

Dynamic Presentation of Synchronised Photo Streams

Sam Zargham
Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, GU2 7XH, UK
s.zargham@surrey.ac.uk

Janko Čalić
Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, GU2 7XH, UK
j.calic@surrey.ac.uk

Due to the proliferation of social media platforms, digital photo sharing has emerged as a new way of establishing chronology of events and reminiscing forgotten experiences. The growing trend of presenting visual content using timelines or shared photo streams offers new possibilities to social media applications. In this paper, a novel interface for visualisation and sharing of multiple photo streams is proposed. A detailed user study into aspects of different temporal presentations used in this interface is presented. In relation to the proposed synchronous visualisation, a novel notion of continuity in photo streams is introduced. A number of continuity detection algorithms are proposed and evaluated. The results of the user study demonstrate good comprehension of user's own and shared photo streams and their temporal structure even if presented at relatively high speeds. Users were able to easily contextualise events, recall specific photos and easily find them using the proposed interface.

Photo sharing, multiple photo streams, visualisation, user studies

1. INTRODUCTION

Digital photo sharing has grown rapidly with the proliferation of capture devices, social networking and file sharing services. In addition to typical photo related activities, such as editing and archiving, new ways of photo browsing and visualisation have emerged as critical aspects of digital 'photoware'. Not only has photo sharing become a means of social communication (Van Dijck (2008)), but also an easy way of establishing a chronology of images for memory, identity, and narrative (Van House (2007)). Furthermore, the emergence of Web services articulating photo sharing through **photo streams** and **timelines**, has highlighted the importance of chronology (Miller and Edwards (2007); Graham et al. (2002)) as the key feature in organising and presenting personal photos. Indeed, the increasing ability we have to document our lives in ever increasing detail, and compare them to others, raises new issues for the design and management of time-related **media streams** in general (Marcus (2004); Davis et al. (2004)).

For example in commercial photo archiving systems such as Picassa or Flickr, the notion of indexing photographs on a timeline and reviewing them at different speeds in a photo slideshow is now

commonplace. The slideshow is also the default mode of presentation on dedicated photo display devices or screen-saver software. Things get more complicated in systems like Facebook which relate photographs and other media for multiple users (friends) over time. The traditional visualisation for this is a transcript or log of interaction, although individual video summaries and have recently been introduced by Facebook to envision selective photo narratives. The manual construction of multimedia narratives in 'digital stories' constitutes another approach to visualising the timecourse of events, and is represented by a new crop of apps such as Adobe Voice, Storehouse and Com-Phone. However, most of these existing approaches are individualistic, in the sense of visualising the photo stream of individual users, rather than relating them to each other. In the work presented here, we seek to address a demand for supporting better photo sharing and visualisation within small groups (Ojala and Malinen (2012)). The work is based on a new paradigm for multi-photo display on a dedicated device or service, reported at the HCI 2012 conference. In particular, we take up a recommendation of that paper to refine the way multiple synchronised photo streams are dynamically displayed.

Our original system supported the presentation of multiple photo streams (MPS) on a photo display (Zargham et al. (2012)). Its motivation was to allow multiple users to keep in touch through a kind of visual twitter feed of concurrent photographs from the mobile phones of members of a small group of four friends. These were displayed in a dynamic collage in the four quadrants of a dedicated ambient display, with photographs to each quadrant arriving in real time as photographs are taken (and uploaded to the system). The same display and system could be used to review a historical collection of photographs from each member of the group, in lock-step. A slider allowed users to rewind the display to a particular point on the timeline, and a replay button allowed them to animate the display again at different speeds. Users enjoyed seeing their photographs come in alongside those of their friends, particularly when they had attended shared events at which they had all taken photos. However, they had some problems controlling the rewinding and re-display of synchronised images at different speeds. Image streams are typically 'bursty' and non-linear (Gargi (2003)) and users found it hard to find a speed setting which was not too boring (with nothing happening in certain quadrants) or too fast (with images speeding past too quickly to see them). Hence we recommended the development of algorithms to warp the time of image display in different ways, to optimise the user experience of image review.

In the rest of this paper we report two lab-based studies exploring seven synchronised photo display algorithms. In the first study we test three ways of varying the speed of presentation and its effect on image perception, comprehension and memory. In the second study we measure the accuracy of four algorithms for identifying the continuity of images from a similar event, and user's initial reaction to the best of these when used to create a burst of similar images. The work is presented as a technical exploration of options for the synchronised photo stream paradigm, resulting in recommendations for this new field.

2. RELATED WORK

There are two related research themes addressing the core aspects of multiple photo stream visualisation presented in this paper: i) spatial arrangement of multiple streams, and ii) time-synchronous delivery of photos. There have been many efforts put in spatial arrangement of visual content, either from photographic collections or video repositories. Some of them maintained entire content from images, such as (Ren and Calic (2009)) who optimised usability, available display space and resulting cognitive load

while maintaining the underlying structure of the represented content; while others have discarded less salient image regions to condense the representation of large-scale photo collections in a visually appealing collage or digital tapestry (Rother et al. (2006)). In their seminal work (Frohlich et al. (2002)) express that an important aspect of photo sharing are the clues of the timeframe and other contextual information about the photos, as well as the symmetrical contributions from users involved in photo sharing. The time-synchronous nature of photo sharing, especially in the mobile scenario, has been addressed in (Clawson et al. (2008)) who developed an application that enables capture and simultaneous photo sharing in real-time within a small group of collocated users. Both research strands have analysed spatial and temporal aspects of sharing photo streams in isolation. In this paper we exploit the potential of a simple spatial layout and photo stream's temporal structure to deliver an intuitive presentation of synchronised photo streams.

3. SYNCHRONOUS VISUALISATION

As initially presented in the work on requirement analysis for visualisation of multiple photo streams (Zargham et al. (2012)), due to emergence of passive photo capture and life-logging practices (Sellen and Whittaker (2010)) there is a clear demand for visualisation of photo streams from different sources. Based on these requirements, a novel interface is proposed that lets users observe their photos chronologically and concurrently in a grid of four adjacent windows. This design enables visualisation of concurrent events and experiences within a small group of users, whether they were collocated or not. The concurrency of presented photo streams is achieved by transforming intervals between capture time stamps of two consecutive photos from the presented streams into intervals between appearances of the respective photos in the interface. In order to optimise the perceived speed of transitions between the photos, three types of transformations or scalings applied to calculate the transition intervals, i.e. *transition modes*, were studied: fixed, proportional and logarithmic.

The *fixed* transition mode provides an experience similar to typical slideshow by assigning fixed intervals between transitions of two photos in the interface. Users can set the fixed interval by using the vertical slider, and the range used in the experiment was set to 0.1 to 10 sec. This mode was the baseline case in our study, since users can easily relate to it as a typical photo slideshow.

In the *proportional* transition mode, transition interval (t_i) is calculated by dividing the difference of the

capture time stamps (Δ_t) with a scaling coefficient (K): $t_i = \Delta_t/K$. The scaling coefficient K can be set by the vertical slider from 1 to $5 * 10^6$. By setting it to 1, interface presents the photos in real time, i.e. by the pace at which they were taken. The proportional mode conveys the notion of time between the captured events, but could result in extremely fast or slow transitions.

In the *logarithmic* transition mode, the presented transition interval (t_i) is calculated as logarithm of the time difference between the capture time stamps $t_i = \log_B(\Delta_t)$. Users can use the vertical slider to set the logarithm base B , thus speeding up or down the playback. This still conveys the feeling of time, yet suppressing the extreme values of transitions.

Before the playback, user selects the photo streams from a drop down menu, and one of the three transition types. A horizontal slider at the bottom part of the interface facilitates temporal browsing through timeline of the photo streams and gives indication of the current position of the photos in the overall dataset. The vertical slider has been placed on the right side of the interface to control the speed of playback.

4. USER STUDY OF TEMPORAL PRESENTATION

In this paper, a twin-stream version of the photo-sharing interface was implemented, as depicted in Figure 1. In order to solely focus on the temporal aspects of this interface, the implementation comprised only two side-by-side windows. The study has focused on three temporal aspects of synchronous visualisation: transition modes, playback speed adjustment and user's comprehension of the visualised dataset.

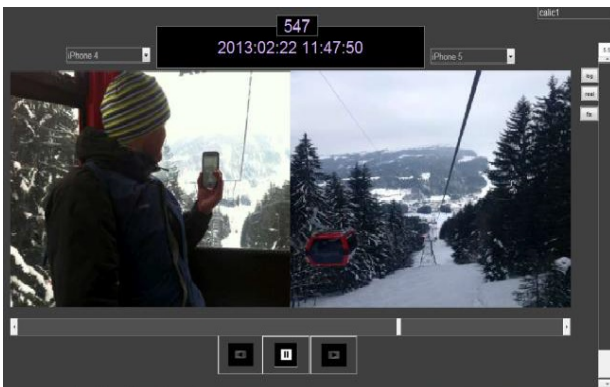


Figure 1: Twin photo stream interface used in the study.

4.1. Experimental Design

To investigate the effect of the the first three transition modes and the proposed interface on

user's photo viewing behaviour, a study of user's perception, comprehension and memory when using this visualisation interface has been designed and conducted. The study recruited 20 participants and instructed them to provide their photo collections comprising at least 500 photos. The photos provided by the users have all been taken within last 12 months. Participants were friends, family or colleagues of the researchers who conducted this study. Each participant's photo stream was paired with that of the related researcher. Before pairing participants, the photos have been checked for the consistency throughout the entire period. All participants were living in the UK at the time the study took place. The participants consisted of 12 females and 8 males, aged 20 to 50.

This study investigated three conditions, i.e. fixed, proportional and logarithmic transition modes, and each condition had two phases. For each condition, the phase 1 comprised participants observing the first 300 photos using relevant transition mode. During observations participants were adjusting the speed of the transitions using the vertical slider for comfort. In the phase 2, a random photo was picked from participants' photo stream and asked: i) if they can remember this photo from the slideshow, ii) do they know what happened next in their stream and iii) what happened next in the other person's stream. If requested by participants, a visual clue of four images was presented to them and they were asked which one of those photos happened next. At the end, the participants have been told to verify their answer by searching for the photo that was shown to them. The same process was repeated by picking a random photo from the other photo stream. At the end of the experiment, participants were asked general open-ended questions for later qualitative analysis.

4.2. Results

4.2.1. Total Slideshow Time

The average slideshow duration for viewing 300 photos was: 181.19 sec for the proportional transition mode, 399.21 sec for the fixed transition mode and 262.23 sec for the logarithmic transition mode. One-way ANOVA statistical test was conducted to assess if the use of logarithmic, proportional or fixed transitions affects the slideshow time duration. The ANOVA result was $F(2,57)=9.5$, $p=0.0002$, which indicates the results were statistically different. Furthermore, a t-test was performed between each set of results. It showed that there is a significant difference between each pair of transition types (all p-values were less than 0.001). Therefore, it can be concluded that each of the transition types affected the length of the total slideshow time. From qualitative analysis, it can be observed that the

use of logarithmic transition mode was preferred for watching multiple photo streams, since this mode was faster yet provided better comprehension. The fixed transition mode has been preferred by participants aiming to watch the photos in more detail and for longer.

4.2.2. Preference of Transition Modes

During the interview we asked participants to rank their favourite transition modes (highest rank scores 3, second 2 and the last 1 point). The highest ranked overall was the logarithmic transition mode with 51 points, followed by the fixed transition mode with 39 points and the proportional mode with 31 points. Most of the participants believed that logarithmic mode conveys best the notion of time and although the speed of slideshow can be very fast, they easily follow the story. In addition, in the logarithmic mode events can be differentiated despite the speed of slideshow. The participants reported that photos taken in bursts were visualised in the time-lapse fashion and that resulted in visually very appealing effect. On the other hand, participants did not prefer the proportional transition mode as much as the other two, because the transitions were often too fast or too slow.

In order to compare the proposed visualisation to the current practices, the participants were asked if they normally used slideshows to watch their photos and most of the answers were negative. The majority of participants (80%) found the slideshows “boring”. After observing the proposed interface, they all enjoyed watching the photo streams in a slideshow-like format but with different transition intervals. This reaction is a result of the contextualisation of personal photos and comparison of concurrent events from different streams. Secondly, factors such as the conveyed notion of time, much shorter observation time in logarithmic mode and the freedom of selecting the transition intervals in fixed transition mode, all improved the experience of viewing the photo streams.

4.2.3. Alternation between Windows

In the proposed interface, after each transition interval, one of the two presented photos would change. In case the transition happens in a window that did not change in the last transition, we have an alternation. The number of alternations between slideshow windows has been calculated for each data set of 300 photos. We found that on average, there were 14.86 alternations in each data set. Out of the 20 participants, 12 used a camera phone and 8 used a normal point-and-shoot camera. In our previous study (Zargham et al. (2012)) it has been found that more alternations between photo streams brings better experience in the visualisation of multiple photo streams. The study

in this paper showed that the average number of alternations in photo streams that have generated by a camera phone (18.22) was more than that of a normal camera (9.91). One-way ANOVA statistical test resulted in a $F(1,58)=19.38$ and $p < 0.0005$, proved that there is a significant difference between the means of the number of alternations. Hence, it can be concluded that the current culture of photography with a camera phone in comparison to a normal camera provides more evenly distributed photo streams and consequently, more alternations between photo streams.

4.2.4. Speed Control

Throughout the experiment, participants were able to change the speed of the transitions using the vertical slider. They selected a wide range of values for the base, coefficient and fixed intervals. The mode of each participant's selected values was calculated. The minimum value of the logarithmic base was 7, while the maximum was 475, with the average of 203 and the standard deviation of 191. In proportional mode, the minimal selected scaling coefficient value was 2282, the maximal was 111,660, while the average coefficient was 52868 with the standard deviation of 39065. In the fixed transition mode, the minimum transition time was 0.4 sec, while the maximum was 3 sec, with average of 1.2 seconds, and standard deviation of 0.84. These diverse values show that irrespectively of the transition mode, speed of visualisation depends on many factors: user's personality and mood, type of content or display device. Therefore no generic values can be defined for the transition speed in any of the three modes.

4.2.5. Comprehension

There were three photos that were chosen randomly from the participant's photo stream and another three were selected from the researcher's photo stream. We asked participants if they remembered those photos from the slideshow. The mean of remembering their photos was 3 out of 3 while the mean of remembering the researcher's photo was 2.7 out of 3. After applying the One-way ANOVA test we got values of $F(1,38) = 8.14$ and $p = 0.006$. From this result it can be concluded that there is a significant difference between remembering user's own images and the researcher's images, after watching the two photo streams using this application. Participants on average remembered 5 out of 6 photos of what happened next to them and the researcher when a photo from their stream was shown. On the other hand, they remembered 4.5 out of 6 photos of what happened next to them and the researcher when they saw a photo from the researcher's stream. One-way ANOVA test resulted in $F(1,38) = 1.7$ and $p = 0.18$, which shows that there is no significant difference between the means. The results show that although

participants sometimes might forget what happened next in different conditions mostly in the researcher's stream, they have a very good recollection of the narrative in the multiple photo streams. This demonstrates that the side-by-side photo sharing application is an effective way for remembering the users' and their friends' photos and events. This application can be utilised as a storytelling tool, enabling relations between different events that have occurred in the life of two friends through their shared photos. The analysis of the influence of transition modes to memory and comprehension showed that there is no significant difference between the three transition modes in recalling events in user's own or friend's photo streams.

4.2.6. Search

After watching the photo streams, one of the tasks that each participant was given was to find a random photo from their own and their friend's photo stream. On average, it took 49 sec to find a photo from the user's own stream and 44 sec to find a photo from their friend's stream. Participants used mainly two techniques to find a photo. They were using the time-line to find the event represented in the photo. Afterwards they located the photo by advancing through that event's photos one-by-one. The other search strategy emerged when participants were unsure which event the photo belonged to, In that situation they used a combination of time-line scanning and playing the photo streams in fast mode, until the required event was found.

5. CONTINUITY IN PHOTO STREAMS

An important discovery emerged from observations of the fast proportional transition mode visualising events whose speed of content change was slower than the rate of capture. In those circumstances the produced photo stream appeared continuous. Triggered by this finding, a notion of "continuity" has been introduced, denoting a finite incremental change between two photos in a photo stream that produces an effect of event's continuity if presented at a fast visualisation rate, i.e. in a time-lapse video fashion. Therefore, a second study was conducted focusing on this aspect of photo stream visualisation.

In order to automatically detect which pairs of photos could be classified as "continuous", three computer vision algorithms were implemented and evaluated. The first algorithm is based on the dense optical-flow estimation method (Lucas et al. (1981)), which tries to calculate the motion between two image frames which are taken at times t and $t + \Delta_t$ at every pixel position. The overall measure of continuity is thus inversely proportionate to the sum of calculated motion vector intensities between two

photos. The second algorithm is based on SIFT feature matching method (Lowe (1999)) where in case of a significant number of matched features between the two photos, with feature displacement within predetermined limits, "continuity" likelihood is very high. Finally, a recent dense correspondence estimator SIFT Flow (Liu et al. (2011)) is used to derive a dense flow using invariant features, while the continuity measure is derived from its energy optimisation function.

5.1. Evaluation of Continuity Detection

In order to evaluate effectiveness of the continuity detection methods, a simple interface was designed to let the user decide if there is a continuity between two adjacent photos from a photo stream. Two consecutive photos were displayed forwards and backwards in time with the delay of 0.4 seconds in the fixed transition mode until the user decides if these two photos were continuous or not. In this experiment we presented 779 photos of a ski trip to participants. The dataset comprised photos that have been often taken in burst, resulting in a good proportion of potentially continuous photo pairs. The manual decisions were logged and used to determine optimal threshold of continuity for each of the proposed algorithms. Using the optimal thresholds, a comparison is conducted to determine which algorithm was the best for detecting continuity in photo streams.

5.1.1. Accuracy of Algorithms

To find the accuracy of continuity detection for each algorithm, the users were labelling continuous photo pairs using three different resolutions: small (60*40), medium (400*300) and large (640*480). The accuracy was calculated using Equation 1.

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (1)$$

Here, the terms true positives (T_p), true negatives (T_n), false positives (F_p), and false negatives (F_n) compare the results of the detector with the manually labelled ground truth. The accuracy results for each of the three algorithms are illustrated in Table 1. The results showed that proposed SIFT had the closest performance to the manually labelled ground truth in determining the continuity of photo streams.

After the selecting SIFT based continuity detector as the closest algorithm to human labelling the number of continues photos in the dataset of the first experiment was calculated. There were three sets of 300 photos for each participant and in total there were 20 participants. The result showed that on average 85 out of 300 photos in each dataset

Table 1: Accuracy of proposed algorithms for different images sizes compared to the ground truth.

Algorithm/Size	Small	Medium	Large
SIFT Flow	81%	79%	79%
SIFT	91%	91%	91%
Optical Flow	69%	71%	75%

were detected to be “continuous”. This shows there is a significant amount of continuous photos in user’s personal photo collections and thus a lot of potential to exploit continuity in visualising photo streams.

6. CONCLUSION

In this study we examined the user behaviour of sharing personal photo streams using a novel photo-sharing interface with four different transitions modes. The study found that the logarithmic transition mode was suitable for watching multiple photo streams quickly without losing notion of time. On the other hand, fixed transition mode was preferred by users watching photo streams in more detail for longer. After observing the two streams, participants could easily recall the photos from both stream and they had an idea about the chronology of events. The study also showed that using time-line and fast play are effective tools for searching through photo streams. At the end, it has been shown that continuity transition mode with proposed detection algorithm based on the SIFT feature matching offers a lot of potential in presenting photo streams. The future work will be focusing on summarisation challenges of multiple photo streams and trials of the proposed interface in real-world environments.

REFERENCES

- Clawson, J., Volda, A., Patel, N., and Lyons, K. (2008). Mobiphos: A collocated-synchronous mobile photo sharing application. In *Proc. of the 10th Inter. Conf. on HCI with Mobile Devices and Services*, MobileHCI '08, pages 187–195, New York, NY, USA. ACM.
- Davis, M., King, S., Good, N., and Sarvas, R. (2004). From context to content: Leveraging context to infer media metadata. In *Proc. of the 12th Annual ACM International Conference on Multimedia*, pages 188–195, New York, NY, USA. ACM.
- Frohlich, D., Kuchinsky, A., Pering, C., Don, A., and Ariss, S. (2002). Requirements for photoware. In *Proc. of the 2002 ACM Conference on CSCW*, pages 166–175, New York, NY, USA. ACM.
- Gargi, U. (2003). Modeling and clustering of photo capture streams. In *Proc. of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, MIR '03, pages 47–54, New York, NY, USA. ACM.
- Graham, A., Garcia-Molina, H., Paepcke, A., and Winograd, T. (2002). Time as essence for photo browsing through personal digital libraries. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 326–335. ACM.
- Liu, C., Yuen, J., and Torralba, A. (2011). Sift flow: Dense correspondence across scenes and its applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):978–994.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. IEEE.
- Lucas, B. D., Kanade, T., et al. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679.
- Marcus, A. (2004). It’s about time. *interactions*, 11(6):16–21.
- Miller, A. D. and Edwards, W. K. (2007). Give and take: a study of consumer photo-sharing culture and practice. In *Proc. of the SIGCHI conference on Human factors in computing systems*, pages 347–356. ACM.
- Ojala, J. and Malinen, S. (2012). Photo sharing in small groups: Identifying design drivers for desired user experiences. In *Proceeding of the 16th International Academic MindTrek Conference*, pages 69–76. ACM.
- Ren, K. and Calic, J. (2009). Freeeye: Interactive intuitive interface for large-scale image browsing. In *Proc. of the 17th ACM International Conference on Multimedia*, pages 757–760, New York, NY, USA. ACM.
- Rother, C., Bordeaux, L., Hamadi, Y., and Blake, A. (2006). Autocollage. *ACM Trans. Graph.*, 25(3):847–852.
- Sellen, A. J. and Whittaker, S. (2010). Beyond total capture: a constructive critique of lifelogging. *Commun. ACM*, 53(5):70–77.
- Van Dijck, J. (2008). Digital photography: communication, identity, memory. *Visual Communication*, 7(1):57–76.
- Van House, N. A. (2007). Flickr and public image-sharing: distant closeness and photo exhibition. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 2717–2722. ACM.
- Zargham, S., Calic, J., and Frohlich, D. (2012). User experience study of multiple photo streams visualization. In *Proc. of the BCS-HCI '12*, pages 416–421, Swinton, UK, UK. British Computer Society.