills performs better than the proposed system, which we could not validate in our study.

The rest of this paper is structured as follows: in Section 2 we present a review of related work in automatic photo event detection and photo collection summarization for storytelling. Section 3 describes the proposed photo storytelling system. In Section 4, we describe our user study and its results, whilst a few implications for the design of multimedia storytelling applications are summarized in Section 5, followed by our conclusions and lines of future work in Section 6.

2. RELATED WORK

Most of the prior art related to our proposed approach relies on the information extracted from the photos to process only – either a personal collection, or a set of images retrieved from the web – by segmenting them into events, either for collection navigation or summarization, in which case representative images are selected from those events.

In [25], Platt presents a simple time clustering algorithm which starts a new cluster if a new photo is taken more than a certain amount of time since the previous photo was taken. Clusters are merged based on content analysis until the desired number of clusters is reached. This algorithm was improved in [26], by means of an adaptive temporal threshold and a new approach to select the representative image of each cluster (the most distinctive image in the Kullback-Leibler divergence sense). Loui *et al.* present in [19] an automatic albuming system in which collection summarization is performed by event detection using time clustering and sub-event clustering based on color similarity; in addition, very low quality images – with underexposure, low contrast and camera de-focus – are discarded. In [10] a browsing interface is presented that summarizes photo collections by exploiting the capture time information in an adaptive way, similar to [26]; the allocated space for each event is roughly proportional to the number of photos taken in that cluster, and the representative images for each event are selected by identifying very close or very distant images in time. In [24] a scalable image collection representation is created by iteratively traversing time clusters and selecting the most relevant image from each –*e.g.* photos of important faces, photos of appealing imagery. Naaman *et al.* [21] present a system that utilizes the time and location information – *i.e.*, GPS coordinates– to automatically organize a personal photo collection in a set of event and location hierarchies.

Additional unsupervised approaches have been proposed for event clustering [7] – using either temporal similarity or temporal and content similarity quantified at multiple temporal scales, and also for photo storytelling [15]. In the latter, semantic keywords are extracted from the story and an annotated image database is searched. Unfortunately, users are typically reluctant to annotate images with text [27], and therefore such a system may not be suited to generate personal photo stories. Finally, there has also been some work in web (*i.e.*, Flickr) multiuser collection summaries. For instance, Simon *et al.* [30] have recently proposed a solution to the problem of landmark summarization, *i.e.*, the Pantheon in Rome. They use multi-user image collections from the Internet, and select a set of canonical views – by taking image likelihood, coverage and orthogonality into account – to form the scene summary.

Unfortunately, none of the previous work approaches address the social aspects of the photo stories, such as their target audience. Given all previous work, the main contributions of this paper are three-fold: (1) We leverage information from the users' photo albums in their OSN, in order to create a personalized and relevant social photo collection summary or album; (2) we propose a novel *photo collection summarization system that takes into account storytelling principles* (*acts*, *scenes*, *shots* and *characters*) and *image aesthetic measures*, including a 2-category image aesthetics metric, for image selection; and (3) we carry out an *in-depth user study* to validate empirically the proposed system.

3. STORYTELLING FOR SOCIAL ALBUMS

The proposed photo summarization system is inspired by principles of dramaturgy and cinematography. Each generated summary, album or photo story[3] contains a set of elements that are first described in this section, followed by a detailed description of the algorithms that compose the proposed system.

3.1 Photo Story Elements

A good story includes essential elements such as a certain narrative structure, with identifiable beginnings, middles and ends, and a substantial focus on characters and characterization which is arguably the most important single component of the story [17]. In the case of personal photo storytelling, many times users want to show off their experiences [9] – emphasizing good/happy times with friends and family, and aesthetic imagery [16].

3.1.1 Narrative Structure

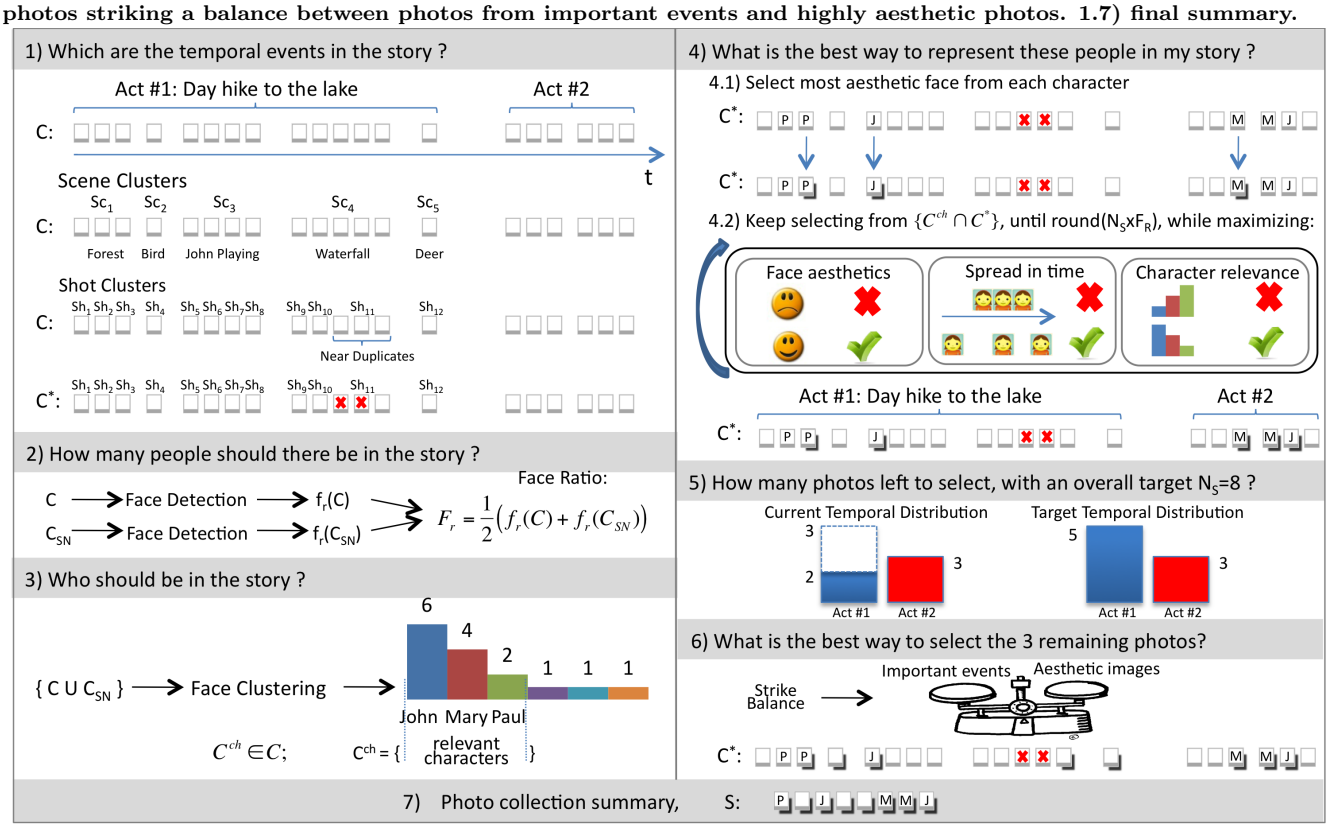
In our approach to storytelling, the photos are grouped into meaningful events, which will generate a certain narrative structure. We divide the story into a three level hierarchy of *acts*, *scenes* and *shots* – see Figure 1 part 1. Since this three level hierarchy provides a good level of granularity, we use relatively non-sophisticated – and hence faster – clustering methods to detect the *acts*, *scenes* and *shots* and still obtain a good overall performance.

1. *Acts*: An *act* is major section of a play (dramaturgy), in which all story elements are related to one another; in our application this can be seen as a relatively large group of photos representing a well defined period in time. Users typically give some structure to their image collection by the temporal patterns (*i.e.*, bursts) with which they take their photos [10]. Hence, *acts* are detected by an algorithm similar to that in [25], where a photo is included into a new *act* if it was captured more than a certain amount of time T_t since the previous photo was captured. This allows us to

[2]Satisficing is a stopping rule for a sequential search, where an aspiration level is fixed in advance, and the search is terminated as soon as an alternative exceeds that level.

[3]In the following, we shall indistinctively refer to photo stories, albums or summaries.

Figure 1: The photo selection process (See section 3.1.5 for notation explanation): 1.1) Event clustering; 1.2) number of images with faces; 1.3) face clustering; 1.4) *people photos* selection, striking a balance among face aesthetics, spread in time and *character* relevance; 1.5) slot allocation for remaining photos to be selected; 1.6) selection of remaining photos striking a balance between photos from important events and highly aesthetic photos. 1.7) final summary.



target a specific number of *acts* just by varying T_t, which is an important feature as explained below.

The number of *acts*, $N_{ActClusters}$, into which the photo collection should be partitioned will depend on the average number of images per *act* $\overline{N_{Act}}$, and the overall number of images in the collection N_C: $N_{ActClusters} = \frac{N_C}{\overline{N_{Act}}}$. Given N_C and $\overline{N_{Act}}$, the proposed *act* clustering algorithm will vary the time threshold T_t until $N_{ActClusters}$ is reached.

2. *Scenes*: Each *act* within a photo story is divided into *scenes*, in which the setting is fixed. In our algorithm a *scene* is composed of images from one specific *act* that are *similar* to each other, using global color similarity.

3. *Shots*: Finally, each *scene* is divided into *shots* – borrowing now from cinematography – which are single video sequences captured by one single camera without interruption. Each consecutive frame in a video *shot* is almost identical to the previous one, and therefore we use this term in our algorithm to refer to a set of near-duplicate photos – *i.e.*, images that were taken from almost the same camera angle, presenting almost identical foreground and background.

Note that we follow a bottom-up approach to accomplish a hierarchical *scene/shot* representation. First, similar images within a specific *act* are clustered into *shots* using the normalized SIFT [20] feature similarity function described in [30]. Next, only *one representative* image from each *shot* is selected using an aesthetic measure (see Section 3.1.3 below). All the representative pictures selected at the *shot* level are then clustered together using a global color similarity function (the normalized histogram intersection in HSV – hue, saturation, value – color space [31]), generating the *scenes* for this particular *act*. Also note that a specific *shots* or *scenes* may be composed of one single image.

In addition, time and content have been previously combined in order to improve event clustering with a similarity measure that linearly relies less on content-based similarity as the inter-photo capture time difference grows [7]. Inspired by this approach, we propose a similarity threshold that varies linearly with the difference between the photo's capture time, ΔT. The similarity threshold is low for $\Delta T \ll \Delta T_{max}$, *i.e.*, images taken close in time and similar to each other will be clustered together, whereas for $\Delta T \geq \Delta T_{max}$, the similarity threshold is 1, *i.e.*, similarity does not apply. As opposed to [7], where the time difference cap ΔT_{max} is set to 48 hours, we make it adaptive to the time duration of the *act* at hand: $\Delta T_{max} = \frac{1}{3} ActDuration$. We apply this approach to both *scene* and *shot* clustering.

3.1.2 Characters

The *characters* in the story are probably its most important element [17]. Hence, it is not surprising that users tend to be very sensitive to the selection of *characters* in their social photo stories. For photo albums to be shared on OSN, users tend to prioritize photos with members of the network.

Our system takes into account three *character* related features: (1) *Face ratio*: the proportion of images with people, *people photos*[4], that should appear in the story; (2) *characters*: who should be the people in the pictures; and (3) *aesthetics*: the aesthetic value of the *characters*' faces in

[4]In the rest of the paper, we shall refer to images with people in them as *people photos* or *images*.

the photos where they appear, including whether they are smiling or not.

Since the goal of our system is to help users create photo stories that will be shared on their OSN, we use two sources of information to determine the target *face ratio* and the *characters* in the story: The specific photo collection to be summarized (C) *and* the set of photos in the user's OSN albums (C_{SN}). This allows our system to approximate the user's style – *i.e.*, average face ratio in an album, which we found to be a very personal trait (see Table 1 in section 4.1) – and adapt to the target audience – *i.e.*, friends that appear prominently in the user's albums are probably socially closer, and should therefore be favored in future summaries.

The *face ratio* is given by the ratio of number of *people photos* in a collection when compared to the total number of photos in that collection. Since different photo collections do not necessarily have the same face ratios, *i.e.*, the user may have lots of people images in one collection and barely any in another collection, the target *face ratio* in the photo story, F_r, is given by a linear combination of the face ratios in C ($f_r(C)$) and in C_{SN} ($f_r(C_{SN})$): $F_r = \frac{1}{2}(\, f_r(C) + f_r(C_{SN}))$ –see Figure 1 part 2. In this way we reach a compromise between the user's social storytelling style and the actual collection to summarize.

In addition, a specific photo collection to be summarized does not necessarily include photos from all the people that are relevant to the user (*e.g.* family, friends). In order to identify the main story *characters*, we combine C and C_{SN} into a single photo collection $\{C \cup C_{SN}\}$, which we use to identify the user's *character* set by clustering the faces using a face detection and recognition engine based on [14] –see Figure 1 part 3. Each face cluster that has at least two images is considered relevant enough to correspond to a *character* important to the user. This gives a good estimation of the people the user cares about. For instance, one of these relevant people may appear only once in C but many times in C_{SN} and hence our system would include that person as a *character* in the summary. In addition, we infer the *importance* of the *characters* from the number of images in each face cluster.

Finally, the aesthetic value of the faces in *people photos* is also computed as described in Section 3.1.3 –see Figure 1 part 4.

3.1.3 Aesthetics

As previously mentioned, users typically share images of important events, relevant *characters*, or images that may be important to them mainly for aesthetic reasons [16]. In addition, if a low quality photograph is selected to summarize an event, it will not be a mnemonic for the user to remember that event [26]. Prior work in computational image aesthetics has focused on automatically classifying aesthetic *vs.* non-aesthetic images [8]. However, in the case of image selection it makes more sense ranking the images within a cluster rather than classifying them. Hence, we use a regression-based computational image aesthetics algorithm based on [23]. Our system also includes regression-based computational face aesthetics algorithm, since it has been shown that different image categories would benefit from different aesthetic metrics [22], and the best high level categorization regarding aesthetics is usually obtained by partitioning the set into *people* and *non-people* photos[5] [5].

a. Face Aesthetics.

There has been some research in trying to understand facial attractiveness [32] using face features including symmetry. Unfortunately, these type of approaches would favor a *character* over another based on their looks, which would go against the storytelling principles described above. In order to avoid this kind of bias, we have used a normalized face aesthetic measure (A_f) that takes into account normalized face sharpness, combined with the relative size of the face with respect to the overall image size [23], and smile detection [33]. This face aesthetic measure turns out to be very effective when comparing aesthetics of the same *character*'s face, *i.e.*, within the same *character*'s face cluster. For the rest of the images with faces, but no *characters* in them, the algorithm rates the aesthetics of the largest face in the photo, since smaller faces might not be relevant or could have been photographed accidentally.

b. Image Aesthetics.

As previously explained, different methods of selecting representative images from within an image cluster have been proposed in the literature [10, 30, 26]. In this work, we take a similar approach to [24] where the representative images within a specific event cluster will be selected based on their aesthetic value [16], and images within a cluster will be ranked based on their aesthetic value as given in [23]. This algorithm measures aesthetics of an image c_i, *i.e.*, $A(c_i)$, on a region by region basis, and takes into account sharpness, contrast, colorfulness and exposure. For compositional purposes the algorithm also measures how well the most appealing region is isolated from the background, and its relative size. The output is normalized between 0 – lowest aesthetic appeal – and 1 – highest aesthetic appeal.

3.1.4 Visual Variety or Diversity

Each summarized *act* should present enough photo variety so as to allow the user to indulge in as many different aspects of the story as possible: relevant people and moments combined with aesthetically beautiful images [16]. Therefore, the photo selection algorithm presented in the next section takes into account these three elements: relevant people and events together with aesthetically beautiful images.

Before delving into the details of our approach, we shall summarize the rest of the notation used in the paper.

3.1.5 Notation

A photo collection C is formed of $N_C = |C|$ images (c_i) in capture time order[6]: $C = \{c_i,\ 0 \leq i < N_C\}$. The photo summary, S, and the collection of photos available in the user's OSN, C_{SN}, are similarly defined.

We shall define next two subsets of C, C^{ch} and C^*:

(1) C^{ch}, which is the subset of C with all the photos that have *characters* in them. It is represented as a collection of M *characters*, or face clusters, that are obtained from the combined set $\{C \cup C_{SN}\}$. Note that some of the clusters might be empty if there are no photos in C where a particular *character* appears – *i.e.*, he/she only appears in C_{SN};

(2) C^*, which is the subset of C that contains no near-duplicate photos, *i.e.*, in C^* all *shots* contain only one image.

As previously explained, C is subdivided into of a series of *acts*, each *act* into a series of *scenes*, and each *scene* into a series of *shots*: $Act = \{Act_i,\ 0 \leq i < N_{ActClusters}\}$, where $N_{ActClusters}$ is the number of *acts* in C. *Scenes* and *shots* are similarly defined.

One of the constraints that we impose on the photo summary S to be created is to preserve the temporal distribution

[5]Note that we will consider photos to be *people* photos if they have at least 1 face detected by the face detection algorithm.

[6]From now on, all our representations of image or event lists will be in capture time order since it is the most common way of ordering personal photos [27].

of photos – characterized by normalized histograms – in *acts*, *scenes* and *shots* of the original collection C, where:

$$H_{Act}(C) = \left\{ \frac{N_{Act_i}}{N_C},\ 0 \leq i < N_{ActClusters} \right\}$$

is the histogram of *acts* in collection C. $H_{Scene}(C)$ and $H_{Shot}(C)$ are similarly defined.

Finally, the generated summary should also approximate the user's *character* normalized histogram, $H_{Character}(C \cup C_{SN})$, while maximizing aesthetics and variety, as explained in the following section.

3.2 The Photo Selection Algorithm

Given a particular user, his/her online photo collection C_{SN} and a photo collection to be summarized C, the goal of the photo selection algorithm is to generate a photo summary S, from C, that contains a pre-defined number of photos, $N_S << N_C$, and conveys the essence of the story to be shared by the user on his/her OSN. This is achieved by ensuring that the photo summary satisfies the following requirements:

1. *People vs. non-people*: The summary's *face ratio* $f_r(S)$ should approximate the target *face ratio* F_r;
2. Characters: $H_{Character}(S) \approx H_{Character}(C \cup C_{SN})$. Note that we use C instead of C^* because near-duplicate photos of *characters* informs us of their importance.
3. Narrative: $H_{Act}(S)$, $H_{Scene}(S)$ and $H_{Shot}(S)$ approximate $H_{Act}(C^*)$, $H_{Scene}(C^*)$ and $H_{Shot}(C)$ respectively. In this case, C^* is used for *acts* and *scenes* because it better represents their event distribution.
4. Aesthetics: High normalized aesthetic value of the summary (A_S), considering both aesthetics of faces in *people photos*, and *non-people photos* image aesthetics.
5. Variety: The selected images are visually diverse.

In order to satisfy the previous requirements, we carry out a two-step process: First select the *people photos* that will appear in S (step 1 below), and then select the rest of images up to N_S images (step 2 below). Both steps are greedy algorithms.

3.2.1 Step 1: People Photo Selection

The goal of this first step is to add to S all the needed *people photos* by selecting $N_S^f = round(N_S \times F_r)$ faces from C^*, according to the following algorithm:

1.a. Rank the face clusters in $\{C \cup C_{SN}\}$ by number of images. Select the image with the most aesthetic *character* face that belongs to $\{C^{ch} \cap C^*\}$ from each of the face clusters – starting from the largest cluster in the rank, which ensures coverage of relevant *characters* in the story while avoiding near-duplicates –see Figure 1 part 4.1. If the image has already been selected – *i.e.*, there are two or more *characters* in the same picture, pick the following image in the aesthetically ordered list from the most popular *character* of the two – *i.e.*, largest histogram bin in $H_{Character}(C \cup C_{SN})$, and so on and so forth.

1.b. Keep selecting face images from $\{C^{ch} \cap C^*\}$ while $|S| < N_S^f$ and maximizing the objective function O_f, see Figure 1 part 4.2:

$$O_f(C, C^*, S, C_{SN}) = \alpha_f\ A_f(S) - \gamma_f\ d(H_{Character}(S), H_{Character}(C \cup C_{SN}) - \delta_f\ d(H_{Act}(S), H_{Act}(C^*))$$

where $A_f(S)$ is the normalized aesthetic value of the considered face in the people images in the summary, and $d(.)$ is the normalized L1 distance metric between histograms. More importance is given to the *character* histogram distance ($\gamma_f = 1$), followed by the face aesthetic value ($\alpha_f = 0.8$), and the *act* histogram distance ($\delta_f = 0.5$). This last term is important to ensure a certain amount of temporal coverage by the *characters*, since images with highly aesthetic people faces may be confined to specific *acts* – *i.e.*, better *vs.* worse light conditions at different times of the day. Note that if not enough *character* images are present, then the photos with most aesthetic faces of other people will be selected. If there are not enough *people photos* in the collection, *i.e.*, $N_S^f < round(N_S \times F_r)$, the algorithm moves on to step 2.

3.2.2 Step 2: Non-People Photo Selection

The previous step has selected the first N_S^f images of S. Now the algorithm will select the rest of the images ($N_S - N_S^f$) from C^*.

From here on, we define a large *scene* (*L-scene*) or large *shot* (*L-shot*), as *scenes* or *shots* with at least 3 images, which ensures the importance of those sub-events, and avoids potentially noisier smaller clusters.

2.a. Similar to [10] and in order to ensure good temporal coverage of all *acts*, we start by ensuring that each *act* has had one representative image selected in step 1 above. If not, we allocate one image slot, out of the ($N_S - N_S^f$) available empty slots, for each of the empty *acts*. If not enough empty slots are available, then the larger *acts* are favored.

2.b. Next, we optimally allocate the rest of the empty image slots to each *act* –see Figure 1 part 5– by minimizing:

$$O_a(C^*, S) = d(H_{Act}(S), H_{Act}(C^*)) \quad subject\ \ to\ \ |S| = N_S$$

For each Act_i in C, we keep selecting images until Act_i has all its empty image slots filled. Similar to [19, 24], the images are selected based on their aesthetic value. The algorithm alternates between *L-shots* or *L-scenes* and highly aesthetic images in order to provide good selection variety, as well as never selecting more than one representative image from a particular *scene* – see Figure 1 part 6:

- 2.b.1. Select the most aesthetic image from the largest unrepresented *L-shot* from an unrepresented *scene* in Act_i. If not available, then select the most aesthetic image from the largest unrepresented *L-scene* in Act_i. If not available, move to the following step.
- 2.b.2. Select the most aesthetic image in Act_i from any of the unrepresented *scenes*.
- 2.b.3. Go to 2.b.1.

We found that giving higher relevance to the largest *L-shot* is important since they usually represent the same object or landscape portrayed from the same viewpoint, implying a certain level of relevance for the user [30]. Conversely, highly aesthetic images tend to appear in smaller clusters or alone, and hence the alternate search for relevant and aesthetic images. Finally, all selected images are reordered chronologically before being presented to the end user. In the following sections, we describe an in-depth user study targeted at understanding the advantages and disadvantages of the proposed photo storytelling system.

4. USER STUDY

We designed and carried out a user study to assess the strengths and weaknesses of the proposed system. In order to motivate the need for an automatic storytelling approach, we also wanted to investigate if users consider the task of creating a photo story to be laborious and time consuming. Previous work supports this assumption [9] and confirms that photo retrieval is neither a fast nor an easy task [34]. With our study, we wanted to verify whether users perceive the effort and time demand associated with the photo storytelling task as the main source of workload instead of their concern to create a good story. In other words, if photo storytelling is too hard and users are not that demanding with the final results, an automatic approach could be appropriate. Hence, we formulate our first hypothesis as:

$\mathbb{H}1$ Users consider the task of creating personal photo stories to be laborious and time consuming, and this effort is more important than their concern to create a good photo story.

With respect to the proposed system, we wanted to evaluate the assumption that its storytelling features improve the quality of the automatically generated photo stories when compared to a simple automatic approach. Hence, the second hypothesis is formulated as:

$\mathbb{H}2$ Users prefer personal photo stories generated automatically by the proposed system *more often than* those generated by a random selection of photos in chronological order.

Furthermore, we carried out a comparison between the stories generated by the proposed system and a human expert in A/V photo story creation. Clearly, a human A/V professional can better filter photos based on aesthetics than our proposed system, such as removing high quality photos where the main *character* looks fat or with the eyes closed. Moreover, (s)he can use his/her storytelling skills to combine photos that were taken with different timestamps but at the same place, or even select photos that compose stylish stories with an artistic taste. Therefore, the third hypothesis is thus formulated as:

$\mathbb{H}3$ Users prefer personal photo stories created by a human A/V professional –who creates video/photo stories for a living– *more often than* those generated by the proposed system.

Our intention with hypothesis $\mathbb{H}3$ is to test whether an automatic approach that is aware of the user's photo sharing patterns can achieve a performance level similar to a human that does not take this information into consideration.

Finally, we believe that the photo stories generated by our system can be appreciated by users as an initial draft instead of creating the entire photo story by themselves from scratch. Therefore, we state our final hypothesis as:

$\mathbb{H}4$ Users prefer to reuse a personal photo story generated by the proposed system and upload it to their OSN after making the appropriate changes instead of creating the photo story from scratch.

Note that $\mathbb{H}4$ introduces the goal of sharing photo stories in social networks. Therefore, we asked participants in the initial questionnaire if they agree with this assumption.

In order to validate these hypotheses, we conducted two lab studies to: (1) measure the workload associated to the photo story creation process and (2) obtain the users' level of satisfaction with photo stories generated automatically and by the A/V professional. The next sections describe in detail the user study design and discuss the main results.

Table 1: Profile of the participants and their stories. Where *Face Ratio* is $f_r(C_{SN})$, and *#characters* is the number of characters with representation in C^{ch}.

Subj.	Sex	Age	Story Title	Story Relev.∗	Face Ratio	#characters
1	F	28	Peru	4	45.1%	8
2	F	25	Mexico	5	84.1%	8
3	M	29	Rome	5	46.5%	9
4	M	32	California	4	18.5%	3
5	F	26	Mjøsa Lake	3	47.7%	3
6	M	23	Rome II	2	57.3%	9
7	M	33	Sardinia	4	31.8%	7
8	M	37	Russia	3	14.6%	1
9	F	29	Bolivia	5	57.8%	6
10	F	33	Calabria	3	68.9%	7
11	M	30	Nepal	2	40.0%	5
12	F	31	Seattle	5	59.4%	5

∗ Relevance rated by the subjects (*Not important:* 1, to *very important:* 5).

4.1 Participants

Twelve subjects (male: 6) were recruited via e-mail advertisement inside a large company. Subjects were considered eligible if they had an account on at least one OSN, were currently sharing photos with peers, and had at least 200 photos in their personal repository from a specific event that they would be willing to use during the user study (*e.g.* a vacation trip, a wedding party, a night out). Each participant was offered 20 Euro (about 27 USD) to be part of the experiment and a prize of 100 Euro (about 135 USD) was raffled among all of them. Mean age was 30 years old ($s = 3.89$) and occupations were somewhat varied, including students, researchers, software developers, a technology expert, a professional from human resources, a secretary, and a teacher. Participants self-assessed their photo shooting skills as slightly below average (1:novice, 5:expert, $\tilde{x} = 2.5, q1 = 2, q3 = 3$), and their ability to differentiate photos by image quality – *e.g.* contrast, sharpness, composition – as average ($\tilde{x} = 3, q1 = 3, q3 = 4$). They typically took photos from one to three times per month.

All participants had a Facebook account and 92% considered it as their main OSN, which they used to access at least two days per week ($\tilde{x}$ = every day). On average, they had 13 online photo albums each ($s = 9.746$), 36 photos per photo album ($s = 18.17$), and about 100 photos per folder in their personal collection ($s = 63.66$). These setting are similar to those from the experiment conducted by Kirk *et al.* [16]. Table 1 characterizes the profile of the participants and the event associated to the 200 photos that they lent to the user study. Note that most of the collections were about trips and holidays, which are a common source for storytelling between friends and family [16].

In addition to the 12 participants, an A/V professional with storytelling skills was recruited with the goal of creating one photo story for each of the participants' collections.

4.2 Apparatus

Photo stories were created and evaluated by the participants using the same apparatus, including a 21.5 inch flat panel monitor with a resolution of 1680x1050 pixels, a standard mouse with a scrolling wheel button and a keyboard. The Windows Explorer application (Windows Vista version) was used by the participants to create photo stories (see Section 4.3.1). Interviews were audio recorded.

4.3 Procedure

The lab study was divided in two trials. In the first trial, participants created a photo story using photos from the personal collection they lent to the user study. Workload was measured both objectively and subjectively to shed some light on $\mathbb{H}1$. In the following week, each subject attended the second trial in which they were presented with photo stories generated automatically and by the A/V professional, and were asked to evaluate how good they were. This procedure was adopted to provide answers to $\mathbb{H}2$, $\mathbb{H}3$, and $\mathbb{H}4$. Next, each trial is explained in more detail.

4.3.1 First Trial: Workload Measurement

Workload is an individual experience and therefore very hard to be quantified effectively in different activities by different subjects. In the first trial, we used the NASA Task Load Index [2] to subjectively measure workload in a storytelling task, thus gathering information in six different dimensions: mental demand, temporal demand, physical demand, performance, effort, and frustration level. Furthermore, we also measured task completion time and logged the participants' interactions with the interface, including keystrokes, mouse clicks, mouse moves and mouse scrolling.

All participants were assigned the task of creating one story of 20 photos – from their initial pool of 200 – that they would be willing to share on their main OSN. The Windows Explorer application was used by the participants to create the photo stories for three main reasons: First, participants were familiar with it, thus reducing the interaction learning curve; second, it implements all interactions available on the Facebook Photo Album webpage (*i.e.,* photo selection, photo maximization, and *drag'n'drop*); and third, its popularity eases replication of the study by the scientific community. Figure 2 shows an example of the interface used by the participants.

Note that the pool size of 200 photos matches the total number of photos that one could share on a Facebook[7] photo album [1] and is also an approximation of the total number of photos per event considered in the work by Kirk *et al.* [16]. Moreover, the 20-photo story might have imposed a challenge to some participants as they shared an average of 36 photos per photo album in their main OSN ($s = 18.17$). As a consequence, if these participants were willing to share the automatic 20-photo generated stories, instead of manually creating them (with on average 36 photos), we would be reducing information overload by 44%.

After accomplishing the storytelling task, participants filled the NASA TLX questionnaire [2] and commented on the experience. Each session lasted an average of 34 minutes.

4.3.2 Second Trial: Evaluation of Stories

In the second trial, three photo stories with 20 photos each were generated for each collection of 200 photos. The following approaches were used to generate the stories:

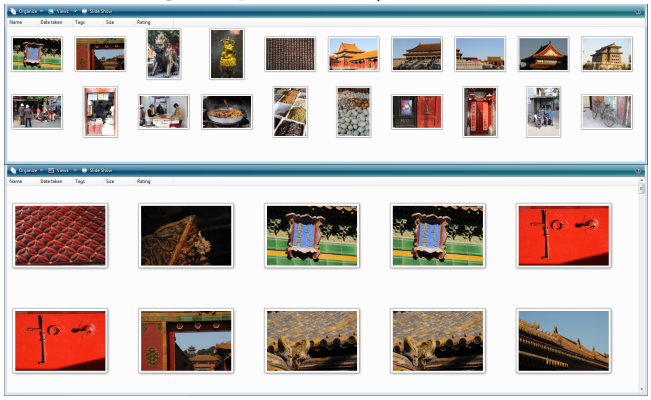
1. **Random:** Photos chosen randomly and presented in chronological order;
2. **System:** Photos chosen and ordered by the proposed system;
3. **Professional:** Photos chosen and ordered by the human A/V professional with storytelling skills, who received instructions to create *appealing* photo stories that best *describe* the titles provided by the participants (see Table 1).

Considering the size of our sample, we opted for a single factorial design where the approach used to generate the photo stories was a within-subjects factor. The second trial was conducted as follows: One week after the first trial, participants attended the second lab session that took an average of 36 minutes per session. First, they browsed their 200 photos to remind themselves of the event and the photos available to compose a story. Next, they were presented with one photo story containing 20 photos from the initial pool of 200 and were asked if they would share it in their social network. This procedure was repeated for the two remaining photo stories. After evaluating each story, subjects were asked to select the stories they would be most comfortable and least comfortable to share on their OSN. Finally, they answered if they would prefer to use the System story as an initial draft to compose the story or if they would rather create the photo story from scratch.

In the beginning of each session, the expert told participants that all photo stories were generated automatically. Deception was used in this case to avoid biasing the results towards either of the approaches. In addition, the presentation order of the photo stories was rotated in a Latin square basis to avoid biasing the results due to the within-subjects design. Finally, participants were gender balanced in each of the presentation ordering groups to avoid gender biases.

[7]Facebook was taken as point of reference because 11 out of 12 participants used it as their main OSN.

Figure 2: Storytelling interface used by the participants. Bottom window contains the initial pool of 200 photos (only 10 thumbnails can be seen at a time, as in Facebook), while the upper window retains the 20 photos that belong to the story created by the participants (ordered from left to right, top to bottom).



4.4 Statistical Analysis

Non-parametric analysis was used for both null-hypothesis testing and measurement of associations/correlations between categorical and ordinal variables. Given that a within subjects design was applied, we used the Friedman test and the Wilcoxon rank sum test to evaluate differences between automatic storytelling approaches. Spearman's Rho (ρ) was used to measure correlations between ordinal related variables and between interval variables that were not homoscedastic (*e.g.* v1: task duration, v2: workload). Finally, associations between categorical variables were evaluated using the Chi-square test, or the Fisher's exact test when more appropriate. The level of significance was taken as $p < .05$.

4.5 Results and Discussion

The results obtained are presented and discussed in this section with the aim of evaluating each of the hypotheses stated in Section 4. The following subsection serves as an introduction to the validation of each of these hypotheses.

4.5.1 Online photo sharing behaviors

When asked about their OSN photo sharing habits, participants reported rarely reordering photos in their shared photo albums (1: never, 5: always $\tilde{x} = 2, q1 = 1.25, q3 = 3$) and mainly using chronological ordering (67% of the participants). This behavior can be explained by their major complains on the difficulties of sharing photos in OSNs:

1. Effort to select photos: This was the most cited issue by the participants. Half of them indicated that there is a lot of effort in this process, including time demand (participants 3, 6 and 9), identification of the most special/appealing photos (participants 4 and 11), and selection of the best photo from the available near-duplicates (participant 5).

2. Delay to upload photos: Four participants mentioned the time demand to upload photos (participants 1, 2, 10, 12), and one of them also pointed the fact that sometimes errors happen in this process.

3. Effort to organize photos: Three participants were concerned about the effort to reorder photos after uploading them (participants 1, 2 and 6). One of the arguments reveals that shared and private photos do not necessarily follow the same organizational schema: *"...you have to reorder them if they are not stored in one single folder before uploading"* (participant 2).

Note that the first and third most cited problems by the participants motivate the study presented herein.

4.5.2 Validation of $\mathbb{H}1$

The task of creating photo stories is laborious and time consuming. As mentioned before, the effort to select/organize photos and the time required to do it are the self-reported main problems that our participants had with online photo sharing. However, participants were neither satisfied nor unsatisfied with current social networks as a means to share their photo stories (1: very unsatisfied, 5: very satisfied, $\tilde{x} = 3, q1 = 3, q3 = 4$). This is somewhat in accordance with our results from the NASA TLX questionnaire, in which the overall workload to create the 20-photo personal story from an initial pool of 200 photos was considered slightly low (0: low, 100: high, $\bar{x} = 35.2$, $s = 3.92, min = 16, max = 58$). Hence, although participants consider the time demand and mental demand in the photo selection process to be relevant problems, they do not seem to be important enough to make them dislike current online related services. These problems were also validated by the analysis of the workload source, in which performance (concern to create a good story), mental demand, time demand, and effort received the highest weights (no significant difference between them: $\chi = 2.798, df = 3, p = .424$). Furthermore, objective workload measures are consistent with these results, given that a strong positive correlation was found between task duration and workload ($\rho = .587, p = .045$). In other words, the longer the task, the higher the workload.

From these results, we reject $\mathbb{H}1$ and rewrite it as:

$\mathbb{H}1_{new}$ Users consider the task of creating personal photo stories to be *mentally* laborious and time consuming, and this effort is *as high as* their concern to manually create a good photo story.

We revisit $\mathbb{H}1_{new}$ after validating $\mathbb{H}4$ to clarify whether the participants' concern to create a good photo story is such that they would persist in doing it by themselves.

4.5.3 Validation of $\mathbb{H}2$

The proposed system performed better than the random approach. A majority of nine participants (or 75%) preferred to share on their main OSN the stories generated by the proposed system instead of the stories by the random approach. Moreover, the difference between the subjects' rankings to the stories generated by these approaches was *significant* ($N = 12; Z = -2.183; p = .029$), thus confirming that the better performance of our system compared to the random approach is not due to chance. After carefully analyzing the participants' comments on the reasons why the Random method generated worse stories, we realized that the key advantages of the proposed system include:

1. Image aesthetics analysis: Several participants wanted to remove photos from the Random story that they considered to have low aesthetic value, *i.e.,* photos that were blurred (participants 2, 3, 4, 6, 8, and 11), had poor composition (participants 3 and 6), were too dark/bright (participant 6), or less colorful (participant 4). Conversely, it was rare the case when participants complained about the image aesthetics of the photos belonging to the System stories (only participants 1 and 6 considered one of the pictures to be, respectively, too bright and out of focus).

2. Balance of photos per act: In the Random stories, participants complained about the absence of photos showing all the places they visited (participants 7, 10 and 12), or the over representation of certain parts of the event: *"This photo I would keep as well... But there are too many of the aquarium... so... I would keep only these two"* (participant 12). There was also the case when participants were presented pictures they did not consider relevant for the event (participants 3, 5, 6, 9 and 10) or even photos they did not remember: *"This one I would remove because... I didn't even know of this, you know? It's a photograph that I almost don't even remember of being part of the 200."* (participant 9). Conversely, the System stories did not include photos from less relevant memories because it balanced the number of photos per *act* according to their relative importance (*i.e.,* ratio of photos taken per *act*, *scene* and *shot*).

3. Near-duplicate detection: Six subjects experienced near-duplicates in their Random story and opted for the one with better image quality (participants 7 and 10), well centered (participants 3 and 6), with their friends (participant 1), or without himself to highlight the landscape (participant 11).

4. Face aesthetics modeling, including smile detection: Poor face aesthetics was mentioned by four participants when analyzing the Random stories. More specifically, participant 2 did not like the way she looked in one of the random photos, participant 7 was concerned with the weird face his son was making, and participants 5 and 9 avoided photos that both themselves and their friends/family were not looking good: *"Oh no! My friend is not going to be happy with that photograph. She looks drunk! And she... and evil! Like, she'd kill me!"* (participant 9). None of the Random stories' rejected photos were selected by the System, with the exception of one from participant 5. Although image aesthetics was good in that photo, the main *character* – her boyfriend – was smiling in a somewhat aggressive way, and thus she would not be comfortable in sharing it with friends.

5. Character selection: By using face detection and generating photo stories with more/less people according to the *face ratio* of the participant's photo collections, our system was able to better adapt to individual preferences –see Table 1. Moreover, the use of face recognition and clustering helped in identifying the most relevant *characters*. This feature was well appreciated: *"This one –System story– is better because there are more photos where I am with my girlfriend."* (participant 3); Conversely, the Random approach had no leverage to opt between photos with different people: *"I don't know these people –Random story. I know I took these photos, but I wouldn't share them in my social network because my friends don't know them."* (participant 5).

Considering the prior observations and the fact that a *significant* difference was revealed between the participants' preference for the System and the Random stories, we corroborate hypothesis $\mathbb{H}2$.

4.5.4 Validation of $\mathbb{H}3$

No evidence was found that the human A/V professional performs better than the proposed system. Seven participants (58%) liked the System stories more than the Professional stories while the remaining five participants (42%) preferred the Professional stories. The difference between these preferences is not significant ($N = 12; Z = -.165; p = .869$) and therefore suggests that participants liked the Professional stories *as often as* they liked the System stories. However, by corroborating this hypothesis we are subject to a significant Type II error due to our sample size. Nevertheless, at least two facts suggest that the Professional approach did not perform better than the System approach: (1) although not significant, the majority of participants preferred the System stories; and (2) the number of participants that would not like to upload the Professional story to their social network –participants 6 and 9– was the same when compared to those that also would not upload the System stories –participants 9 and 11[8]. Therefore, we neither corroborate nor deny $\mathbb{H}3$, but instead highlight the tendency to obtain similar results with both Professional and System approaches.

Furthermore, after carefully listening to the audio recorded sessions, we confirmed that the Professional stories were lacking information regarding the participants photo sharing preferences, which was better captured by System stories:

☺*"This one –System story– is better because there are more photos where I am with my girlfriend."* (participant 3)

☺*"I prefer this story –System– because it gives me a more... warm feeling about it. The combination of colors, the brightness of pictures, and the... there is more people here, but usually I don't like much to have people in my pictures, but maybe for sharing"* (participant 4)

☺*"I like it –System story– because there are more people that I know."* (participant 5)

☹*"This story –Professional– is not exactly what I would do because it lacks pictures with people. The pictures are really nice, but as far as the social network, I am much more interested in looking for some people out there. What I am looking for is some sort of experience."* (participant 6)

☹*"This one –Professional story– is too focused on me, isn't it? Very selfish. I prefer to do it like half and half. For instance, there is no photo with my husband. Oh, no. There's one here. In the end. But he's alone."* (participant 12)

The statements above confirm the importance of the *character* selection technique used by the System stories, which was based on both C and C_{SN}.

4.5.5 Validation of $\mathbb{H}4$

Participants preferred to use the System story as an initial draft instead of manually creating the entire photo story from scratch. Although results from $\mathbb{H}1$ reveal a slightly low effort to manually create the photo stories from scratch, 75% of the participants reported preferring to reuse the System story and make changes to it. Some of the reasons included the good aesthetics of the photos chosen by System (participants 1, 3, 4, 5, 8 and 12), the summarization/arrangement of sub-events (participants 1, 2 and 12), the possibility of reducing the effort associated to the photo selection process (participants 2, 6 and 10), and the lack of time (participants 2, 5, 6, 8, 10 and 12). These observations validate $\mathbb{H}4$ and confirm that both mental and time demands are indeed the sources of the storytelling workload, as indicated by the first trial.

Finally, from the results obtained for $\mathbb{H}4$, a final version of $\mathbb{H}1_{new}$ can be written as:

$\mathbb{H}1_{final}$ Users consider the task of creating personal photo stories to be *mentally* laborious and time consuming, and this effort is *as high as* their concern to ~~manually~~ create a good photo story.

5. IMPLICATIONS FOR DESIGN

The findings presented herein support a few guidelines that might help designers and multimedia technology experts to build social storytelling solutions, including:

Focus on face aesthetics. Seven participants (58%) complained about face aesthetics in the System stories (participants 1, 2, 5, 7, 9, 11 and 12). While some of the reasons might be easily covered in future automatic algorithms (*e.g.* eyes closed detection), others are harder (*e.g.* detection of "weird" smile, goofy face, fat face, *etc.*). Face aesthetics was considered relevant not only for the main *character*, but also for the peers: *"When it comes to people, I draw the line."* (participant 9). We believe participants were concerned with their aesthetics mostly because of the main goal of the task: create a story to share on their social network. Given that self-promotion is one of the main reasons for sharing multimedia content in social networks [6], users will definitely appreciate automatic approaches that highlight photos in which they look better.

Reduce information overload automatically, but support the user's creativity with story customization. Even though the proposed system has proven to be effective, sharing photo stories is a social activity, as exemplified by a comment from participant 2: *"You know, some things you want to share only with your friends, not your family"*. Photo story personalization is key, specially because no one else knows the event captured by the photos better than the users themselves. The validation of hypotheses $\mathbb{H}2$, $\mathbb{H}3$, and $\mathbb{H}4$ confirms that users would benefit from the proposed approach, but would want to control the storytelling task. This is supported by the fact that, from the nine participants that would upload the System story to their OSN, eight would either remove, reorder or swap photos from the generated stories. Note that we considered *personal* photo stories where our participants were the main *characters*. That might have been the reason for the extra motivation to edit the stories. Future work shall evaluate the combination of automatic and manual approaches towards increasing productivity, reducing information overload and supporting the user's creativity.

Combine automatic approaches with collaborative storytelling. The participants preference for the System story instead of the Professional story was strongly associated with their preference to reuse the System draft instead of creating the story from scratch ($\phi = .683, p = .045$). In other words, those who preferred the Professional story also preferred to create the story by themselves instead of reusing any of the generated stories. Hence, our findings suggest that there is a subset of users that do not seem to benefit from automatic solutions, but are more likely to benefit from human-generated ones. Therefore, we expect automatic multimedia storytelling solutions to benefit from a collaborative component in order to better fulfill the users' needs. Previous work has already tackled the problem from

[8]In this case, participants were allowed to consider both removing and reordering any of the 20 photos, but not including others from the initial photo collection, which would require them to browse the remaining 180 photos and thus be exposed to the information overload problem.

a purely collaborative perspective [13, 28, 29], but our results suggest that the combination of automatic and collaborative approaches might lead to a more appropriate balance between increased productivity, information overload reduction and subjective satisfaction with the final story.

6. CONCLUSIONS AND FUTURE WORK

We have presented a novel approach for social photo storytelling which takes advantage of the users' storytelling behaviors by analyzing the images in their OSN photo albums. Some of the key features included in the proposed system include face clustering and a characterization of the percentage of images with faces, event detection and image aesthetics ranking. In an in-depth user study, we have shown that our approach can be of help to users in creating a first draft of a photo album to be shared online and that users can improve in a less demanding way than if starting from scratch. Areas of future work include extending the mining of the OSN photo albums in order to further enhance the personalization of the final album, including photo category detection along with new aesthetic measures tailored to those categories and face expression recognition. We would also like to carry out a longer longitudinal user study to better understand the pros and cons of the proposed approach and identify additional unfulfilled user needs when creating photo albums.

7. ACKNOWLEDGMENTS

We thank all the participants of our user study for their valuable feedback, as well as the A/V professional, Adria Peral, for his help. Telefonica Research participates in the Torres Quevedo subprogram (MICINN), cofinanced by the European Social Fund, for researchers recruitment.

8. REFERENCES

[1] Facebook increases photo album limit to 200 pictures. Retrieved in 4/2010 from http://reface.me/news/facebook-increases-photo-album-limit-to-200-pictures.

[2] Nasa task load index. Retrieved in 4/2010 from http://humansystems.arc.nasa.gov/groups/TLX.

[3] B. Adams, D. Phung, and S. Venkatesh. Extraction of social context and application to personal multimedia exploration. In *Proc. of the 14th annual ACM international conference on Multimedia*, pp. 987–996, 2006.

[4] F. Bentley, C. Metcalf, and G. Harboe. Personal vs. commercial content: the similarities between consumer use of photos and music. In *CHI '06: Proc. of the SIGCHI conference on Human Factors in computing systems*, pp. 667–676. ACM, 2006.

[5] C. Cerosaletti and A. Loui. Measuring the perceived aesthetic quality of photographic images. In *Intl. Workshop on Quality of Multimedia Experience*, 2009.

[6] M. Cherubini, A. Gutierrez, R. Oliveira and N. Oliver. Social tagging revamped: Supporting the users' need of self-promotion through persuasive techniques. In *Proc. ACM Int. Conf. on Human Factors in Computing Systems, CHI'10*, 2010.

[7] M. Cooper *et al.* Temporal event clustering for digital photo collections. *ACM Trans. Multimedia Comput. Commun. Appl.*, 1(3):269–288, 2005.

[8] R. Datta *et al.* Studying aesthetics in photographic images using a computational approach. *Lec. Notes. in Comp. Sci.*, 3953:288, 2006.

[9] D. Frohlich *et al.* Requirements for photoware. In *Proc. of CSCW'02*, pp. 166–175, 2002.

[10] A. Graham *et al.* Time as essence for photo browsing through personal digital libraries. In *Proc. of the second ACM/IEEE-CS Joint Conference on Digital libraries*, pp. 326–335. ACM, 2002.

[11] J. Hayes and L. Flower. *Identifying the organization of the writing process*. Lawrence Erlbaum associates, Hillsdale, New Jersey, 1980.

[12] N. V. House *et al.* From "what?" to "why?": The social uses of personal photos. In *Proc. of CSCW'04*. ACM, 2004.

[13] M. Jones *et al.* "narrowcast yourself": designing for community storytelling in a rural indian context. In *Proc. of the 7th ACM conference on Designing interactive systems*, pp. 369–378. ACM, 2008.

[14] M. J. Jones and P. Viola. Face recognition using boosted local features. *Tech. Report MERL TR-2003-25. Mitsubishi Electric Research Lab.*, 2003.

[15] D. Joshi, J. Z. Wang, and J. Li. The story picturing engine—a system for automatic text illustration. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2(1):68–89, 2006.

[16] D. Kirk *et al.* Understanding photowork. In *Proc. of the SIGCHI conference on Human Factors in computing systems, CHI'06*, pp. 761–770. ACM, 2006.

[17] D. Lodge. *The art of fiction*. Secker & Warburg, London, U.K., 1992.

[18] A. Loui *et al.* Multidimensional image value assessment and rating for automated albuming and retrieval. In *Proc. IEEE Intl. Conf. Image Proc.*, pp. 97–100, 2008.

[19] A. C. Loui and A. E. Savakis. Automated event clustering and quality screening of consumer pictures for digital albuming. *IEEE Trans. on Multimedia*, pp. 390–402, Sept. 2003.

[20] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. on Comput. Vision*, 60(2):910–110, 2004.

[21] M. Naaman *et al.* Automatic organization for digital photographs with geographic coordinates. In *Proc. of the 4th ACM/IEEE-CS joint conference on Digital libraries, JCDL '04*, pp. 53–62. ACM, 2004.

[22] P. Obrador *et al.* The role of tags and image aesthetics in social image search. In *Proc. of the* 1^{st} *SIGMM workshop on Social media, WSM '09*, pp. 65–72. ACM, 2009.

[23] P. Obrador and N. Moroney. Low level features for image appeal measurement. In *SPIE, Electronic Imaging, Image Quality and System Performance VI*, volume 7242, pp. 72420T-1–12. IS&T/SPIE, 2009.

[24] P. Obrador and N. Moroney. Automatic Image Selection by means of a Hierarchical Scalable Collection Representation. In *SPIE, Electronic Imaging, Visual Communications and Image Processing*, volume 7257, pp. 72570W-1–12. IS&T/SPIE, 2009.

[25] J. C. Platt. Autoalbum: Clustering digital photographs using probabilistic model merging. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pp. 96–100, 2000.

[26] J. C. Platt, M. Czerwinski, and B. A. Field. Phototoc: Automatic clustering for browsing personal photographs. *Microsoft Tech. Report*, 2002.

[27] K. Rodden and K. R. Wood. How do people manage their digital photographs? In *Proc. of the SIGCHI conference on Human factors in computing systems*, pp. 409–416. ACM, 2003.

[28] L. Schäfer, C. Valle, and W. Prinz. Group storytelling for team awareness and entertainment. In *Proc. of the third Nordic conference on Human-computer interaction, NordCHI'04*, pp. 441–444, 2004. ACM.

[29] C. Shen *et al.* Sharing and building digital group histories. In *Proc. of CSCW'02*, pp. 324–333. ACM, 2002.

[30] I. Simon, N. Snavely, and S. M. Seitz. Scene summarization for online image collections. *Proc. 11th IEEE International Conference on Computer Vision*, pp. 147–155, 2007.

[31] J. R. Smith and S.-F. Chang. Tools and techniques for color image retrieval. In *SPIE, Electronic Imaging, Storage & Retrieval for Image and Video Databases IV*, pp. 426–437. IS&T/SPIE, 1996.

[32] R. Thornhill and S. W. Gangestad. Facial attractiveness. *Trends in Cognitive Science*, 3:452-460, 1999.

[33] X. wen Chen and T. Huang. Facial expression recognition: a clustering-based approach. *Pattern Recognition Letters*, 24:1295–1302, 2003.

[34] S. Whittaker, O. Bergman, and P. Clough. Easy on that trigger dad: a study of long term family photo retrieval. *Journal Personal and Ubiquitous Computing*, 14,1:31–43, 2010.