Multimedia Tools and Applications manuscript No.
(will be inserted by the editor)

Time and space for segmenting personal photo sets

Nuno Datia · João Moura Pires · Nuno Correia

Received: date / Accepted: date

Abstract A personal collection of photos shows large variability in the depicted items, making difficult a fully automated solution to cope with sensory and semantic gaps. Emotions and non-visual contextual information can be very important to address those problems. Manual annotations are key, but their time-consuming nature alienate users from doing them. One solution is to lower the annotation effort, building solutions on top of algorithms that prepare a context separation, making possible the reuse of annotations. In this paper we present a segmentation algorithm that uses spatio-temporal information to segment personal photo collections. The algorithm is assessed in a user study, using the participants own photos. The results show users make none or few changes to the proposed segmentations, indicating an acceptance of the algorithm outcome.

Keywords Empirical User Study · Segmentation Algorithm · Formalisation · Personal Photo Collections

1 Introduction

A picture is worth a thousand words, some may say. In fact, humans can see beyond what is depicted, using their knowledge and skills to fulfil gaps, and to hypothesise what a photo represents. For personal use, photos serve the purpose of holding memories for personal or collective purposes (Gye 2007; House 2009; Kirk and Sellen 2010). According to Lux et al (2010), millions of new photos are taken everyday, some of which are uploaded to online sites, like Flickr[1]. How to properly manage a personal collection of photos, without effort, is a problem that thousand of users often face, and has been addressed by (Sun et al 2002; Latif et al 2006;

N. Datia
Instituto Politécnico de Lisboa, ISEL, Rua Conselheiro Emídio Navarro, 1 1959–007 Lisboa
E-mail: datia@isel.ipl.pt

J. Moura-Pires · N. Correia
Universidade Nova de Lisboa, FCT, 2829–516 Caparica, Portugal
E-mail: {jmp,nmc}@fct.unl.pt

[1] `http://www.flickr.com/`

Zhao et al 2006; Kang et al 2007; Viana et al 2008; Sun et al 2013). The solutions need to include user-driven methods (Datta et al 2008). Recent studies (House (2009); Whittaker et al (2010)) show that locating specific photos in a personal collection becomes harder, as times passes by. The size of collections also makes it difficult (Whittaker et al 2010). People do not spend much time maintaining their collections, nor annotating them, although the personal information is important for long term retrieval (Whittaker et al 2010). The key is on reducing the manual labour for structuring and for annotating the collections, raising its benefits.

Time and space are the most basic dimensions in human life, being the major influence on peoples actions (Kellerman 1989). Thus, any context in a personal photo collection should involve them. A major problem separating photos with similar context is the trade-off between annotation effort and the relevance of information gathered from, and presented to, the users. This means that, on the one hand, a collection with fewer divisions is easier to annotate but can lead to generic annotations, with less informative value. On the other hand, an over-divided collection presents an additional effort of annotation that can stop users from annotating. The aim of this research is to demonstrate that is possible to present the users a segmented collection of photos, that is coherent chronologically and it is accepted by the users as a proper structure of their memories. The segmentation requires no human intervention, as it uses only information gathered from the photos' metadata. On top of it, users can insert annotations in batch to the segments, lowering the manual annotating labour. Thereby, our contributions are:

1. A segmentation algorithm that uses spatio-temporal information presented in each photo to autonomously separate a personal photo collection into segments of similar context;
2. The demonstration that temporal cycles are needed to properly separate the context;
3. The proposal of 4 binary relations between segmentations, addressing a lack in the literature for a theoretical support for qualitative comparisons.

This paper is organised in the following way. Section 2 presents related work on photo segmentation. In Section 3, the problem is formalised and the segmentation algorithm presented. The experimental set-up is described in Section 4, followed by the results in Section 5. Finally, we draw conclusions and point out directions for future work.

2 Related Work

In Graham et al (2002); Gargi (2003), the authors suggested that users take photos in bursts. Such behaviour can be used to separate different groups of photos. Many psychological studies, for example Janssen et al (2006), shows the importance of time in our process of recalling past events. Besides, people describe and understood time using a set of cycles that govern their lives, which includes the natural cycle of days (Zerubavel 1985). Considering such facts, it is not odd to see the temporal information is the most used in the algorithms whose objective is to divide a collection of photos. Besides time, the content of photos (Platt et al

2003; Loui and Savakis 2003; Cooper et al 2005; Gozali et al 2012) and the spatial location information (Naaman et al 2004; Cao et al 2008; Bruneau et al 2010; Cooper 2011) are the most used features to divide photo collections. Before we go into details, it is worth clarifying the meaning of two terms that will be used frequently in this section: segmentation and clustering. A segmentation is a division of a temporal totally ordered set into smaller parts, the segments, keeping the temporal order. We found examples of segmentations in Platt et al (2003); Loui and Savakis (2003); Naaman et al (2004); Cooper et al (2005); Gozali et al (2012). Clustering is a process that divides a set into clusters of similar objects, not necessarily ordered. The work of (Graham et al 2002; Platt et al 2003; Loui and Savakis 2003; Naaman et al 2004; Cao et al 2008; Bruneau et al 2010; Cooper 2011) use clustering techniques applied to photo collections. In the literature we can find three main reasons for segmenting/clustering photos:

- to support browsing over a large collection of photos (Graham et al 2002; Platt et al 2003; Naaman et al 2004; Cooper et al 2005; Bruneau et al 2010),
- to assist the creation of photo albums (Loui and Savakis 2003; Cooper 2011; Gozali et al 2012), and
- to support annotation (Cao et al 2008).

Despite the purpose and techniques used, there are two common approaches to settle the segments (or groups):

1. using simultaneously a set of features, or
2. using an N-step algorithm, where the solution is refined in each step.

The first approach is used by Gozali et al (2012). The authors mix temporal information, EXIF metadata, and visual information to settle a segmentation of a photo stream. They use an Hidden Markov Layer, whose parameters are learned from a set of unlabelled, unsegmented event photo streams and from the event photo stream they want to segment.

However, most of the works use the second approach. In Graham et al (2002), the authors use the time gap between photos to build an hierarchy of clusters, with an increasing level of detail. In each step, the clusters are divided, using a fitness function, that uses the quartiles of the gap distribution inside each cluster. The work of Platt et al (2003) uses time gaps between two consecutive photos, compare it to the neighbourhood average differences, and decide to start a new segment when such gap is greater than average. When the number of photos within a segment exceeds a maximum value, it is divided using a content-based clustering technique.
A similar approach is used by Loui and Savakis (2003), where the divisions are settled by clustering the time gaps between photos with K-Means ($K = 2$), choosing the cluster with higher values to perform the segmentation. The gaps are calculated from a temporal ordered set of photos. The segmentation is followed by a second level clustering, using a block-based color histogram correlation, that produces fine grain sub-groups for each temporal segment. Naaman et al (2004) do a first segmentation, initiating a new segment whenever the gap between consecutive photo exceeds a predefined value in time and space. The initial segmentation is used to build clusters, grouping photos with similar location. Cooper et al (2005), use a multi-scale approach to determine the temporal structure in a photo collection.

They take into account the time gap between photos, to calculate a photo-indexed novelty score. The score uses a parameter K, enabling different time resolutions. The boundaries of clusters are identified by analysing each novelty scores first difference. Higher differences means higher novelty scores. The novelty score, following Foote (2000), is calculated from a similarity matrix by photo (in time order). The principal diagonal contains the intra-cluster similarity. The authors also build a content-based matrix using low-frequency Discrete Cosine Transform, to detect similarities between photos. The temporal and content-based matrices are combined to settle the final clustering solution.
Other example is Cao et al (2008). The first clusters from temporal information are fixed by the Mean-Shift algorithm (Comaniciu and Meer 2002). There is an additional step, where the initial clusters are refined using information about the location. Cooper (2011) also starts to settle hierarchical temporal clusters, maintaining high pairwise similarity between photos. Simultaneously, another hierarchy of clusters is built using the distance between photos. The final photo clustering is selected from the two hierarchies of initial clusters. They use dynamic programming to minimize a fitness function considering the temporal range and the distances between clusters.

Bruneau et al (2010) used both approaches. They used Gaussian Mixture models to identify clusters from photos in a 3D space taking $(time, latitude, longitude)$. They start with an arbitrary high K value and, using Bayesian estimation of model parameters by means of a variational approximation, an effective K is obtained. The clusters are then refined using agglomerative clustering, considering only the spatial location.

It seems that researchers agree the temporal information plays an important role towards event detection from collections of photos. However, their use of time do not take into account its peculiarities and how it influences our daily life. Namely, it is known the temporal cycles affects our social life (Zerubavel 1985; Zuzanek and Smale 1993) and thus, time constrains the generation of photos, whose purpose is holding memories for personal or collective purposes (Zerubavel 1996; Gye 2007; House 2009; Kirk and Sellen 2010). In this paper, we consider the daily cycle as a natural separator of events. However, we model the day from a social perception, that can be different from the established range of a day.

From the work described above, it is possible to see an increase usage of geographic location of the photos for grouping purposes. The technological evolution, specially the adoption of smartphones worldwide (Gye 2007; Do et al 2011), escalate the availability of this type of information. Nevertheless, the spatial location in a photo collection can be sparse, with variable precision, depending on the source of information (von Watzdorf and Michahelles 2010). Researchers take a binary approach when considering location information: it is either available (and thus correct), or absent. In this work we consider that location information can be available, but it may be wrong. We use an outlier detection method to treat such cases as missing information.
Most of the segmentation algorithms in literature are used for presentation purposes. Their design aims at presenting collections in a compact way, with a strong focus on the content. However, those algorithms are not sufficient to keep balanced important features in a solution that encourages manual annotations. Namely,

- segments should follow a *natural* temporal organisation of the activities. By natural, we mean in accordance with predefined cultural divisions of time, that drive people to do their activities;
- the number of segments for each day should be small, but large enough to allow the introduction of non trivial annotations.

Those features provide the reduction of the manual labour. On top of it, annotations can be suggested in behalf of the user, using an algorithm similar to the one proposed by Datia et al (2014).
The new segmentation algorithm, described in the next section, includes this balance by design and addresses the sparse and wrong location information.

3 Segmentation algorithm

In the support medium of a personal camera, the stored photo set spans through different time periods, depicting personal and social events, and also photos related to the ordinary day life. The separation of the photos, according those different contexts are done using a segmentation algorithm, described in this section. Before we give details about how it works, we formalise the notions of segments, segmentations, and a set of relations between segmentations. Those formal definitions will support the comparison of different segmentations for the same set of photos.

3.1 Formalisation of the problem

Let P be a sequence of N photos, ordered non-descendingly by their creation timestamps. P is represented by the pairs (t_n, g_n), where t_n is the timestamp for photo n, and g_n is the spatial location for photo n, or, if the location is unavailable, $g_n = null$. Without loss of generality, for segmentation purposes, we represent all photos taken at the same second (the time grain available in EXIF) as a single instant, thus

$$\forall n \in [1..N-1] : t_n < t_{n+1} \tag{1}$$

Let $T = [t_1, \ldots, t_n, \ldots, t_N]$ be the sequence of t_n in P, satisfying (1).

Definition 1 (successor in T) An element $t_j \in T$ is the successor of $t_i \in T$, denoted by $(t_i)^{\succ} = t_j$, when there is no element in T between t_i and t_j, and thus $j = i + 1$.

The elements in T can be arranged to form non-empty, temporal contiguous subsequences of T, leading to the notion of segment.

Definition 2 (a segment) A segment, denoted by $s = [\, t^-, t^+]$, is a non-empty, sub-sequence of T. t^- is the lower limit of the segment and t^+ is the upper limit, where $t^- \leq t^+$, holding

$$\forall t_n \in T\ ,\ t_n \in s \Leftrightarrow t^- \leq t_n \leq t^+$$

Notice that a segment can be singular, when $t^- = t^+$, and can be equal to T, when $t^- = t_1$ and $t^+ = t_N$. A segmentation is a set of segments that covers every $t_i \in T$, where each t_i belongs to exactly one s.

Definition 3 (a segmentation) Given $T = [t_1, \ldots, t_n, \ldots, t_N]$, a segmentation S, represented by $S = [s_1, \ldots, s_k, \ldots, s_K]$, is a non-empty ordered set of segments from T, where:

i) $\forall k \in [1..K-1], (t_k^+)^{\succ} = t_{k+1}^-$
ii) The lower limit of s_1 is t_1
iii) The upper limit of s_K is t_N

The cardinality of a segmentation S ranges from 1 to N. In the first case, $t_1^- = t_1$ and $t_1^+ = t_N$. In the later, S contains N singular segments, where $t_n^- = t_n^+$.
For sake of simplicity and comprehension, for this point forward we will denote the limits of a segment s, t^- and t^+, by s^- and s^+ respectively.

3.1.1 Relations between segments

The relations between segments follows Allen's interval algebra (Allen 1983). However, for sake of completion and coherence in the formalism, we describe them here with an adjusted lexicon to this problem domain.

From Definition 2, it is easy to see that two segments of T, s_1 and s_2, are equal when both of their limits are equal, $s_1^- = s_2^-$ and $s_1^+ = s_2^+$. We will denote the `equal` relation by the symbol $=$. For simplicity, to represent $\neg(s_1 = s_2)$, we will use the symbol $\neq$. To ease the understanding of the upcoming definitions, we will use three segments of T.

Definition 4 (precede relation) Given s_1 and s_2, s_1 `precede` s_2, denoted as $s_1 \prec s_2$, iif $s_1^+ < s_2^-$

Following Definition 4, it is possible to define a more strict precedence order, the `contiguous precedence`.

Definition 5 (precede contiguous relation) Let $\prec_c$ represent that s_1 `precede contiguous` s_2, where

$$s_1 \prec_c s_2 \Leftrightarrow (s_1^+)^{\succ} = s_2^-$$

Definition 6 (contained relation) A segment s_2 is `contained` in s_1, denoted by $\sqsubset$, when all elements of s_2 are also elements of s_1,

$$s_1 \neq s_2 \wedge \forall t \in s_2 \Rightarrow t \in s_1$$

For simplicity, to represent $\neg(s_1 \sqsubset s_2)$, we will use the symbol $\not\sqsubset$. Two segments overlapped, if they have some common elements and they are not equal, neither one contains the other.

Definition 7 (overlap relation) Two segments `overlap` each other, denoted as $s_1 \; O \, s_2$ iff

$$\begin{aligned} \exists t \in T \; : \; (t \in s_1 \; \wedge \; \; t \in s_2) \wedge \\ s_1 \neq s_2 \wedge \\ s_1 \not\sqsubset s_2 \; \wedge s_2 \not\sqsubset s_1 \end{aligned}$$

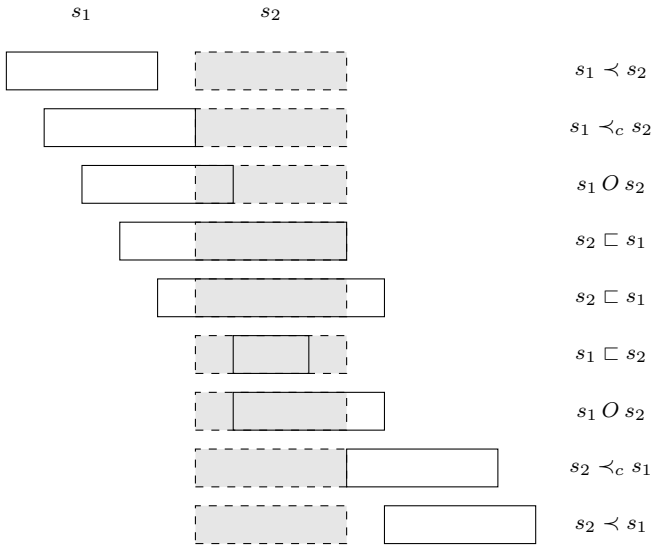


Fig. 1: Illustration of the relations between segments.

	Symbol	Symmetric	Asymmetric	Reflexive	Irreflexive	Transitive
equal	$=$	•		•		•
precede	$\prec$		•		•	•
precede contiguous	$\prec_c$		•		•	
contained	$\sqsubset$		•		•	•
overlap	O	•				

Table 1: Summary of the binary relations between segments.

Figure 1 illustrates the relations between segments presenting their symbols and properties. The rectangles represent the timestamps range of a segment. The size of the rectangles are not representative of any feature or property. They are used only to illustrate the relations between segments. Table 1 summarises the relations between segments, presenting their symbols and properties.

3.1.2 Relations between segmentations

We will present some binary relations supporting the comparison between two segmentations. The only requisite for this comparison is that they are defined over the same photo set. Since there are many ways to segment a set of personal photos, the binary relations allow us to understand how segmentations differ. Since segmenting a set of personal photos is an important step towards annotation, for instance knowing that two segmentations present different grains of segmentations (although compatible), reveals the annotations of one segmentation are more detailed than the annotations of the other.
There are four scenarios of comparison between segmentations: the *equality*, and other three, divided in *compatible* (Figures 2 and 3) and *incompatible* (Figure 4).

The rectangles in the figures represent the time span of a segment.

Definition 8 (equal segmentations) Two segmentations S and S' are `equal`, denoted by $S \equiv S^{'}$, when all the segments at the same index, are equal. Thus

$$\forall s_k \in S, \forall s_l \in S' \ : \ k = l \ \Rightarrow s_k = s_l$$

The `equal` relation, defined between segmentations, is reflexive, symmetric and transitive. For simplicity, to represent $\neg(S \equiv S')$, we will use the symbol $\not\equiv$.

Figure 2 represents the case of two `compatible` segmentations. In such cases, there is no overlapping between segments of the two segmentations.

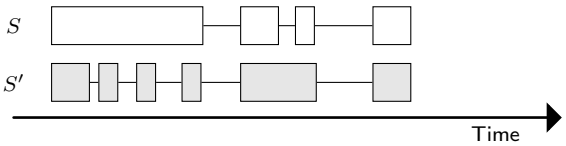


Fig. 2: Compatible segmentations.

Definition 9 (compatible segmentations) Let S and S' be two segmentations of T. S and S' are `compatible`, denoted by $\lesseqgtr$, when they are not equal and there is no overlapping between segments, such

$$\forall s_k \in S, \exists\, s'_k \in S' \ : \ s'_k = s_k \vee s_k \sqsubset s'_k \vee s'_k \sqsubset s_k.$$

The `compatible` relation is irreflexive and symmetric.

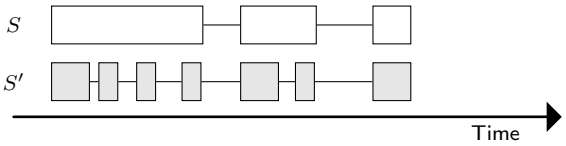


Fig. 3: S' is a refinement of segmentation S.

Figure 3 illustrates the `refinement` relation, that is a special case of compatibility between two segmentations, where the segments on the refined segmentation are equal or contained in the segments of the other segmentation.

Definition 10 (refinement relation) Let S and S' be two segmentations of T. S' is said to be a `refinement` of S, denoted by $S' \lhd S$, when $S \not\equiv S'$ and each segment of S' is equal or contained in one segment of S, holding

$$\forall s_k \in S', \exists^1 s_l \in S \ : \ s_k = s_l \vee s_k \sqsubset s_l$$

For simplicity and clarity, we will denote the $\neg(S' \lhd S)$ as $S' \not\lhd S$.

The relation shown in Figure 4, is the `incompatible` relation, denoted by $\|$.

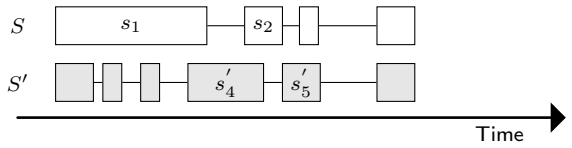


Fig. 4: Incompatible segmentations.

Definition 11 (incompatible segmentations) Let S and S' be two segmentations from the same T. They are `incompatible`, if some segments of S overlaps some segments of S'

$$\exists\, s_j \in S, \exists\, s_k \in S' \; : s_j \; O \, s_k$$

For the example shown in figure 4, there are many cases of of overlapping between segments: $s_1 \, O s'_4$, $s_2 \, O s'_4$, and $s_2 \, O s'_5$. The `incompatible` relation is irreflexive and symmetric.

The binary relations between segmentations are summarised in Table 2. It resumes their symbols and properties. The presentation order, from top to bottom, indicates their descending level of compatibility, where `equal` is more compatible than `incompatible`.

	Symbol	Symmetric	Asymmetric	Reflexive	Irreflexive	Transitive
`equal`	$\equiv$	•		•		•
`refinement`	$\lhd$		•		•	•
`compatible`	$\leqslant$	•			•	
`incompatible`	$\|$	•			•	

Table 2: Summary of the binary relations between segmentations.

Using the formalisation above, the problem is stated as follows. Given a set of photos P, represented by

$$[(t_1, g_1), \ldots, (t_i, g_i), \ldots, (t_n, g_n)]$$

we want to produce a segmentation S such that each segment maps to a distinct context from the user's perspective.

3.2 The LDES algorithm

In this work we call an event to a happening that spans through time and space, forming sequences of happenings that do not overlap, with a delimited context. The context separation can be done either by clustering or segmenting a set of photos. Since the temporal order of happenings is key, segmenting is a way to separate a set of photos. The segmentation algorithm is supported in assumptions:

(i) Timestamps are ubiquitous in todays digital photos;
(ii) Personal photo collections exhibit a bursty nature of the shots taken by the photographer (Graham et al 2002; Gargi 2003);
(iii) Personal and social activities are scheduled, performed, and dictated by temporal rhythms, among them the natural cycles of days and years (Zerubavel 1985);
(iv) Temporal order is important, as people tend to recall events using their order of occurrence (Friedman 2004; St. Jacques et al 2008; Kwok et al 2012);
(v) Personal memories are about happenings in time and space (Tulving 2002).

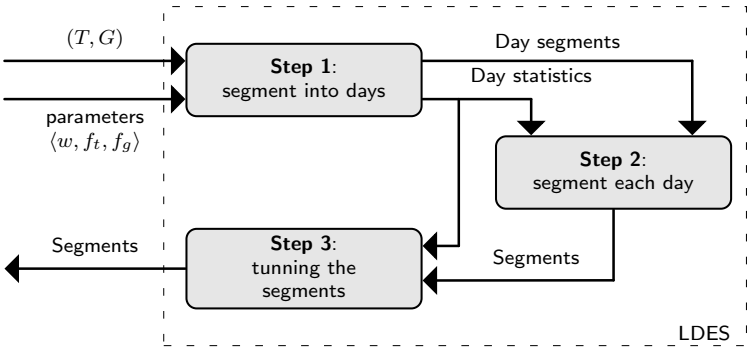


Fig. 5: Overview of the LDES algorithm.

The Logical Day Event Segmentation (LDES) algorithm uses the creation time of photos to fulfil a base segmentation. The spatial information is used to fine-tune the limits of each segment. We take advantage of the bursty nature of the shots made by the photographer to detect gaps in the collection, both in time and space. Figure 5 shows a diagram of the algorithm, representing its three steps — *(i)* Day finding; *(ii)* Event finding; and *(iii)* Event tunning . Parameters w, f_t and f_g are used in different phases of the algorithm and will be described next.

3.3 Step 1 — Day finding

The first step of the algorithm has two goals:

1. produce a segmentation S_{days} where each segment contains the photos for a single *logical day*;
2. produce a set of statistics, denoted by ξ, for each segment, that will drive the later steps.

A *logical day* is the calendar day assigned to all the photos taken in one day of activity. It gets its name from the fact that, for us humans, daily activities may span two calendar days. For example, if someone gets out of bed at 10 a.m. in the 1^{st} of May and goes to bed at 2 a.m. in the next day. For recalling purposes, the day that matters is the 1^{st} of May, because our notion of *day* only ends when we rest. Thus, 1^{st} of May is the logical day assigned to all photos taken in that period. Since the notion of *logical day* follows closely the daily cycle and the activities people do, the assignment changes on a daily basis, depending on the photos we have in the collection.

To settle a logical day, it is necessary to have a sequence of timestamps that spans from the last hours of one "standard" day into the early hours of the next one. If the timestamps fall within a window w, a parameter of the algorithm, they are considered to belong to the same logical day — the first day. If not, the logical and "standard" day are considered to be the same. The rational behind this window is that people need to rest a few hours, and this usually happens at night. Thus, after a "day" in people's activity documented by a set of photos, comes a larger gap. The assignment of one logical day ends when such gap is reached. LDES does not address exceptional cases, like New Year's eve; those cases should be treated at the application level, for example, setting w accordingly, for special days. If the gaps between photos are regular in two consecutive days, thus falling inside the window w, the logical day is the same as the standard day. We also assume the temporal information can be wrong, with errors reflecting a constant delta from the correct temporal reference. Since the gap between photos is unchanged by those errors, the LDES is not affected.

Besides the segmentation into logical days, this step produces the daily statistics ξ. These are important for the next steps of the algorithm, namely, to detect segments within the day. For each segment, representing a logical day, the statistics include:

1. the number of photos;
2. the maximum time gap,
3. the minimum time gap, and
4. the average time gap between two consecutive photos.

With such information, the algorithm can adjust the segmentation to the shot behaviour the photographer had each day.

3.4 Step 2 — Event Finding

The second phase of the algorithm takes the segmentation S_{days} and the statistics ξ to produce a segmentation S_{evt}, where S_{evt} refines S_{days}, i.e. $S_{evt} \lhd S_{days}$. This refinement divides each segment of S_{days} into fine grained segments that are close to events (e.g. to visiting a museum). Since each day has its own statistics, the algorithm adapts the cut points to each daily set of photos. The decision to create a new segment is based on a reference value, denoted by Δ_t. Whenever the time gap between two consecutive photos is bigger than Δ_t, a new segment is created.

Δ_t is calculated as

$$\Delta_t = f_t \times \Big[\text{average time gap } + (\text{maximum time gap} - \text{minimum time gap})\Big] \quad (2)$$

where $0.1 \leq f_t \leq 0.9$. Setting $f_t = 0.5$ is a recommended design value, that stands in the middle of the scale, providing a good separation of bursts in several scenarios. A lower value of f_t will produce more segments, and on the contrary, an higher value tends to join low density bursts, producing few segments. The rational behind the formula of Δ_t is the following. In the personal domain, each day is usually

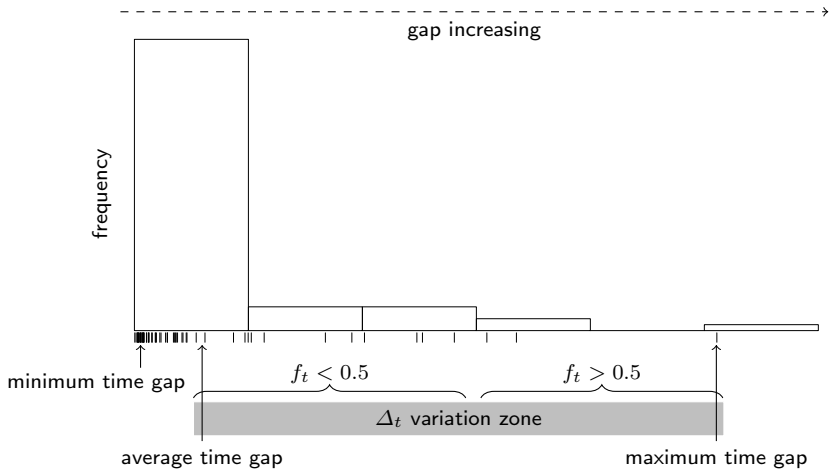


Fig. 6: Representation of a typical gap spread, in a personal photo collection.

dedicated to few specific events, that are worth being photographed. It seems natural that time gaps between consecutive photos change in different periods of the day, even for the same event. Fatigue, new subjects, and pauses, among others, influence the shooting behaviour. Although, typically, time gaps follow an exponential distribution, with low values more frequent than higher ones, there are different shot patterns that must be attained. Those are becoming more frequent as the usage of camera phones increase, and the photo taking habits change (Cobley and Haeffner 2009). We consider the following:

A. *Burst of photos* — two or more group of photos, with the inner group temporal gap lower than the outer group gap;
B. *Just two photos* — one or two spatio-temporal contexts;
C. *A single, uniformly separated group of photos* — the existence of just one spatio-temporal context;
D. *Low density (sparse) groups of photos* — Several spatio-temporal context, with few photos in each.

The maximum, the minimum and the average time gaps are used to handle these situations, contributing differently in each case, as encoded in (2). Shot pattern A is the most common. The typical gap distribution in this situation is

presented in Figure 6. There are many small gaps, representing the time separation within the bursts of photos, and few large gaps, the potential separations between events (Gargi 2003). In this scenario, the minimum time gap tends to zero, the average time gap tends to be small, and the maximum time gap tends to be much larger than the others, leading to $\Delta_t \approx f_t \times$ maximum time gap. Since maximum time gap $\gg$ average time gap, the most frequent situation is

$$\text{average time gap} < \Delta_t < \text{maximum time gap}.$$

For $f_t = 0.5$, the close dense bursts are separated from the ones that are far apart. If there are just two photos (shot pattern B), there is only one time gap and thus, the maximum, the minimum and the average time gap are all equal. In such cases, $\Delta_t = f_t \times$ time gap. Knowing that $0.1 \leq f_t \leq 0.9$, Δ_t is always lower than the gap between the two photos, separating them into different segments. If the gap is small, turning out they belong to the same event, the next step of the algorithm will use the spatial location of photos to correct such situation, joining the two photos.
In shot pattern C, the temporal information could be insufficient to identify the segments. Since there is a steady shooting pattern, all the time gaps are very similar. As such, the minimum and the maximum time gaps tend to cancel out in (2), making $\Delta_t \approx f_t \times$ average time gap. This means the $\Delta_t <$ average time gap, increasing the cardinality of the segmentations. The next step of the algorithm will use the spatial location to fine tune the segmentation, joining segments with similar spatial locations.
Shot pattern D happens when there are few photos documenting each event. In such cases, the gap between photos of the same event is larger but still smaller than the gap photos separating events. Assuming the minimum time gap does not tend to zero in such situations, the difference (maximum time gap$-$minimum time gap) tends to approximate the average time gap, leading to $\Delta_t \approx f_t \times 2 \times$average time gap. With $f_t \geq 0.5$, only gaps greater than average produce new segments.

Despite (2) uses different temporal statistics to adapt Δ_t to different shooting patterns, it is possible to generate under- or over-segmented collections. In such situations, the next step uses spatial information to guarantee spatio-temporal coherent segmentations.

3.5 Step 3 — Event tunning

The temporal information may not be sufficient to separate events. Since each photo n contains temporal and spatial information, represented by the pairs (t_n, g_n), and knowing that it takes time to move in space, there are two situations that need further analysis:

- there is a time gap between two consecutive segments, but their spatial location is similar;
- time gaps inside a segment exhibit a regularity, but the spatial locations indicate two or more places.

This last step takes care of these situations, analysing segmentation S_{evt} and producing the final segmentation, S_{final}. The first task is to validate the spatial coordinates. Although the location based services, available in many digital devices (e.g. smartphones), are becoming more accurate, such accuracy depends on the hardware characteristics and on environmental conditions. The location precision can vary from a few meters up to 3 km (von Watzdorf and Michahelles 2010). The spatial information is used to compute the distance of each g_n to the predecessor, g_{n-1}, denoted by $\delta_n = dist(g_n, g_{n-1})$. By definition, the distance for the first photo, δ_0, is 0, and so is the distance for all photos without spatial coordinates. Then, the outliers in the spatial data are detected. Three statistical methods were tried: one using Tukey's rule (McGill et al 1978), other the Local Outliers Factors (Breunig et al 2000), and another using a speed based solution. The development test shows the first is too sensitive, given the location's range of variation, and the second has an impact on the LDES performance, producing results similar to the latest. The proposed solution, speed based, employs a simple moving average over the spatial and temporal distances, considering a small window of size $2k+1$. The central point of the window is the photo whose coordinates are being assessed. Do note that, for the initial and last photos, the window is only $k+1$. This is because there is no previous or successive photos, respectively. The speed based method assumes that most of the photos in a personal photo collection are taken on foot. As for travelling long distances, it takes times. The average spatial gap between photos inside the $2k+1$ window is compared to a spatial reference value S_{ref},

$$S_{ref} < \frac{\delta_{n-k} + \ldots + \delta_n + \ldots + \delta_{n+k}}{2k+1} \tag{3}$$

and the average speed between photos inside the same window is compared to a temporal reference value V_{ref}

$$V_{ref} < 120 \times \frac{\left(\frac{\delta_{n-k}}{t_{n-k}-t_{n-k-1}} + \frac{\ldots}{\ldots} + \frac{\delta_n}{t_n-t_{n-1}} + \frac{\ldots}{\ldots} + \frac{\delta_{n+k}}{t_{n+k}-t_{n+k-1}}\right)}{(2k+1)} \tag{4}$$

The constant 120 is an approximation to the kilometres in one degree. A photo p_n is marked as an outlier when (3) and (4) are both `True`. The reference values were set as $S_{ref} = 5km$ and $V_{ref} = 100km/h$, making all photos taken at high speed and with large[2] gaps between them marked as outliers. In all the experiments we use $k = 1$. With this procedure at most 1% of spatial coordinates are marked as outliers for each collection used in the experiments. When the coordinates are missing, or are considered outliers, the photos are marked accordingly. This means the only information used from them is the temporal information.

After the outlier detection, the spatial information is used to fine-tune the limits of each segment, using the *split* operation in first place, and then using the *join* operation.

The *split* operation evaluates each segment s from S_{evt} to check if it can be refined, producing a set of segments denoted by $split_s = \{s_1, s_2, \ldots, s_n\}$, such $s^- = s_1^- \wedge s^+ = s_n^+$. If the distance between two consecutive photos is greater

[2] Considering a person on foot

than a reference value Δ_s, a new division is set. The reference value is calculated as

$$\Delta_s = (1.5 - f_g) \times \sigma \tag{5}$$

where σ is the standard deviation of δ_n for the valid photos[3] in s, and $0 \leq f_g \leq 1$. The refinement $split_s$ is kept if it lessens the distance's variance between photos inside each segment, producing more compact segments than the original one, verifying the strict inequality

$$\frac{\left(\sum_{s \in split_s} \sigma_s\right) \times \frac{1}{|split_s|}}{\sigma} < f_g \tag{6}$$

The numerator is the mean of the standard deviation of δ_n in each $s \in split_s$ and the denominator is the standard deviation of δ_n in the original s_k. The first spatial gap of each segment is left out of the calculus, as it represents a cutting point, thus it is not representative of the gaps inside a segment. Do notice that when $f_g = 0$, no split is done. On the other hand, with $f_g = 1$ small reductions in the standard deviation will produce more splits.

Figure 7 shows two examples of a split of a segment, using $f_g = 0.5$. The x-axis and the y-axis represents the number of photos in the segment and the spatial gaps between consecutive photos, respectively. The first case, denoted by ❶, illustrates a segment where there are large spatial gaps, with $\Delta_s = 4.38$. In this example, positions 6 and 11 verify (5). Thus, $split_s$ gets three segments, s_1, s_2 and s_3 starting at positions 1, 6 and 11 respectively, as illustrated. To make the split, it is necessary that new segments lessen the distances variance between photos. The depicted spatial gaps δ_1, δ_2, δ_3 and δ_k are used in (6). Do note they omit the first gap of each segment. Since $split_s$ verifies (6), the split is done.
The second case, denoted by ❷, illustrates a scenario where there are no spatial gaps that stand out for difference, with $\Delta_s = 0.67$. All positions, except the first, verifies (5). In this case, $split_s$ gets thirteen singular segments, whose δ_s is undefined. Therefore it does not verify (6), and no split is done.

The *join* operation is performed, after the *split* operation is completed for every segment $s \in S_{evt}$. It takes two contiguous segments, s_1 and s_2, and produces a new segment s_3 where $s_3^- = s_1^- \wedge s_3^+ = s_2^+$. However, for segments belonging to $split_s$, same restrictions are imposed. Namely, only the first and last ones are evaluated. The rational is as follows. It does not make sense to re-join something that was split using a similar criteria — to enhance the spatial cohesion inside each segment. However, the extremes segments can be spatially close to the contiguous segments, indicating they belong to the same event. Considering the original segments in S_{evt}, joining segments can benefit the context coherence, in situations where the spatial location is very similar, thus removing temporal gaps inside an event. The advantage of combining two segments s_1 and s_2 are verified using

$$\frac{\sigma_{s_3}}{(\sigma_{s_1} + \sigma_{s_2}) \times \frac{1}{2}} < 1 - f_g \tag{7}$$

[3] Photos with spatial coordinates neither null nor outliers

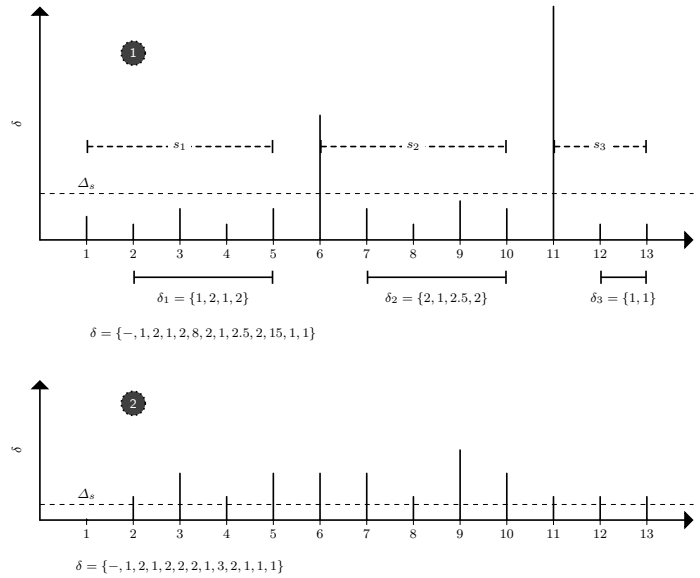


Fig. 7: Examples illustrating the *split* operation.

where s_3 is the segment resulting from the join of s_1 and s_2. The numerator represents the standard deviation of δ_n after the join and the denominator represents the mean of the standard deviation of δ_n in the segments before the join. By definition, whenever the numerator is zero, the left hand side of the equation in (7) is also zero, independently of the value for the denominator. With $f_g = 1$ no join is performed. On the other hand, when $f_g = 0$, just a minor decrement in the standard deviation results in the combination of two segments. Figure 8 shows the conditions where segments can be joined. If they result from a *split*, represented as dashed rectangles, only the first and last segments in $\{s_1, s_2, \ldots, s_n\}$ can be joined with others. For the original segments in S_{evt}, no restrictions are imposed. Cases when the join operation is tried

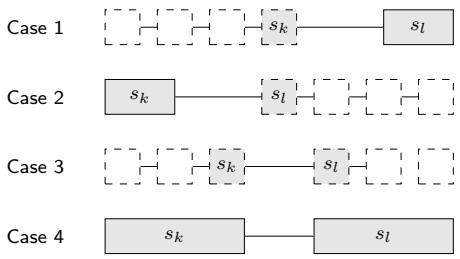


Fig. 8: Cases where the join operation is possible. The segments resulting from a *split* operation are dashed.

Figure 9 shows two examples of the *join* operation, using $f_g = 0.5$. The x-axis

represents the number of photos in the segments and the y-axis represents the spatial gaps between consecutive photos. The first case, denoted by ❶, illustrates the two segments with no obvious advantage to join them, so the division settled using the temporal information is kept. As illustrated, the standard deviation of the spatial gaps in the original segments, δ_{s_1} and δ_{s_2} is very similar to the standard deviation of the joined segments, denoted by δ_{s_3}. Thus, Equation (7) is not satisfied, and the two segments are left untouched.
The second case, denoted by ❷, shows a situation where there is an advantage to join the two segments, since the spatial information represents a similar pattern for both. The join is done, since (7) is satisfied as the standard deviation of the spatial gaps in the resulting segment in zero.

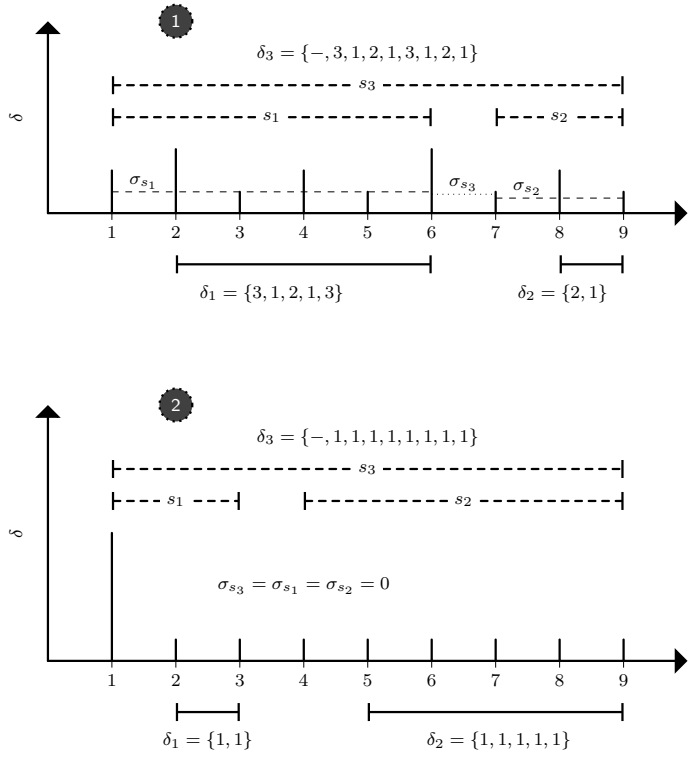


Fig. 9: Examples illustrating the *join* operation.

From the above explanations, we can see that lower values of f_g will produce less segments, and higher values will produce more segments. Setting $f_g = 0.5$ will produce a balance between the splits and joins. An important remark is that no *split* or *join* operation can break the notion of *logical day*.

3.6 Complexity and performance

Given a set P with N photos, the LDES produces a segmentation S. The three steps in the algorithm perform linear operations in P to get S, so the complexity is $\mathcal{O}(n)$, assuming an ordered set P satisfying (1). If we consider this requirement as part LDES complexity, then it changes to $\mathcal{O}(n \log(n))$. In our experiments we use LDES with a stable sort algorithm, and found no performance issues.

4 Experimental set-up

The comparison between LDES and the algorithms described in section 2 was contemplated, but since their goals are different from the LDES, the comparison was not appropriate. Since the gold standard in this domain is the acceptance of the algorithms by the users, we decided to set up a user study to asses LDES segmentation. The empirical study consists in presenting to the participants their photo sets, one by one, segmented using the LDES algorithm, and recording the changes made to the proposed segmentation. There were 14 participants who have provided us with some of their photo sets (35 in total). There was a balance in terms of gender, with eight males and six females. Their age ranged from 21 to 55 years. We provided them with some guidelines to help the photo set selection, namely:

1. The photos should have location information;
2. The temporal range of each photo set should have more than 1 day;
3. The photo sets should reflect real sequences of photos, without a pre-selection.

The participants were responsible for selecting the photo set. For the record, we used all of the provided sets, even though some do not fulfil all three guidelines.

4.1 Characterisation of the photo sets

Table 3: Descriptive statistics for photo sets used in the experimental test.

Stats.	No. Days	No. Photos
Min.	1	5
1st Qu.	1.5	28.5
Median	3	48
Mean	4.2	101.3
3rd Qu.	4	185
Max.	23	351

The 35 photo sets are summarised in Table 3. As we can see, for half of the datasets, the number of photos range from 5 to 351 photos. On average, each dataset has about one hundred photos. Most of them have location information, with 68% of georeferenced photos. However, five photo sets contain only photos with temporal information. In terms of temporal range, about 50% of the photo

sets range between 1 to 4 days approximately. From an observation of the photo sets, we can tell that participants selected them to be, mostly, from weekend trips and one-week holiday. The camera types used to capture the photos are showed in Figure 10. As we can see, more than 60% of the participants choose photo sets taken from their smartphones. Thus, the 68% of geo tagged photos can be explained by this fact, since the smartphone models used are equipped with localization services and have a built-in GPS tracker.

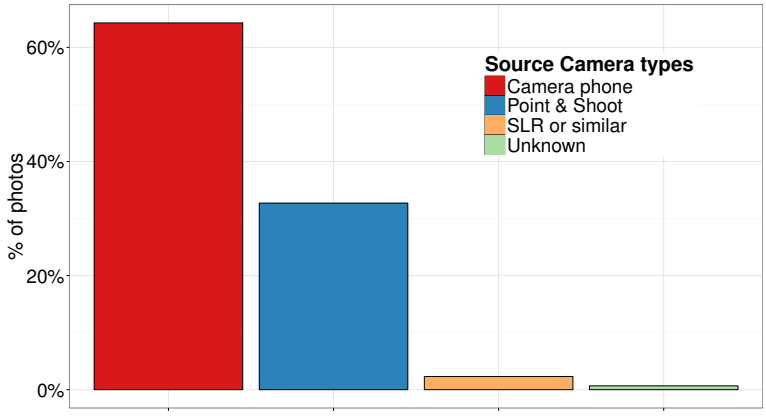


Fig. 10: Type of cameras used to take the photos in the user's photo sets.

4.2 Process description

The experimental unit in this empirical user test is the pair $<participant,photoset>$. There are some participants that share some photo sets. Two are shared between two participants each, and four photo sets are shared between two participants. Those represent situations where the participants are simultaneously the photographer and the subject. The segmentations are presented independently to each user. Thus, the 14 participants, interacting with 35 different photo sets, makes the 43 experimental units in the test.

Figure 11 describes the flow used in the experimental tests. It includes:

1. a *questionnaire*,
2. a *training phase* and;
3. the *test* itself.

4.2.1 Step 1: The questionnaire

In a first phase of the test, users need to complete a questionnaire with general questions about how they manage their photo collections. Most of them are closed-ended questions, presented as follows:

1. *Sex*: ○F ○M;

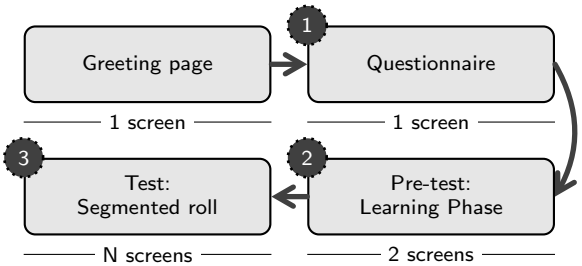


Fig. 11: Representation of the test flow.

2. *Age*: _____([18..120]);
3. *What type of camera you use most?*:
   ☐ Cameraphones (iPhone, Android, ...), ☐ Point and Shoot Cameras, ☐ Reflex or similar;
4. *When you geotag your photos...*:
   ☐ I have no idea what is geotagging, ☐ the camera does it for me, ☐ with a GPS tracker, ☐ by hand, with less detail, ☐ by hand, with as much detail as possible, ☐ I do not geotag photos;
5. *What program do you use to manage your collection of photos?*:
   ☐ None, ☐ Aperture, ☐ Digikam, ☐ F-Spot, ☐ iPhoto, ☐ Picasa, ☐ Lightroom, ☐ Shotwell, ☐ Windows Live Photo Gallery, ☐ Other: _____;
6. *What online library do you use to publish your photos to?*:
   ☐ None, ☐ Facebook, ☐ Flickr, ☐ Instagram, ☐ PicasaWeb, ☐ SmugMug, ☐ Other: _____;
7. *To which online storage service do you save your photos?*:
   ☐ None, ☐ Dropbox, ☐ Google Drive, ☐ iCloud, ☐ Microsoft OneDrive, ☐ MyShoebox, ☐ Other: _____;

With the exception of the first two questions, all the others allow multiple answers.

4.2.2 Step 2: The training phase

After the questionnaire, the participant passes to a *pre-test learning phase*. The goal is to let the participant learn how photos are presented in the segments and how to change the segmentation, permitting an exploratory interaction. The learning phase has two steps, that share a similar layout:

1. a first one, introduces some simple terms and calls the participant attention to the way the segmentation is presented, showing the basic interaction, and
2. a second one, where the participant can freely interact with the user interface (UI), changing the segmentation at will. This includes the creation of new segments, and changing photos from one segment to another.

A participant can repeat the learning steps, as it is possible to go back and forth thorough them. Figure 12 depicts the first step of the learning phase. The left-hand side, marked as ⓐ, displays a representation of a segmentation. The tooltips indicate the locations of the labels, what a segment is, and identifies the contents of a segments — photos. The photos are labelled with letters, so the order can

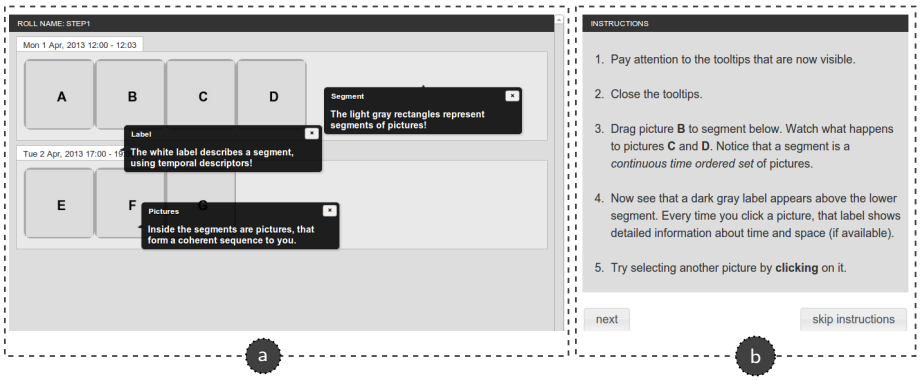


Fig. 12: Interface for the learning phase.

be checked, after the participant changes the segmentation. On the right-hand side, marked as b, the instructions are presented. They are just a simple script of actions, pointing out the artefacts depicted and the way we can interact with them. The left-hand side layout is, in every way, a representation of the test interface. The participant can, at any time skip the instructions. However, before the test begins, a pop-up asks to confirm that action.

4.2.3 Step 3: The test

We take great care on how the participants "see" the segmented collections during the test. We have developed a simple interface, yet, enabling changes to the segmentation. The interface development cycle included several empirical tests (Nielsen 1994) using real users. Those tests allow us to tune the interface and the flow of the dialogues. Figure 13 shows the test interface, depicting a sample photo set. We use some of the Gestalt principles (Johnson et al 2010) for representing a segment. The *proximity* law was applied to the photos of one segment, reinforcing the idea of group. The *figure/ground* law was also applied, making contrast between the background of the test and the foreground of the segments. Both principles reinforce the perception of a group. Each segment is annotated with a short description of the temporal information of the photos in a segment. The annotation consists of the day and hour range, representing the timestamps, to the minute, of the first and last photos in the segment (see Figure 13, a1). If there are photos from more than one day in a single segment, the annotation depicts a range of two dates (see Figure 13, a2). The photos are chronologically presented from left to right, and the segments are chronologically presented from top to bottom. When a photo is selected, it is possible to see more of its spatio-temporal information, as illustrated in Figure 13, b. The interface was also designed to take several guidelines into account (Reeves et al 2004). Namely, the UI:

1. provides *feedback* on the user's action. For example, selecting a photo shows more spatio-temporal information about it (see Figure 13, b);
2. it is *consistent*, as a change in a segment produces changes in annotations;
3. *prevents users from making errors*. The drag and drop facility is available to modify the segmentation. However, the temporal order of the photos is held,

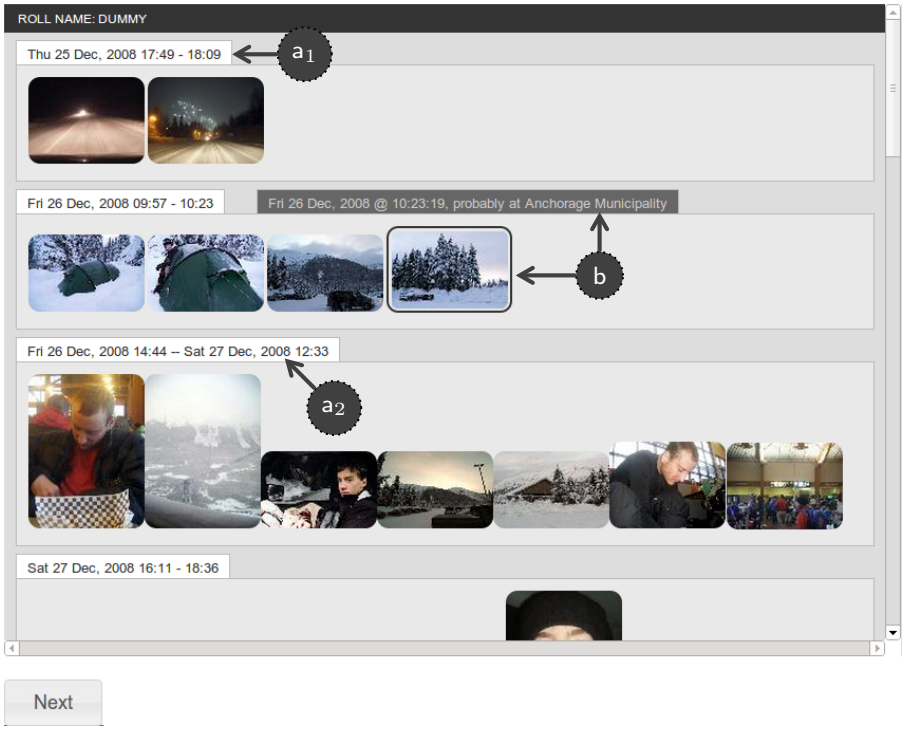


Fig. 13: Example of the test interface.

i.e., if the user drags a photo to the next segment, all the following photos are also moved.

A participant can take two types of actions in the proposed segmentation:

1. `Split`
2. `Join`
   (a) `Move-One`
   (b) `Move-Many`

The `split` action is triggered by selecting a photo and hitting the 's' key. The segment with the selected photo is split in two. The photos until the selected photo are kept in the existing segment. The selected photo and the ones that follows in the same segment are moved to a new segment.
The `join` action is done by drag and dropping photos from one segment to another. The `Move-One` action is a special case of a join, where only one photo is moved. Notice this action consists of moving the first photo from one segment into the previous one, or moving the last photo in a segment into the next one. The `Move-Many` action is another special case of a join. When the selected photo is dragged from one segment and dropped into another (that must be contiguous), other photos are also moved, guaranteeing the temporal order intra- and inter-segment. If the photo is dragged to the predecessor segment, all the preceding photos of the selected one are moved. If the selected photo is moved to the suc-

cessor segment, then all the photos following the selected one are also moved. If all the photos in one segment are moved into another segment, a complete `Join` is performed and the empty segment is removed from the interface. Otherwise, a partial join is done (`Move-One` or `Move-Many`). All the actions are presented and explained to the participant in the learning phase, that precedes the test itself, where the participant in invited to explored them.

During the test, when a participant finishes a test screen, he must rate the LDES proposed segmentation using a 4+1 Likert item, with the options *(i)* 1 (Disliked), 2, 3 and 4 (Liked). The extra option is the *Don't know* choice. Researchers found significant differences when this option is omitted (Lietz 2010), namely, an higher increase on the weak agree/disagree than for the agree/disagree. Nevertheless, we want the participants to take a position, either positive or negative, and since they own the photos, we believed that none will go undecided. Our assumption was confirmed by the results. The question used is short, in a neutral tone, to improve comprehension and to avoid bias, respectively (Lietz 2010).

5 Results

In this section we will show the survey results, towards a characterisation of the participants, and present the key results of this experiment. We have performed

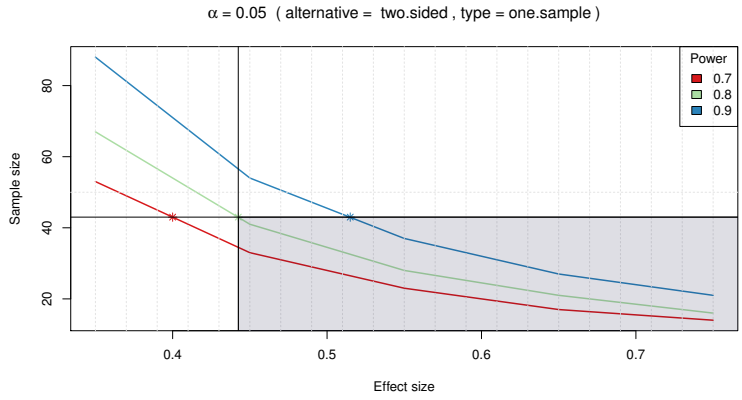


Fig. 14: Power analysis for the given study.

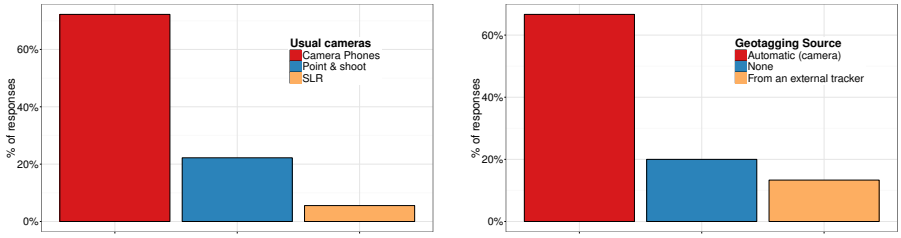
a power analysis for the empirical test, using $\alpha = 0.05$, varying the power from 70% to 90% and the effect size from small (0.2) to large (0.8). As Figure 14 shows, for the current number of experimental units, we can detect medium to large effects sizes[4] with a power greater than 80%, $\beta = 20\%$, (Cohen 1992). Typically, researchers agree this value is the accepted value for a good power (Cohen 1988; Seltman 2012). The shaded area on the bottom right, represents the reachable zone, considering the number of experimental units we tested. Depending on the

[4] The effect size values are for a student's t-distribution.

effect size, we can achieve up to 99% power, for a high effect size (equal to 0.8). Since the research statement is that users do accept the segmentation proposed, something that was confirmed by the experimental test, we have an effect size from medium to high. In fact, for high effect sizes, a sample size of 15 experimental units would suffice.

5.1 Survey analysis

The analysis of the survey enables us to better describe the group of participants in the study, in terms of their habits of capturing/archiving photos. Figures 15(a) and 15(b) show the results of the questions "*What type of camera do you use most?*" and "*When you geotag your photos...*" respectively. As we can see, the

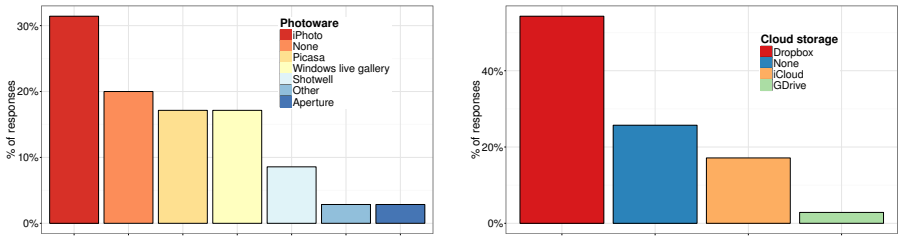


(a) Usual cameras (b) Geotagging sources

Fig. 15: Characterisation of the usual cameras used by the participants and the geotagging source.

cameraphone is the participants primary camera, followed by the point and shoot ones. In fact, almost 93% select *cameraphone*, as a single response or in conjunction with another option. This result (*what users say*) is aligned with the information about the cameras used to capture the photo sets handed for testing purposes *(what users do*, depicted in figure 10). Thus, it reinforces the evidence of the increasing usage of smartphones as the primary camera. Only one participant does not use a camera phone. Thus, it is not surprising that users answer the geottaging source is mostly done by the camera itself, as depicted in Figure 15(b). From the EXIF data, we confirmed that all camera phones are smartphones equipped with a built-in GPS tracker.

Figures 16(a) and 16(b) shows the results for the questions "*What program do you use to manage your collection of photos?*" and "*To which online storage service do you save your photos to?*", respectively. The results show that, despite the fact that most participants manage their photos using specifics programs, 20% do not manage their photos using "photoware". A closer look on their responses reveal that 25% do not use any online storage service, but 85% use social networking sites, specially Instagram and Facebook. This means they still select the photos to post them online. Figure 17 resumes the responses to question "*What online library do you use to publish your photos?*". As we can see, about 15% of the participants do not publish their photos online.



(a) Software used to manage photos (b) Cloud storage used to backup photos

Fig. 16: Storage and organisation.

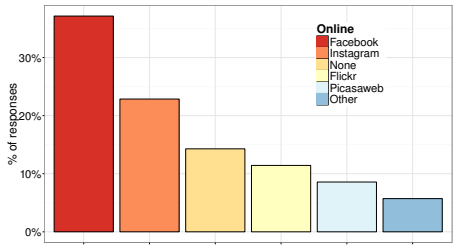


Fig. 17: Social networking sites used by the participants.

From the above results, we can describe the participants as people who use their smartphones to take photos, they later share online. Most of them also use an online storage service to backup their photos, besides the usage of a desktop program to locally store and manage their photos.

5.2 User test analysis

As stated earlier, the research statement that we want to assess in this empirical study is that users will accept the temporal coherent segmentation provided by LDES, with none or minor changes. During the empirical study, we collected several data, namely:

1. the participant's modified segmentation;
2. the stream of actions made on the segmentation, by the participant;
3. the perceived quality of the segmentation.

The LDES parametrisation[5] is displayed in Table 4.

5.2.1 Acceptance of LDES segmentation

From the collected data we are able to derive other indicators, some of which will be used next. One of the indicators is the *number of segments* in the segmenta-

[5] The values for f_t and f_g were settled after testing with 39 personal collections of photos, publicly available at Picasa Web Albuns. Those collections are different from the ones provided by the participants in the study.

Table 4: Parametrization of LDES used in the empirical user test.

LDES	
Parameter	Value
f_t	0.5
f_g	0.5
w	4

tion. The distribution for the number of segments, in the suggested segmentations and in the modified ones, is very similar. This can be observed in the Quantile-Quantile (Q-Q) plot, showed in Figure 18. The values are positively correlated, with $\rho = 0.94$, with a p-value < 0.001. Besides the number of segments, we anal-

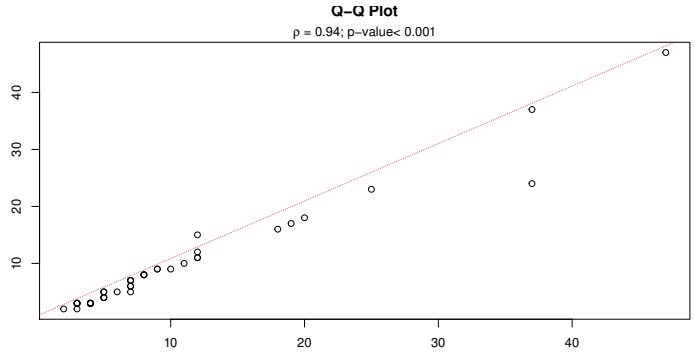


Fig. 18: Quantile-Quantile plot for the no. segments for the suggested segmentations and user-modified segmentations.

ysed the changes made by the participants to the segmentations using 4 actions:

1. `Split`: One segment is divided in two. A new segment is created;
2. `Join`: Two segments are merged into a single segment. One segment is removed;
3. `Move-One`: One photo is moved from one segment to another. No segment is created. One segment may be removed;
4. `Move-Many`: Many photos are moved from one segment to another. No segment is created. One segment may be removed.

Figure 19 shows the distribution of the number of actions made by the users to the proposed segmentations. One result that stands out is that the most common behaviour is not to change the segmentation, representing about 30% of the cases. However, in a few segmentations we can see an higher number of actions. Looking at the data, those cases represent a misalignment between the level of detail (LoD) the users want in certain parts of their photo sets, and the LoD given by LDES. This represents a user preference that cannot be derived from the spatio-temporal information. One of the participants that made more actions said the following

> *Participant 8*: "If the photo set contains photos from a longer vacation (more than a week), I do not want to see it divided lower than the day. However, for a weekend holiday it seems OK."

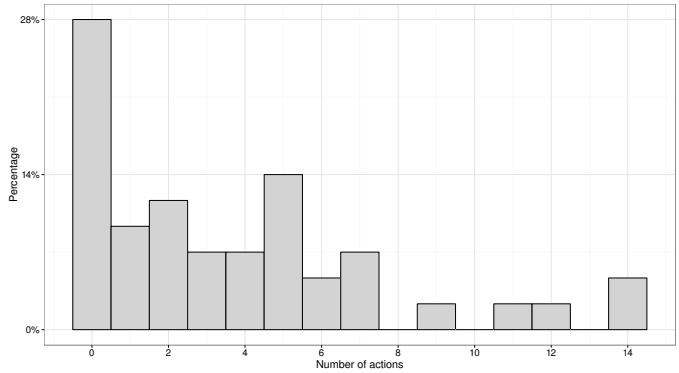


Fig. 19: Distribution of the number of user actions made in the experimental units.

This result is confirmed by the share each type of action has in the changes made by the user, as depicted in Figure 20. The `join` action was the most common, used in 46% of the changes, followed by the `Split` action (23%). Those two are responsible for the higher number of actions made to some of the segmentations. Nevertheless, the over-segmentation (corrected by the `Join`) and the under-segmentation (corrected by the `Split`) are made by the users in small portions of their photo sets, generally confined to one logical day. An important finding is the `Move-One` and `Move-Many` are used occasionally. This shows that, most of the time, important cut points are well identified by LDES. Together, Figures 18, 19 and 20 show a strong evidence that segmentations are accepted by the participants. The number of segments proposed by the algorithm are highly correlated with the ones that are accepted by the users. The actions made by the participants are small, with median < 2.

From the data in Figure 21, it is apparent that participants like the segmentations produced by LDES, since more than $\frac{2}{3}$ of the responses are positive (one sample t-test, $p < 0.001$). Only in 17% of the responses, the participants select

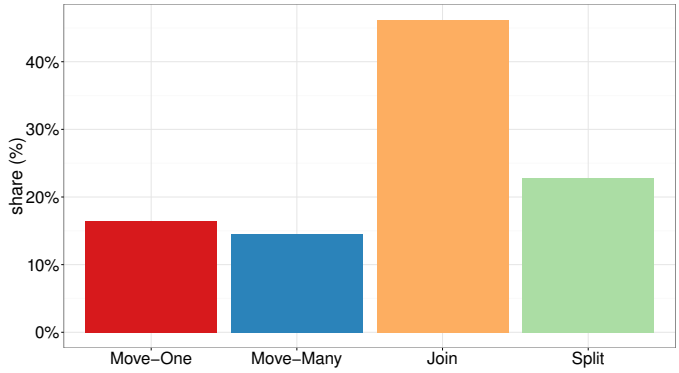


Fig. 20: Share of actions type made by the users to the proposed sgmentations.

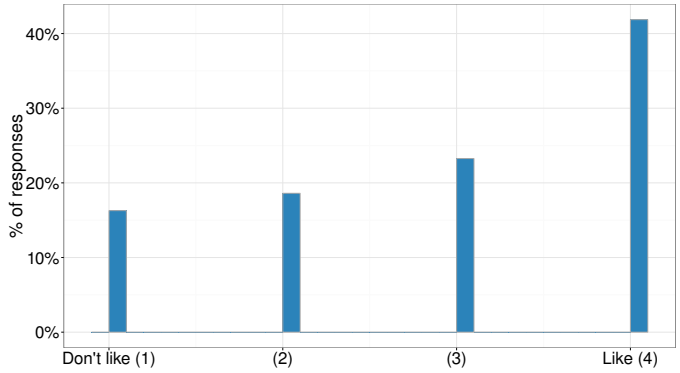


Fig. 21: Responses to the quality of the proposed segmentation.

don't like. The responses for the quality of the segmentation are more diverse, towards the lower part of the scale, when the spatial information is absent. The perceived quality is lower when the geographic information is absent (one sample t-test, $p < 0.001$).

5.2.2 Logical day

LDES introduces the notion of *logical day*, as an important mechanism to match people's behaviour to the notion of the day cycle. We analysed the photo set to see whether the logical days are different from the standard day. It was found that on 7% of the cases they differ. The analysis of the segmentations modified by the participants shows that 100% of the logical days were kept. These results suggest the concept is important to maintain a temporal coherence in the segments, going beyond the strict boundaries of temporal cycles. Further analysis showed that some participants join, sparsely, photos from two days, producing *multi-day* segments. The statistical tests revealed that multi-day segments appear specially in segmentations having higher cardinality[6] (one sample t-test, p-value < 0.01). However, this was done sparsely, without an apparent criteria. This may indicate the size of the segmentation (which may or not be directly related with the temporal and spacial range) may influence the perception users have of the context.

5.2.3 Singular segments

Other important finding is that singular segments are kept by the participants. The distribution for the number of singular segments in each experimental unit, considering the suggested segmentation and the ones that exist in the final segmentation, after the participants made their changes are very similar. They have a similar distribution, and they are positively correlated, with $\rho = 0.97$, with a p-value < 0.001. These are interesting results, as it seems that LDES is capable to isolate photos that have their own context. This situation is becoming more

[6] No. of segments greater than the median

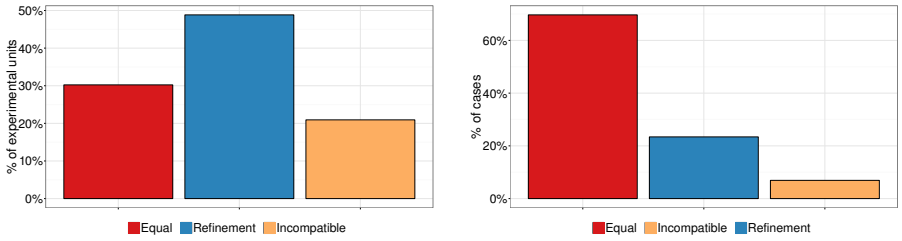
frequent in personal photo collections, as there are more photos taken with smartphones. Further statistical tests revealed that singular segments in the participants segmentations are more frequent in the segmentations with higher cardinality (one sample t-test, p-value < 0.01).

5.2.4 Relations and measures

We have compared the segmentations before and after the participants intervention, using qualitative and quantitative analysis. In the first case, we used the binary relations introduced in Section 3.1. In the second case, we used the PR_{error} (Georgescul et al 2006) and the WindowDiff (Pevzner and Hearst 2002) measures to compare the segmentations. The PR_{error} was set in three different scenarios:

1. *equal costs* for miss and false positive (FP);
2. *FP costs are three times higher* than miss costs;
3. *miss costs are three times higher* than FP costs.

Figures 22(a) and 22(b) present the results of the qualitative comparison between LDES segmentations and the ones the participants submitted during the test. In case (a), the segmentations are compared as a whole. However, to better understand the behaviour of participants, we decided to analyse separately each logical day inside the segmentations (b). In this case we do not consider the situations of multi-day segments. It can be seen from Figure 22(a) that almost 80% of the



(a) Qualitative comparison between LDES segmentations and modified segmentations

(b) Qualitative comparison between LDES segmentations and modified segmentations, considering the division in logical days

Fig. 22: Qualitative comparison between suggested segmentations and user-modified segmentations.

segmentations are compatible, 30% of which are `equal`. This means that most of the time, the important cut points are well identified by LDES. The `refinement` relation, which represents almost 50% of the cases, tells us the participants need to insert or remove cut points. In either way, such behaviour represents a difference in the level of detail the participants want to see, and the level of detail LDES provides. However, as shown in Figure 22(b), these changes in cut points happen in few segments, since almost 70% of the logical days were perfectly segmented by LDES. These results are in line with the previous evidences found when analysing

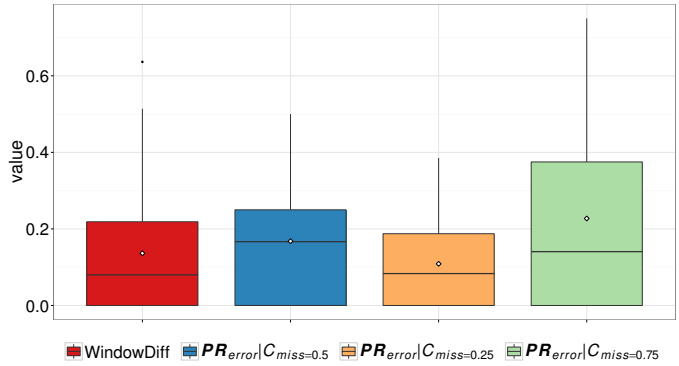


Fig. 23: Distance measures between LDES segmentations and participant's modified segmentations.

the number and type of actions made to the segmentations by the users. The quantitative analysis confirms the results shown above. From the data in Figure 23, it is apparent that the segmentations are similar. The medians, in all scenarios, are above 0.2. The upper quartiles for the two measures, considering PR_{error} with equal costs, are under 0.3. The sizes of the first two boxplots are similar, indicating that there is not such great difference between measures.

Another interesting observation is there are more misses than false positives, given the values for different parametrisations of PR_{error}. This is in line with the characteristics of the measures. WindowDiff penalises all pure false positive the same way, regardless of how close they are to an actual boundary (Pevzner and Hearst 2002) and misses are less penalised than false positives (Georgescul et al 2006). Those characteristics explain the lower values in WindowDiff and the behaviour of PR_{error} when the cost of a miss is changed. This also reinforces the results depicted in Figure 22(a), where almost 50% of the segmentations have a `Refinement` relation between them, as a result of the `Join` being the action most used by the participants (see Figure 20). Thus, when participants changed the segmentations, they mainly lowered the level of detail for some days in their photo sets. This was done removing cut points, but maintaining the ones that are key to separate the context.

6 Conclusions and future work

This paper has given an account of the importance of temporal and spatial information for separating contexts in personal photo collections. In this investigation, the aim was to assess if users accepted an automatic segmentation of their own photo sets, that is temporal dominant, and incorporates spatial information. The segmentation algorithm, LDES, incorporates the notion of logical day to adapt the segments to the perception people have about where a day ends.

During the test, if the users are not satisfied with the segmentation of their photo sets, they can change it, dragging photos from one segment to another. The most obvious finding to emerge from this test is that the segmentations pre- and

post-user action are almost equal. This is supported not only by the similar number of segments in each, but also in the small amount of changes the users made. It was also verified that, among the four actions available to the users to change the segmentations, `Join` was the most common. The relevance of LDES is clearly supported by the current findings, where users have a positive reaction to the segmentation of their photo sets. Together, these results suggest the users accept the LDES segmentation. They made few, or no modifications to the segmentations, and when they did, it was mainly to correct the *Level Of Detail* (*LoD*) in some parts of the photo set. Changing the *LoD* is made by reducing the number of segments, maintaining the key cut points. Nevertheless, this was done sparsely. This is confirmed by the response to the perceived quality of the segmentation, that was positive, with a tendency towards the highest value.

The results also support the relevance of the *Logical Day*, as a key piece to settle context boundaries. This indicates that incorporating temporal cycles is important, if they are modelled to the way users perceive them. The test has demonstrated the acceptance of singular segments by the users. Such result is important to assess the LDES capability to detect isolated photos, a common feature in photo sets gathered from smartphones.

To our best knowledge, this is the first time that binary relations are use to explore the relation between segmentations in personal photo collections, exploring a qualitative approach complementary to a quantitative one. During the analysis, it became clear they are an important tool to understand or confirm the results of the tests. The quantitative metrics, despite their importance, are insufficient for a complete analysis of the data.

This research has thrown up some questions in need of further investigation. One relates to the over-segmentation in some parts of the photo set. It would be interesting to assess what causes such behaviour. Other question we need to answer is if the time frame of the photo set influences the need for a less segmented photo set. And if so, are a proper parametrization of the algorithm capable to deal with it or not.

A natural progression of this work is to analyse how LDES can be extended to support hierarchical segmentation, incorporating larger cycles, in particular the weekly cycle. There should be more research to understand if the logical day concept can be extended to the upper cycle or, if at that level of detail, the use of standard boundaries is enough for users. This is an important issue if the number of photos to archive is high, and spans through a large time frame.

References

Allen J (1983) Maintaining knowledge about temporal intervals. Communications of the ACM 26(11):832–843

Breunig MM, Kriegel HP, Ng RT, Sander J (2000) Lof: Identifying density-based local outliers. SIGMOD Rec 29(2):93–104, DOI 10.1145/335191.335388

Bruneau P, Pigeau A, Gelgon M, Picarougne F (2010) Geo-temporal structuring of a personal image database with two-level variational-bayes mixture estimation. In: Detyniecki M, Leiner U, Nrnberger A (eds) Adaptive Multimedia Retrieval. Identifying, Summarizing, and Recommending Image and Music, Lecture Notes in Computer Science, vol 5811, Springer Berlin Heidelberg, pp 127–139, DOI 10.1007/978-3-642-14758-6_11

Cao L, Luo J, Kautz HS, Huang TS (2008) Annotating collections of photos using hierarchical event and scene models. In: CVPR, IEEE Computer Society, DOI 10.1109/CVPR.2008.4587382

Cobley P, Haeffner N (2009) Digital cameras and domestic photography: communication, agency and structure. Visual Communication 8(2):123–146

Cohen J (1988) Statistical power analysis for the behavioral sciences. Psychology Press

Cohen J (1992) A power primer. Psychological bulletin 112(1):155

Comaniciu D, Meer P (2002) Mean shift: A robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(5):603–619

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

Cooper ML (2011) Clustering geo-tagged photo collections using dynamic programming. In: Proceedings of the 19th ACM International Conference on Multimedia, ACM, New York, NY, USA, MM '11, pp 1025–1028, DOI 10.1145/2072298.2071929

Datia N, Moura-Pires J, Correia N (2014) Summarised presentation of personal photo sets. In: Gurrin C, Hopfgartner F, Hurst W, Johansen H, Lee H, OConnor N (eds) MultiMedia Modeling, Lecture Notes in Computer Science, vol 8325, Springer International Publishing, pp 195–206, DOI 10.1007/978-3-319-04114-8_17

Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: Ideas, influences, and trends of the new age. ACM Comput Surv 40(2):1–60, DOI 10.1145/1348246.1348248

Do TMT, Blom J, Gatica-Perez D (2011) Smartphone usage in the wild: a large-scale analysis of applications and context. In: Proceedings of the 13th international conference on multimodal interfaces, ACM, New York, NY, USA, ICMI '11, pp 353–360, DOI 10.1145/2070481.2070550

Foote J (2000) Automatic audio segmentation using a measure of audio novelty. In: Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, IEEE, vol 1, pp 452–455

Friedman W (2004) Time in autobiographical memory. Social Cognition 22(Special issue):591–605, DOI 10.1521/soco.22.5.591.50766

Gargi U (2003) Consumer media capture: Time-based analysis and event clustering. Tech. rep., Technical Report HPL-2003-165, HP Laboratories

Georgescul M, Clark A, Armstrong S (2006) An analysis of quantitative aspects in the evaluation of thematic segmentation algorithms. In: Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, Association for Computational Linguistics, pp 144–151

Gozali J, Kan M, Sundaram H (2012) Hidden markov model for event photo stream segmentation. In: Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on, IEEE, pp 25–30

Graham A, Garcia-Molina H, Paepcke A, Winograd T (2002) Time as essence for photo browsing through personal digital libraries. Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries pp 326–335

Gye L (2007) Picture this: the impact of mobile camera phones on personal photographic practices. Continuum 21(2):279–288

House NAV (2009) Collocated photo sharing, story-telling, and the performance of self. International Journal of Human-Computer Studies 67(12):1073 – 1086, DOI 10.1016/j.ijhcs.2009.09.003

Janssen S, Chessa A, Murre J (2006) Memory for time: How people date events. Memory and Cognition 34(1):138

Johnson J, et al (2010) Designing with the mind in mind: Simple guide to understanding user interface design rules. Morgan Kaufmann

Kang H, Bederson BB, Suh B (2007) Capture, annotate, browse, find, share: Novel interfaces for personal photo management. International Journal of Human-Computer Interaction 23(3):315–337, DOI 10.1080/10447310701702618

Kellerman A (1989) Time, space, and society: geographical societal perspectives. Kluwer Academic Pub

Kirk DS, Sellen A (2010) On human remains: Values and practice in the home archiving of cherished objects. ACM Transactions on Computer-Human Interaction (TOCHI) 17(3):10, DOI 10.1145/1806923.1806924

Kwok SC, Shallice T, Macaluso E (2012) Functional anatomy of temporal organisation and domain-specificity of episodic memory retrieval. Neuropsychologia 50(12):2943 – 2955,

DOI http://dx.doi.org/10.1016/j.neuropsychologia.2012.07.025
Latif K, Mustofa K, Tjoa A (2006) An approach for a personal information management system for photos of a lifetime by exploiting semantics. In: Bressan S, Kng J, Wagner R (eds) Database and Expert Systems Applications, Lecture Notes in Computer Science, vol 4080, Springer Berlin Heidelberg, pp 467–477, DOI 10.1007/11827405_46
Lietz P (2010) Research into questionnaire design. International Journal of Market Research 52(2):249–272
Loui A, Savakis A (2003) Automated event clustering and quality screening of consumer pictures for digital albuming. IEEE Transactions on Multimedia 5:390–402
Lux M, Kogler M, del Fabro M (2010) Why did you take this photo: a study on user intentions in digital photo productions. In: Proceedings of the 2010 ACM workshop on Social, adaptive and personalized multimedia interaction and access, ACM, New York, NY, USA, SAPMIA '10, pp 41–44, DOI 10.1145/1878061.1878075
McGill R, Tukey JW, Larsen WA (1978) Variations of box plots. The American Statistician 32(1):12–16
Naaman M, Song YJ, Paepcke A, Garcia-Molina H (2004) Automatic organization for digital photographs with geographic coordinates. In: JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, ACM Press, New York, NY, USA, pp 53–62
Nielsen J (1994) Usability inspection methods. In: Conference companion on Human factors in computing systems, ACM, pp 413–414
Pevzner L, Hearst MA (2002) A critique and improvement of an evaluation metric for text segmentation. Computational Linguistics 28(1):19–36
Platt J, Czerwinski M, Field B (2003) Phototoc: automatic clustering for browsing personal photographs. Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia Proceedings of the 2003 Joint Conference of the Fourth International Conference on 1:6–10 Vol.1
Reeves LM, Lai J, Larson JA, Oviatt S, Balaji TS, Buisine S, Collings P, Cohen P, Kraal B, Martin JC, McTear M, Raman T, Stanney KM, Su H, Wang QY (2004) Guidelines for multimodal user interface design. Commun ACM 47(1):57–59
Seltman HJ (2012) Experimental design and analysis. Online at: http://www stat cmu edu/, hseltman/309/Book/Book pdf
St Jacques P, Rubin D, LaBar K, Cabeza R (2008) The short and long of it: Neural correlates of temporal-order memory for autobiographical events. Journal of Cognitive Neuroscience 20(7):1327–1341, cited By (since 1996)29
Sun F, Li H, Wang X (2013) Photo 4w: Mobile photo management on what, where, who and when. Neurocomputing 119:59 – 64, DOI http://dx.doi.org/10.1016/j.neucom.2012.03.038, intelligent Processing Techniques for Semantic-based Image and Video Retrieval
Sun Y, Zhang H, Zhang L, Li M (2002) Myphotos: a system for home photo management and processing. In: MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia, ACM Press, New York, NY, USA, pp 81–82, DOI http://doi.acm.org/10.1145/641007.641022
Tulving E (2002) Episodic memory: From mind to brain. Annual review of psychology 53(1):1–25
Viana W, Bringel Filho J, Gensel J, Villanova-Oliver M, Martin H (2008) PhotoMap: from location and time to context-aware photo annotations. Journal of Location Based Services 2(3):211–235
von Watzdorf S, Michahelles F (2010) Accuracy of positioning data on smartphones. In: Proceedings of the 3rd International Workshop on Location and the Web, ACM, p 2
Whittaker S, Bergman O, Clough P (2010) Easy on that trigger dad: a study of long term family photo retrieval. Personal and Ubiquitous Computing 14(1):31–43
Zerubavel E (1985) Hidden Rhythms: Schedules and Calendars in Social Life. University of California Press
Zerubavel E (1996) Social memories: Steps to a sociology of the past. Qualitative Sociology 19(3):283–299
Zhao M, Teo Y, Liu S, Chua TS, Jain R (2006) Automatic person annotation of family photo album. In: Sundaram H, Naphade M, Smith J, Rui Y (eds) Image and Video Retrieval, Lecture Notes in Computer Science, vol 4071, Springer Berlin Heidelberg, pp 163–172, DOI 10.1007/11788034_17

Zuzanek J, Smale J (1993) Life-cycle variations in across-the-week allocation of time to selected daily activities. SOCIETY AND LEISURE-MONTREAL- 15:559–559