HAL
open science

Organizing a personal image collection with statistical model-based ICL clustering on spatio-temporal camera phone meta-data

Antoine Pigeau, Marc Gelgon

To cite this version:

Antoine Pigeau, Marc Gelgon. Organizing a personal image collection with statistical model-based ICL clustering on spatio-temporal camera phone meta-data. Journal of Visual Communication and Image Representation, 2003, 15 (3), pp.425-445. hal-00415968

HAL Id: hal-00415968

https://hal.science/hal-00415968

Submitted on 11 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Organizing a personal image collection with statistical model-based ICL clustering on spatio-temporal camera phone meta-data

A. Pigeau, M. Gelgon

IRIN / INRIA ATLAS group, Nantes university
2, rue de la Houssinière - BP 92208, 44322 Nantes cedex 03 - France
Tel: +(33)2.40.68.32.57 Fax: +(33)2.40.68.32.32

Abstract

This paper addresses the issue of automated organization of a personal image collection, in particular to respond to the emerging needs from a mobile camera phones. The issues related to browsing through large image collections acquired from such devices are first discussed. In contrast with retrieval in metadata-less collections, which necessarily relies on image content, we propose a collection organization technique based on picture geolocation and timestamps. These descriptors are indeed available and generally reliable in the proposed context. Collection organization is formulated as an unsupervised classification problem, in both space and time. The statistical integrated completed likelihood criterion is chosen, providing effective solutions both to model complexity determination and the cluster separability objective, in a setting which limits arbitrary algorithm parametrization. Reliability of space and time partitions obtained are then assessed to select a segmentation, which may then provide a calendar-type structured view for navigating in the picture collection.

Key words: Image retrieval, mobile applications, spatio-temporal metadata, statistical clustering, mixture modelling

1 Problem and context

Content-based image retrieval problems have been dealt with for the past few years, and the field is now equipped with both recent contributions [1] and surveys [2]. It was stressed in [3] that extending the scope of this field to

Email address: `surname@irin.univ-nantes.fr` (A. Pigeau, M. Gelgon).

address personal image collections was an important stake, and that there was little existing. Very recently, proposals have emerged from most leading industrials in the field of multimedia [4–8], as well as from academics [9,10]. A first survey on the matter is to appear in [11].

We advocate that, within personal image collections, the following interesting niche is emerging. Compared with workstations, or even digital cameras, *camera phones* are always carried by their user, and are hence advantageous both as an image acquisition device and as an image retrieval terminal [12]. Indeed, despite the technical issues image retrieval sets on such devices, their permanent availability and the ability to easily share retrieved pictures (eg. through Multimedia Mobile Messaging) makes this context propitious for building large collections. A further quality of camera phones resides in their excellent geolocation potential, which founds our proposal. Indeed, they can switch between, or fuse, measurements from an embedded GPS receiver with infrastructure-based (GSM/UMTS) position estimation tools, such as E-OTD. Interestingly, these network techniques enable fast initialization of the GPS receiver, which is a required feature for comfortable usage without battery-consuming always-on GPS. Finally, let us point out that these means fortunately complement each other well, in terms of availability and accuracy (e.g. poor GPS reception in "urban canyons" is generally made for by a high density cell zone in the telecom network).

As images are gathered, they progressively build up a valuable memory of one's life, which can be later searched, whether for practical, emotional, or pass-time purposes. The system should offer users not only the possibility of quickly, reliably and comfortably retrieve a well-defined piece of information in their potentially large collection, but also functions for browsing, simply to get an overall idea of the content of the collection. Providing such overviews is actually also beneficial, let alone necessary, for retrieving a well-defined piece of information.

A survey of technical issues pertaining to camera phones is provided in [6], while insight in their usage may be found in [12]. One of the conclusions drawn in [6] is that the handling (storage and retrieval) of image collections in this context is an important industrial need and overall an open problem. Furthermore, we believe this issue is both well-defined (probably better than, e.g., all world-wide-web content-based image retrieval), and is of practical significance in the years to come.

Personal image collections may be distinguished from the ordinary "digital library" viewpoint by :

- the content itself (nature of the scenes, structure of the image collection, attached meta-data);

- the partial memory that the user has of the collection : the user does not discover a wholly new collection, but progressively recalls his past as navigation progresses, thus determining incrementally browsing directions. Browsing presents advantages over querying to this respect;
- the desired search/browse criteria : user studies [13,14] suggest that the most convenient browsing criteria are time, geographical location, and "image annotations/semantic image content/topic" (topic would often be in the form of an audio note). The identity of people present is also among appealing criteria. On the other side, classical features such as color, shape, layout and texture cues are rated of little relevance.

Research in content-based image retrieval has often been justified by the lack of meta-data. With camera phones, both time and location measurements may be assumed, corresponding to very criteria that users wish to use.

The focus of the present paper is the automated generation of a structured representation of the image collection, attempting to recover meaningful episodes and areas. This should allow the user to effectively browse through time and space, yet keeping manual organisation only optional. In this paper, we solely consider time and geolocation meta-data attached to each image, while the image content itself is ignored. Further, we wish to make the scheme as unsupervised as possible, i.e. the temporal and spatial bounds of the image groups, and the number of these groups should be, as much as possible, driven by the data.

Finally, human-computer interactions considerations are important, as they may define or affect content analysis goals. Research contributions that have been put forward mainly target the workstation/e-book context, and propose a variety of strategies for laying out image sets [15,10]. In the case of small mobile devices with stringent input and display constraints, hardly more than one to three images may be displayed simultaneously. We argued in [16,17] that a crux is the ability to generate summaries according to the criteria identified above as relevant, so as to suit visualization and browsing needs. The present work proposes a contribution in this direction.

Let us point out that structuring the personal image collection may also benefit also system-level efficiency. Indeed, a personal image collection should be accessible both from the small mobile terminal and from e.g. a home-based, larger terminal. For secure storage with multi-terminal access, a distributed storage system should be envisaged [18], with some degree of laziness in replication, taking into account the variation of network conditions and transmission costs (eg. WLAN vs. GPRS/GSM). In this context, handling only higher levels of the image collection hierarchy (that provide collection overviews) can help reduce traffic. Besides, pre-fetching and caching strategies can exploit this collection structure, in conjunction with multi-terminal content adapta-

tion layers, as described in *Universal Multimedia Access* MPEG-7 effort [19]. The present paper does not discuss this direction further, but attention is drawn on the benefit one may expect on this aspect.

The remainder of this paper is organized as follows. Section 2 surveys existing proposals that exploit time or space for retrieval in image collections. In section 3, we propose a technique for spatio-temporal organization of one's image collection, first as an overview and then in more detail. Section 4 provides experimental results, while section 5 depicts a typical user interface exploiting the results. Finally, section 6 is devoted to concluding remarks.

2 Related work

2.1 Time-based structuring

Structuring an image collection according to the time stamp of each picture is intuitively appealing, practically quite cheap and reliable. As noted in [9], the generative process of pictures (i.e. behaviour of users) is likely to exhibit time clusters and, furthermore, often in a hierarchical fashion. Overall, two types of techniques may be distinguished. First, change detection techniques, such as in [4], possess the advantage of not setting a particular parametric model on the intra-cluster time distribution. A combination of these alternatives is proposed in [9], in which (preset size)-gap detection leads to initial groups for clustering. However, how the classical limitations of clustering techniques are addressed is not detailed (number of clusters, arbitrary intra/inter-class separation thresholding). In order to cope with the variety of time scales present in the image collection, solutions such as log-scaling of inter-frame time gaps have been examined in [4,8]. Finally, besides direct use of time for image grouping, it was recently proposed [7] to combine time linearly with camera settings features and image content information, within an 'image similarity' measure.

2.2 Geolocation-based structuring

As recently reviewed in [20], geolocation technologies are progressively being integrated in mobile phones and networks. In practice, geolocation is mainly pushed on the market by navigation and context-aware services [21], rather than our image retrieval purpose, but we nevertheless benefit from it and it is of utmost importance.

The importance of the location for image collection organization is stressed

in [9] but, to our knowledge, there are currently few systems that seem to have considered the matter closely. Regardless of image consideration but still with a view to providing structured calendar-type views, we proposed in [22] a technique towards unsupervised learning of meaningful locations. Geolocation is measured continuously in time, and partitions of time and space are extracted at multiple scales, based on a piece-wise parametric trajectory model. By this means, one attempts to recover significant temporal episodes and areas. A work close in spirit is [23], although the modelling formalism differs.

3 Spatio-temporal organization with model-based ICL clustering

3.1 Meta-data used and overview of the proposed approach

We formulate the recovery of the image collection spatio-temporal structure as a model-based unsupervised classification problem. Probabilistic model-based clustering is a favourite framework for identifying meaningful groups in data [24]. Its adequacy for our needs is justified by *i)* the intra-cluster homogeneity criterion, which is explicit in the form of a parametric model (the drawback arising from this aspect is discussed further down), *ii)* the possibility of assessing classifications obtained, *iii)* the manner in which the process scales well with large amounts of data (say, a few hundreds or thousands of pictures, in our case) and has good perspectives towards an incremental version of the scheme.

Overall, this approach requires setting a functional form on the probability distribution for the data arising from each cluster, defining a statistical optimality criterion and searching for a suitable solution accordingly.

In our case, the data D is assumed to be drawn from a random Gaussian mixture process with probability density :

$$p(D) = \sum_{k=1}^{K} \alpha_k . \mathcal{N}(D|\mu_k, \Sigma_k), \tag{1}$$

where the probabilities α_k are the mixing proportions and $\mathcal{N}(D|\mu, \Sigma)$ indicates a Gaussian distribution with mean μ and covariance Σ.

Here are the main features of the proposed scheme, which will be detailed in the remainder of this section :

- distinct classifications are built for time and space;

- we resort to the statistical integrated completed likelihood criterion (ICL), initially proposed in [25]. With its evidence-like aspect, it provides an effective solution to model complexity determination, i.e. determining a suitable number of clusters. Besides, it drives the search towards favouring cluster separability, thereby providing partitions that are more meaningful for our purpose. An important resulting advantage for our application is improved robustness to mismatch between the (Gaussian) form of probabilistic model components and the actual clusters;
- optimization of this criterion is conducted using an Expectation-Maximization technique, with a dedicated search procedure (implying multiple short searches and cluster split/merge trials);
- the best spatial and temporal classifications found are assessed, with respect to how well they succeed in separating the data into clear groups. Among these two, the partition retained should then provide a structured, calendar-type (or map-type) organization and view on the picture collection.

3.2 Optimality criterion

By taking a Bayesian hypotheses testing viewpoint, it can be shown that an effective manner of evaluating the ability of a clustering hypothesis H_K to explain the data D, taking into account the need for comparing hypotheses with various numbers of clusters, is provided by the so-called *evidence*, or marginalized likelihood :

$$P(D|H_K) = \int P(D|\Theta_K, H_K)P(\Theta_K|H_K)d\Theta_K, \tag{2}$$

where Θ_K indicates the model parameter vector associated to hypothesis H_K. In our case, $\Theta_K = (\theta_1, \theta_2, \ldots, \theta_K)$ with $\theta_i = (\mu_i, \Sigma_i, \alpha_i)$ $1 \leq i \leq K$.

The goal is thus to find the mixture model leading to the greatest evidence for clustering the data. In order to evaluate this marginalized likelihood, a variety of computation strategies and approximations exist, as reviewed in [26]. We opt here for the approximation known as the Bayesian Information Criterion (BIC) [24], expressed as follows :

$$BIC = -ML + \frac{1}{2} \cdot N(K) \cdot log(n), \tag{3}$$

where ML is the maximized mixture loglikelihood, $N(K)$ is the number of independent parameters in the model with K components and n is the number of data elements. An intuitive observation on this approximation is that it acts as a likelihood criterion penalized by model complexity, illustrating the use of marginalized likelihood as the Bayesian implementation of Occam's razor.

Overall, the BIC criterion aims at identifying both model parameters and the number of clusters. Yet, the parametric form enforced for clusters (henceforth, in practise, Gaussianity) leads to poor and/or over-segmented clusters, should their data strongly infringe on this assumption. In our case, groups of pictures forming meaningful areas or time extents do not generally exhibit very Gaussian distributions, as their physical generative process is not so.

To improve this point, we resort to the Integrated Classification Likelihood (ICL), proposed in [25]. Overall, this criterion penalizes the BIC criterion with the entropy of the data-to-model assignment in the mixture, i.e. a mixture which separates well the data has lower entropy and is explicitely favoured. This criterion may be summarized as follows:

$$ICL = BIC - \Phi(K), \tag{4}$$

where $\Phi(K)$ is an entropy-based criterion, defined by:

$$\Phi(K) = -\sum_{k=1}^{K}\sum_{i=1}^{n} t_{ik} \cdot log(t_{ik}), \tag{5}$$

where K is the number of components in the mixture, n the size of the data set, and t_{ik} is the posterior probability for an observation i of originating from cluster k. These t_{ik} values are supplied at convergence of the optimization phase, which is described in the next section.

We provide hereunder a few experimental examples illustrating the benefit of the proposed technique. Fig. 1 depicts a situation where artificial data is generated from a mixture of Gaussians (fig. 1a) (the numbers of samples per component differ). The clustering corresponding to the true underlying mixture is indicated on fig. 1b. It differs slightly from the optimal clustering recovered with the ICL criterion (fig. 1d), as in the latter case, the grouping of two clusters '5' and '6' is preferred. The various model complexities between 1 and 10 clusters are scanned systematically and it is seen that 5 clusters is found to be optimal (fig. 1c). From a clustering perspective of our application, this conclusion seems to lead to more meaningful groups in the data, showing some degree of robustness to non-Gaussianity of the ('5+6') cluster. The next two experiments, reported on fig. 2, illustrate the same idea, but on real data, in the context of our application.

3.2.1 Addressing the small sample problem

The number of data elements in a cluster may be small, leading to inaccuracy in the estimation of the covariance matrices describing the intra-cluster spread,

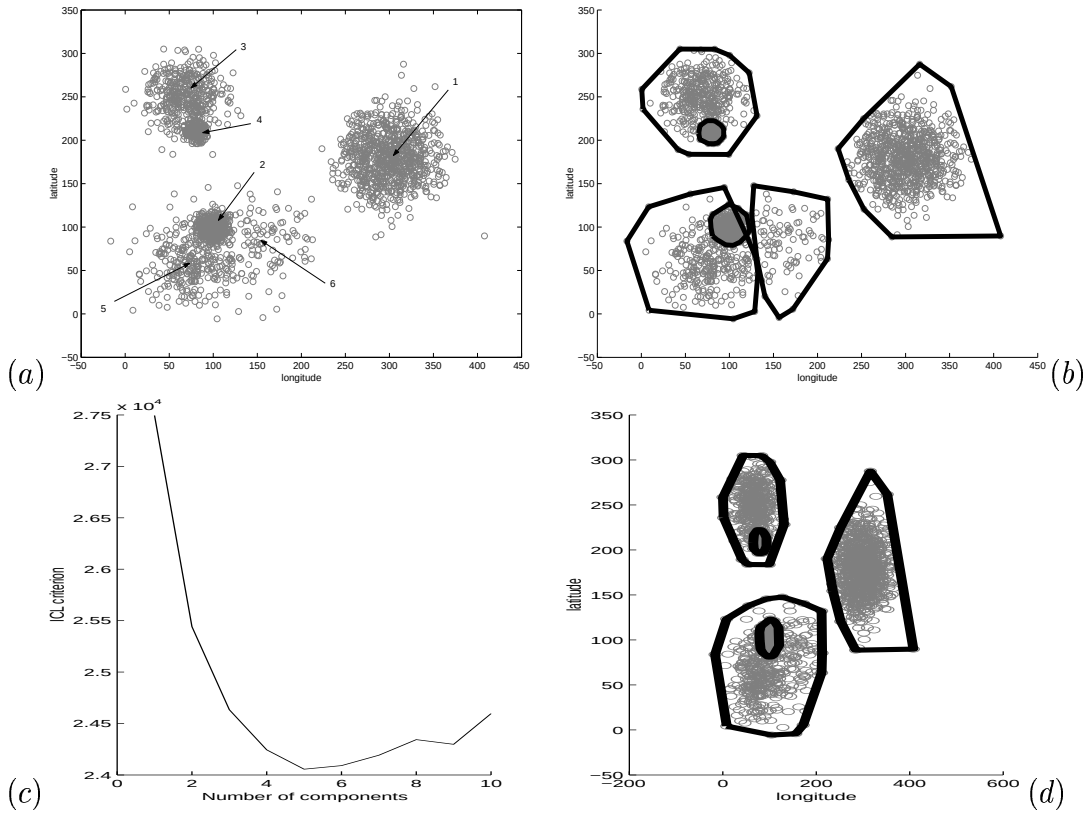

Fig. 1. (a) Artificial data from a mixture of 6 Gaussians (the digits identify each component) and (b) the corresponding clusters (convex hulls) identified by EM, given that there are 6 components. (c, d) respectively describe the optimal value of the ICL criterion obtained for each tentative number of clusters and the overall solution retained using the ICL criterion.

let alone ill-definition if the number of data elements is small than, or equal to, the dimension of the space at hand. We focus here on the geolocation-based clustering, as time-based clustering may be tackled in a similar (yet simpler) way.

Several techniques for regularization of covariance matrix estimation are presented in e.g. [27], which defines the regularized matrix as a linear combination of the sample (full-form) matrix estimate and some other, more constrained, form of estimate. Alternatively, one may select one among several competing covariance matrix forms (such as full, diagonal and spherical), and trade-off model complexity for data fitting.

The solution we introduce consists in estimating a full covariance matrix, yet imposing two constraints :

- *a "minimum area" to each cluster is imposed, by setting minimum covariance eigenvalues.* This enables effective handling of single-data clusters in the classification, through the limitation of the likelihood increase such zero-

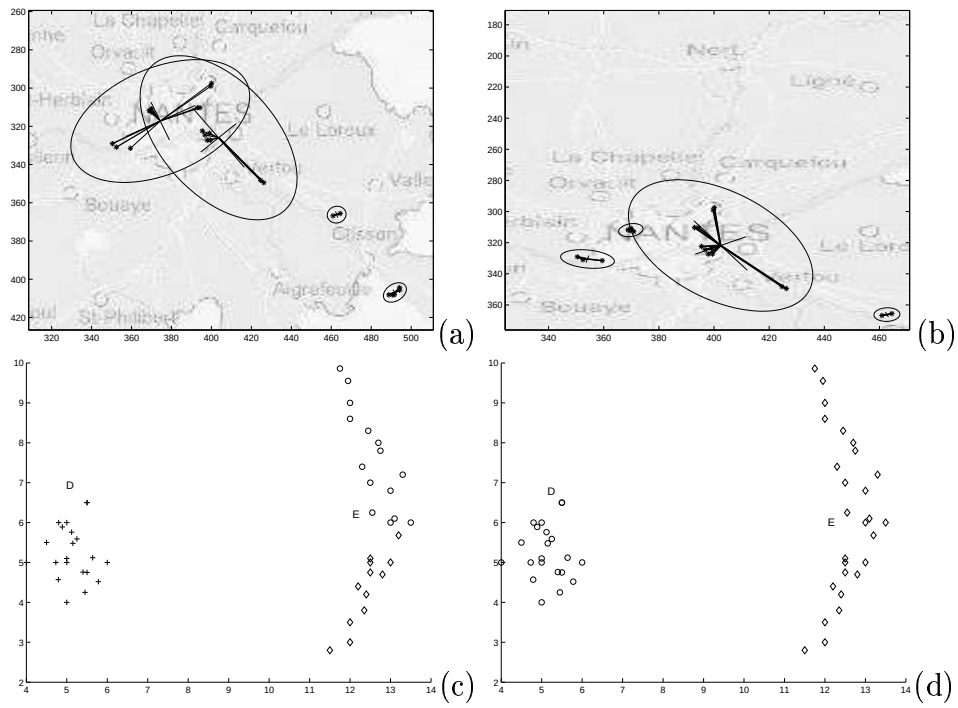

Fig. 2. Illustrating the comparison of BIC vs. ICL : cases (a) and (c) describe optimal solutions found using the BIC criterion, while cases (b) and (d) indicate the corresponding conclusions for the ICL criterion. In the case (a,b), where the number of clusters is forced to 4, BIC leads to interwoven clusters, which ICL avoids (we deem this latter solution better). In the case (c,d), where the number of clusters is free, BIC leads to undesirable splitting of an excessively non-Gaussian cluster (on (c), the cluster on the right is in fact composed of two clusters $\circ$ and $\diamond$) . The actual likelihood values are similar in both cases, it is the entropic term $\Phi(K)$ that makes the difference.

variance critical situations imply. Besides, this feature may integrate measurement noise on geolocation data.

- *a maximum ratio between the two covariance matrix eigenvalues.* A full covariance matrix is more flexible and desirable than a diagonal matrix, as the main inertia axis orientation is free. Still, with only two data elements (or more, if data elements are almost aligned), the two eigenvalues have very different magnitudes, and these data having excessive likelihood. The critical issue due to lack of regularization is illustrated by the example result provided in fig.3.

Let the sample covariance matrix S for a cluster be diagonalized as

$$S = U.\Lambda.U^{-1}, \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{6}$$

where $\lambda_1 > \lambda_2$ and U is the matrix composed of the eigenvectors of S.

The principles indicated above result in the following definition for the regularized covariance matrix :

$$\hat{\Sigma} = U.\begin{bmatrix} max\ (\beta', \lambda_1) & 0 \\ 0 & max\ (\beta', \lambda_2, \beta\lambda_1) \end{bmatrix}.U^{-1}, \tag{7}$$

where β', β are two parameters. β' depends on the measurement noise, while β may be set to 0.3.

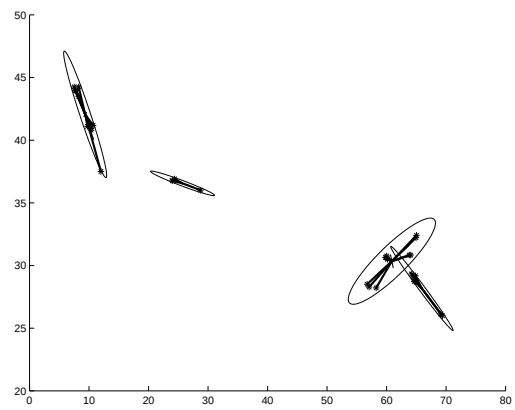

Fig. 3. Typical poor result obtained when applying full-covariance model-based clustering on small samples, without any regularization. Local configurations with strongly unequal eigenvalues are generally obtained, allowing excessive intra-cluster Euclidian distance.

Another issue related to the small sample case is that the BIC approximation theoretically sees its validity reducted. We conducted experiments with an alternative dedicated to small samples ($AICc$ [28]), but did not conclude to significant improvement.

In the next section, we present the technique for attempting optimization of this criterion.

3.3 Search for the optimal classification

3.3.1 The EM algorithm

The Expectation-Maximization (EM) algorithm [29,30] is employed here for (locally) optimizing the criterion described above. This algorithm is widely used in association with mixture modelling. It iterates between two steps. In the E-step, the unknown data-to-model assignements are replaced by their expectation, given the current model parameter estimates; this strategy breaks the combinatorial expense of the general clustering problem. Then, the M-step

updates the model parameters, given the current data-to-model assignement estimates.

We describe hereunder several problems the EM algorithm has to cope with, and some solutions proposed in our case.

As the EM technique assumes the number of components is known, we conduct the search for each hypotheses subspace associated to a given number of components. Among all the solutions found, one retains the solution which maximizes the ICL criterion. An improvement on this rather exhaustive search strategy is proposed in [31]. However, our technique is also motivated by the interest, in future work, in providing partitions at multiple granularities, i.e. finding several plausible partitions with various complexities, to enable multi-scale browsing.

The second issue is due to the local character of the optima found with the EM algorithm. Two complementary techniques are used to improve this point:

- an "em-EM strategy" [32] which is processed during the initialization step. This procedure seems more effective than of the commonly used k-means-based initialization and is described in the "Algorithm 1" below,
- a Split-Merge EM (SMEM) algorithm, inspired from [33], which is applied after local optimization by the em-EM algorithm. This technique consists in evaluating and (possibly) applying joint split & merges among current clusters, to attempt pertinent semi-local jumps in the search space, while keeping constant the number of clusters. Throughout, ICL remains the determining criterion. This SMEM algorithm is described in detail below.

Algorithm 1 "em-EM" strategy

for i=1 to NB_INITIALIZATIONS **do**

1. Make random data-to-model assignments. The set of random CEM initializations is constrained to form connected temporal components, thus generally starting closer from desirable solutions.
2. Run Classification EM (CEM is a hard-assignement version of EM, which convergence is fast)
3. Run a "short EM" (max. 20 iterations, weak convergence test)

end for

Run a "full EM" (up to 500 iterations, strict convergence test), initialized from the one among the above solutions that optimizes the ICL criterion.

3.3.2 Split and merge EM algorithm

The SMEM algorithm [33] aims at improving the result of a locally optimized model. By attempting several successions of division/fusion of components, it can help escape a local optimum. This algorithm is applied after the optimiza-

tion of the model with the em-EM searches. At each iteration of the SMEM algorithm, three components i, j and k are chosen, based on criteria described below. A tentative semi-local jump in search space is attempted by merging components i and j and splitting k, should the ICL criterion be improved. The overall model complexity remains unchanged.

Split and merge criteria

Because of the high number of split and merge possibilities, some criteria are used to rank the candidates. Both criteria proposed here differ with those described in [33].

The merge criterion proposed in [33] is based on the posterior probability of components: the idea is to compare the inter-component posterior probabilities and to merge the components accordingly. This technique is not pertinent with components associated with few observations: for example, if we compare several components that only have a single observation each, they are not deemed good candidates for a merge. In our context, this configuration is frequently encountered. Alternatively, we propose the use of the Mahalanobis distance to compare components:

$$J_{merge}(i, j, \Theta) = min\{dist_{Maha.}(\mu_i, \Sigma_i, \mu_j), dist_{Maha.}(\mu_j, \Sigma_j, \mu_i)\}, \tag{8}$$

where $dist_{Maha}(\mu_j, \Sigma_j, \mu_i) = (\mu_i - \mu_j\}^T \cdot \Sigma_j^{-1} \cdot (\mu_i - \mu_j)$.

Our split criterion is based on the entropy of each component. As a component with a high entropy suggests that the component does not fit well its associated data, or that another model also somewhat fits this data, we consider splitting the components according to this principle.

$$J_{split}(k, \Theta) = \sum_{i=1}^{n} t_{ik} \cdot log(t_{ik}) \tag{9}$$

Finally, the J_{merge} and J_{split} are sorted as follows: first, the merge candidates are sorted based on J_{merge}. Then, for each merge candidates $\{i, j\}$, the split candidates excluding $\{i, j\}$ are sorted as $\{k\}$. The result is then combined to obtain the list of candidates.

Initialization of the new parameters

Parameter initialization for the new models after a split and merge, given the current parameter Θ^*, are defined as follows. The initial parameters for the

merged components i, j are expressed by:

$$\alpha_{i'} = \alpha_i^* + \alpha_j^* \quad \text{and} \quad \theta_{i'} = \frac{\alpha_i^* \theta_i^* + \alpha_j^* \theta_j^*}{\alpha_i^* + \alpha_j^*} \tag{10}$$

For the division of a component k in two components j' and k':

$$\alpha_{j'} = \alpha_{k'} = \frac{\alpha_k^*}{2}, \quad \mu_{j'} = \mu_k^* + \epsilon, \quad \mu_{k'} = \mu_k^* + \epsilon' \quad and \tag{11}$$

$$\Sigma_{j'} = \Sigma_{k'} = det(\Sigma_k^*)^{(1/d)} / I_d, \tag{12}$$

where ϵ and ϵ' are some small pertubation vectors, $det(\Sigma)$ denotes the determinant of Σ and I_d the d-dimensional identity matrix.

After initialization, the SMEM algorithm uses a *partial* EM phase to update the parameters of the new models. This consists in re-estimating the parameters of modified components without affecting the others. Thus posterior probabilities of the new model m' are modified according to (13) during the expectation step of the EM algorithm:

$$P(m'|x, \Theta) = \frac{\alpha_{m'} p(x|\Theta_{m'})}{\sum_{l=i',j',k'} \alpha_l p(x|\alpha_l)} \cdot P(k|x; \Theta^*), \tag{13}$$

where $m' = i', j', k'$. The usual EM algorithm is then iterated until convergence.

SMEM algorithm

The SMEM algorithm consists in computing the list of candidates for the division and fusion of components and then, for the first candidate, to initialize the parameter and optimize the new model with the *partial* EM and classic EM algorithm. If ICL criterion is improved, the new model is retained. In the case it is rejected, a new candidate is tested. The SMEM algorithm is summarized by Algorithm 2 below, and is applied systematically after each em-EM run.

3.4 Comparison of spatial and temporal partitions

Let us recall that the classifications are carried out independently in space and time. Due to the lack of natural clusters in the data, to insufficiency of the optimality criterion or poor local minima, one or both of the partitions obtained would occasionnally poorly capture the data structure. We propose to

Algorithm 2 Split and merge algorithm

1. Perform the usual EM updates from some initial parameter value Θ until convergence. We note Θ^* and Q^* respectively the estimated parameters and the function value after the EM algorithm has converged.
2. Sort the split and merge candidates by computing split and merge criteria based on Θ^*. We note $\{i, j, k\}_c$ the cth candidate.

for $c = 1, \ldots, C_{max}$ **do**

After making the initial parameters settings based on Θ^*, perform the *partial* EM algorithm for $\{i, j, k\}_c$ and then the usual EM algorithm until convergence. Let Θ^{**} be the obtained parameters and Q^{**} the new function value after convergence. If $Q^{**} > Q^*$ then $Q^* \leftarrow Q^{**}$, $\Theta^* \leftarrow \Theta^{**}$ and return to 2.

end for

compare the values of entropy $\Phi(K)$ found for each of the two optimal classifications, and select the one with the smallest value, i.e better-defined clusters. This generally leads to selecting the most relevant and reliable partition for viewing and browsing the picture collection.

4 Experimental results

4.1 Trips in Loire-Atlantique

To illustrate the variety of situations that the proposed scheme can handle, the results of two first experiments are provided on fig.4. Only results on the geolocation data are reported here. Both scenarios are trips in Loire-Atlantique (Nantes district), corresponding to real touristic sights.

In the first case (fig.4a), there exist numerous clusters (in most locations, several pictures have been taken at the same or at nearby spots). Eleven clusters are obtained (including a strongly non-Gaussian one on the coast). In the second scenario, pictures are taken in a sparse manner, at two scales : on the whole district and more locally, around Nantes. The clustering result reveals this hierarchical structure, that is quite sensible with a browsing perspective.

4.2 Two holiday trips

These experiments correspond to two tourist trips, A and B, during which photos are taken at interesting sights. Figures 5 and 6 indicate the location of pictures.

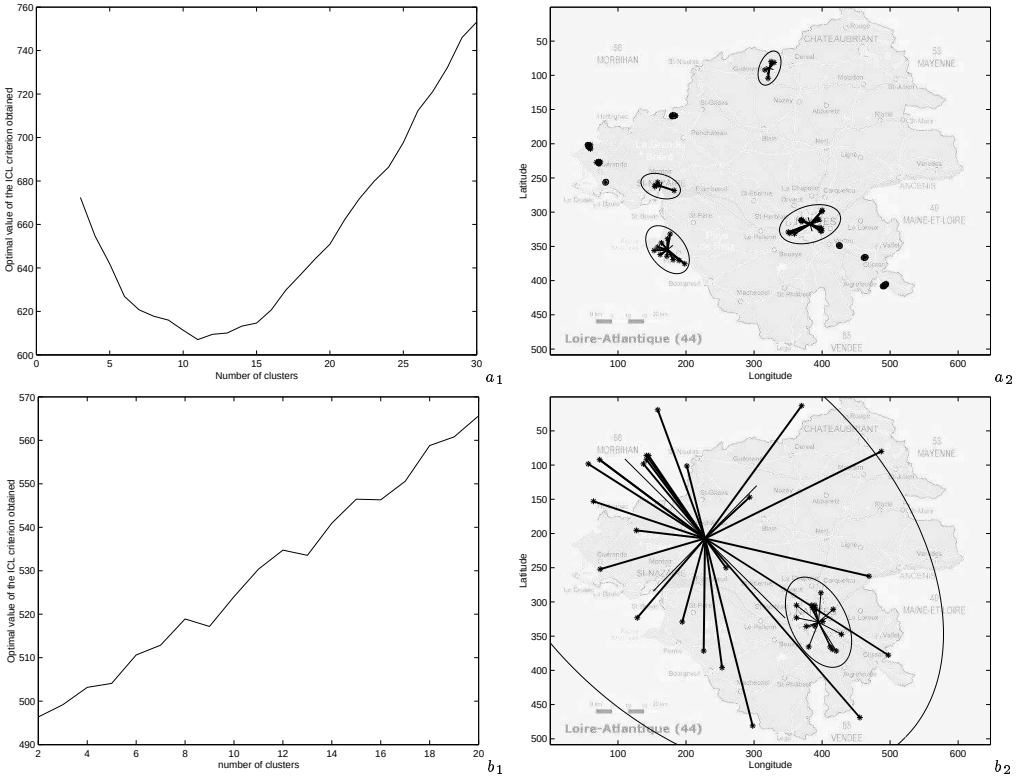

Fig. 4. Two experimental results showing extreme cases of geolocation picture distribution. In (a), the distribution is formed of numerous dense clusters, while in (b), a hierarchy of two sparse clusters is found. Ellipses represent the 2σ equiprobability line, and lines join data points to their cluster center. The optimal ICL value obtained for each candidate model complexity is resp. shown on a_1 and b_1.

During trip A, the user takes 300 pictures during ten days at different locations. Figure 7 and 8 represent the picture classification obtained, respectively for the time and the location. We test models between 1 and 15 components. The temporal classification retained is composed of 14 components and presents a acceptable result. Our constraint on the size on matrix covariance is effective, as single-data clusters are well handled. There remain the problem of data over-segmentation in the interval $[9000, 11000]$, which presents a bad separability: finding three components would be better. Still, all relevant time episodes are identified. The spatial classification has 12 components and the structure is overall identified, although north-western clusters are not well separated. Separabilities of the data in time and space are assessed as $\Phi_{time} = 12.46$ and $\Phi_{location} = 7.45$, and consequently the spatial classification is selected.

During trip B, the user takes 500 pictures within ten days. We test models between 1 and 15 components. 14 and 8 components are respectively found for the temporal classification (fig. 9) and the spatial one (fig. 10). The temporal classification obtained is satisfactory, as all relevant time episodes are recovered. The partition is formed of many little clusters, with good separabil-

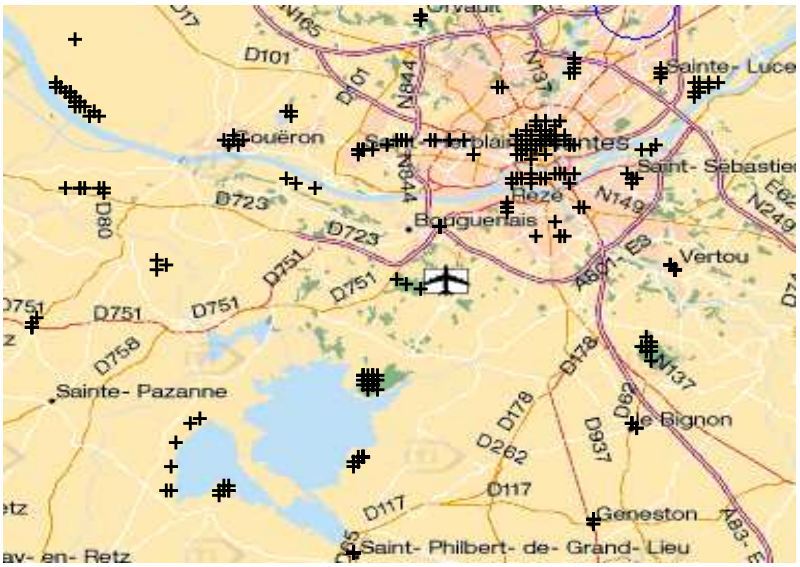

Fig. 5. Scenario A: each cross represents the location of a picture.

Fig. 6. Scenario B: each cross represents the location of a picture. This scenario involves trips along the river and around lakes.

ity. In the spatial classification, most pertinent locations are well determined, while the 3-cluster structure obtained for the trip around the lake (latitude $[0, 4000]$) is debatable, although it remains an exploitable result. This split is certainly due to the excessive non-Gaussianity of the cluster, as hinted by the 'likelihood elbow' for the 5-cluster hypothesis on fig 10. Yet, the separability of the data in time and space is assessed as $\Phi_{time} = 31.45$ and $\Phi_{location} = 34.12$, i.e. the relatively high value of $\Phi_{location}$ captures the abovementioned ambiguous situation and makes the scheme dismiss this partition, in favour of the temporal one. This suggests two extensions for the current scheme : (i) the sequential feature of geographical data should be taken into account further than it is currently, (ii) a time-space switching strategy could improve the

current binary time-or-space decision.

Although experiments were so far conducted on data sets corresponding to the scale of days rather then globally to weeks and months, they provide a first validation of the proposed scheme.

For the sake of comparison, experiments were repeated without the SMEM phase. The resulting classifications were generally visually similar to the previous ones, but the ICL criterion was most of the time improved when the SMEM algorithm was used. Overall, this suggest that, to obtain a clustering of similar quality, SMEM helps reduce significantly the number of random initializations needed of the em-EM scheme.

As a final remark, let us point out that we did not examine experimentally the drawbacks of using the BIC approximation of the evidence, over more refined options such as Laplace's approximation, and hence cannot yet identify precisely this share of the responsability in occasionnal classification mistakes.

5 Application to image collection browsing

In this section, we illustrate the practical interest of the proposed scheme, from an image collection browsing perspective, as applied to a mobile phone/PDA.

We consider here time-oriented views, such as would be found on familiar calendar managers available on PDAs. The proposed divisions of time for browsing may be either the temporal or spatial partition, depending on the abovementioned criterion. In the case space is selected, temporally disconnected components are separated. Restriction to one of the partitions ensures a small number of episodes and consistency of the division criterion along the view. The choice of suitable representative images for image groups is out of the scope of this paper, and some solutions are proposed in works referred to in section 2.

Alternatively, one may exploit both partitions simultaneously. Fig. 11 illustrates a possible calendar-type view. Different background colors or separation marks between episodes indicate from what perspective the cluster is homogeneous. Input keys can be associated to the following functions " jump to next temporal episode","jump to the next occurrence of photos in this spatial zone".

6 Conclusion

In this paper, the problem of organizing personal digital image collections has been addressed. The interest towards temporal and spatial meta-data was stressed, especially for camera phone applications. To focus on purely data-driven structuring, we formulated the problem as an unsupervised classification issue. The ICL criterion was proposed, as it fulfills several requirements of the application (unknown number of clusters, poor Gaussianity). The scheme is further improved by a technique for estimating covariance matrices in the case of clusters containing only a few pictures. The ICL criterion is optimized with an efficient version of the EM technique and a split/merge procedure SMEM for semi-local optimisation. Finally, we proposed a criterion for assessing the two partitions obtained and selecting the most reliable one. Among the qualities of the chosen formalism are the lack of delicate parameter tuning, and perspectives for multi-scale structuring. Overall, we believe it provides an effective and realistic direction towards organizing an image collection, providing a structure that may be mapped onto a variety of navigation schemes.

In current work, we are examining lowering the cost of the optimization strategy, by reducing redundant optimization work when exploring various model complexity hypotheses. Among other perspectives, better joint spatio-temporal structuring on one side, and turning to an incremental system on the other side, are essential. Although the EM algorithm adapts quite well to this case [34], more global reorganization of the collection sometimes has to be carried out.

References

[1] P. Gros, R. Fablet, P. Bouthemy, New descriptors for image and video indexing, Kluwer, 2001, Ch. in State-of-the-Art in Content-Based Image and Video Retrieval, H. Burkhardt, H.-P. Kriegel, R. Veltkamp (eds), pp. 213–234.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Trans. on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380.

[3] P. Anandan, Personal digital media: It's about sharing experiences, in: MMCBIR 2001 - Multimedia Content-based Indexing and Retrieval, INRIA Rocquencourt, France, 2001.

[4] J. C. Platt, M. Czerwinski, B. A. Field, PhotoTOC: Automatic clustering for browsing personal photographs, Tech. Rep. MSR-TR-2002-17, Microsoft Research (Feb. 2002).

[5] T. J. Mills, D. Pye, D. Sinclair, K. R. Wood, Shoebox: A digital photo management system, Tech. Rep. MSR 2000.10, AT&T Labs., Cambridge, England (2000).

[6] Y. Neuvo, J. Yrjanainen, Wireless meets multimedia, Wireless communications and mobile computing 2 (2002) 553–562.

[7] U. Gargi, Y. Deng, D. R. Tretter, Managing and searching personal photo collections, Tech. Rep. HPL-2002-67, HP Laboratories, Palo Alto (Mar. 2002).

[8] A. Loui, A. E. Savakis, Automatic image event segmentation and quality screening for albuming applications, in: IEEE Proceedings Int. Conf. on Multimedia and Expo (ICME'2000), New York, USA, 2000, pp. 1125–1128.

[9] A. Graham, H. Garcia-Molina, A. Paepcke, T. Winograd, Time as essence for photo browsing through personal digital libraries, in: ACM Joint Conference on Digital Libraries JCDL, 2002, pp. 326–335.

[10] H. Kang, B. Shneiderman, Visualization methods for personal photo collections: Browsing and searching in the photofinder, in: IEEE Int. Conf. on Multimedia and Expo ICME'2000 (III), New York, USA, 2000, pp. 1539–1542.

[11] P. Mulhem, J. Lim, W. Leow, M. Kankanhalli, Advances in digital home image albums, Ideal publishing, 2003, Ch. in Multimedia systems and content-based image retrieval.

[12] A. Soronen, V. Tuomisto, Mobile image messaging - anticipating the outlines of the usage culture, in: 4th Int. Symp. on Mobile HCI'2002, Pisa, Italy, 2002, pp. 359–363.

[13] K. Rodden, How do people manage their digital photographs?, in: ACM Conference on Human Factors in Computing Systems, Fort Lauderdale, 2003, pp. 409 – 416.

[14] A. E. Savakis, S. C. Etz, A. Loui, Evaluation of image appeal in photographic images, in: SPIE Conf. on Human Vision and Electronic Imaging V, 2000.

[15] C. Chen, N. Lesh, F. Vernier, Personal Digital Historian : story sharing around the table, ACM interactions 10 (2) (2003) 15–22.

[16] M. Gelgon, Using face detection for browsing personal slow video in a small terminal and worn camera context, in: IEEE Int. Conf. on Image Processing (ICIP'2001), Thessaloniki, Greece, 2001, pp. 1062–1065.

[17] M. Gelgon, K. Tilhou, Automated multimedia diaries of mobile device users needs summarization, in: 4th Int. Symp. on Human Computer Interaction with Mobile Devices (Mobile CHI'2002), Lecture Notes in Computer Science, Pisa, Italy, 2002, pp. 36–44.

[18] A. Myka, J. Yrjänäinen, M. Gelgon, Enhanced storing of personal content, Finnish Patent PCT/FI02/00277, Nokia Research Center, Nokia corp., Finland (Apr. 2002).

[19] J. Smith, P. Van Beek, T. Ebrahimi, T. Suzuki, J. Askelof, Metadata-driven multimedia access, IEEE Signal Processsing magazine, special issue on Universal Multimedia Access 20 (2) (2003) 40–52.

[20] Y. Zhao, Standardization of mobile phone positioning for 3G systems, IEEE Communications Magazine (2002) 108–116.

[21] N. Marmasse, C. Schmandt, Location-aware information delivery with commotion, in: Handheld and Ubiquitous Computing HUC'2000, Second Int. Symp., Bristol, UK, 2000, pp. 157–171.

[22] M. Gelgon, K. Tilhou., Structuring the personal multimedia collection of a mobile device user based on geolocation, in: IEEE Int. conf. on Multimedia and Expo (ICME'2002), Lausanne, Switzerland, 2002, pp. 248–252.

[23] D. Ashbrook, T. Starner, Learning significant locations and predicting user movement with GPS, in: IEEE Int. Symp. on Wearable Computing (ISWC'2002), Seattle, USA, 2002, pp. 101–108.

[24] C. Fraley, A. E. Raftery, How many clusters? Which clustering method? Answers via model-based cluster analysis, The Computer Journal 41 (8) (1998) 578–588.

[25] C. Biernacki, G. Celeux, G. Govaert, Assessing a mixture model for clustering with the integrated classification likelihood, in: IEEE Transaction on pattern analysis and machine intelligence, Vol. 22, 2000, pp. 719–725.

[26] D. Chickering, D. Heckerman, Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables, Tech. Rep. MSR-TR-96-08, Microsoft Research (Mar. 1996).

[27] S. Tadjudin, D. Landgrebe, Covariance estimation with limited training samples, IEEE Trans. on Geoscience and Remote sensing 37 (4) (1999) 134–149.

[28] N. Sugiura, Further analysis of the data by Akaike's information criterion and the finite correction, Communications in statistics, theory and methods A7 (1978) 13–26.

[29] C. M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.

[30] A. P. Dempster, N. M. Laid, D. B. Rubin, Maximum likelihood for incomplete data via the EM algorithm, J. R. Stat. Soc. (1977) 1–38.

[31] M. Figueiredo, A. K. Jain, Unsupervised learning of finite mixtures, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 24 (3) (2002) 381–396.

[32] C. Biernacki, G. Celeux, G. Govaert, Choosing starting values for the em algorithm for getting the highest likelihood in multivariate gaussian mixture models, Computational Statistics and Data Analysis (2003) 561–575.

[33] N. Ueda, R. Nakano, Z. Gharhamani, G. Hinton, SMEM algorithm for mixture models, Neural computation 12 (9) (2000) 2109–2128.

[34] R. Neal, G. Hinton, Learning in graphical models, Dordrecht : Kluwer academic publishers, 1998, Ch. A view of the EM algorithm that justifies incremental, sparse and other variants, pp. 355–368.

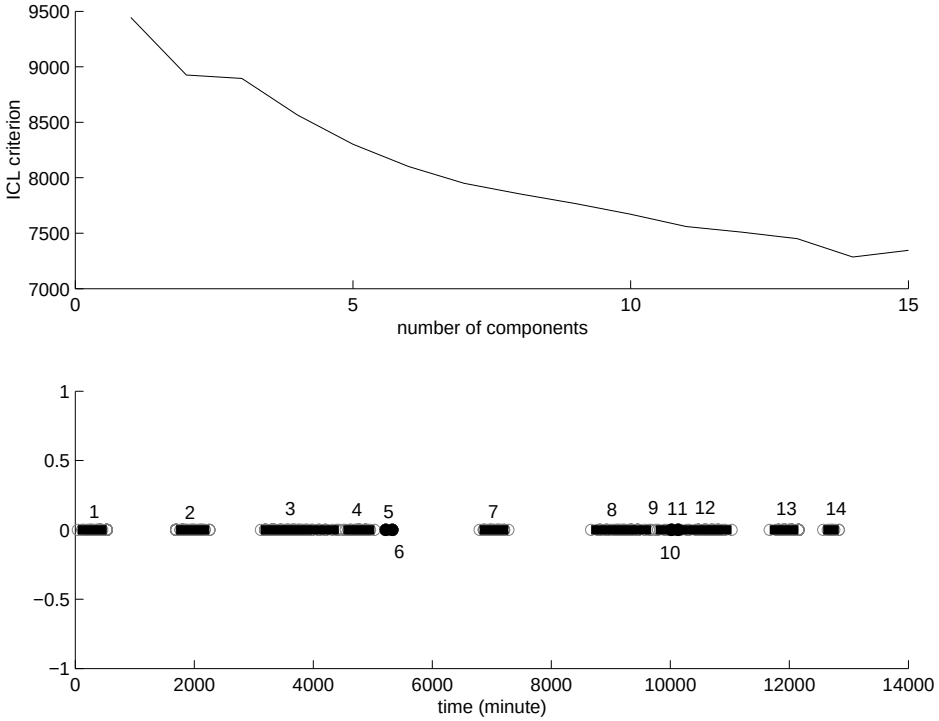

Fig. 7. Temporal classification of trip A: on the top, optimal ICL criterion obtained for each tentative model complexity. On the bottom, the obtained classification in the best configuration among all model complexities. Black lines represent the classes and each component is numbered.

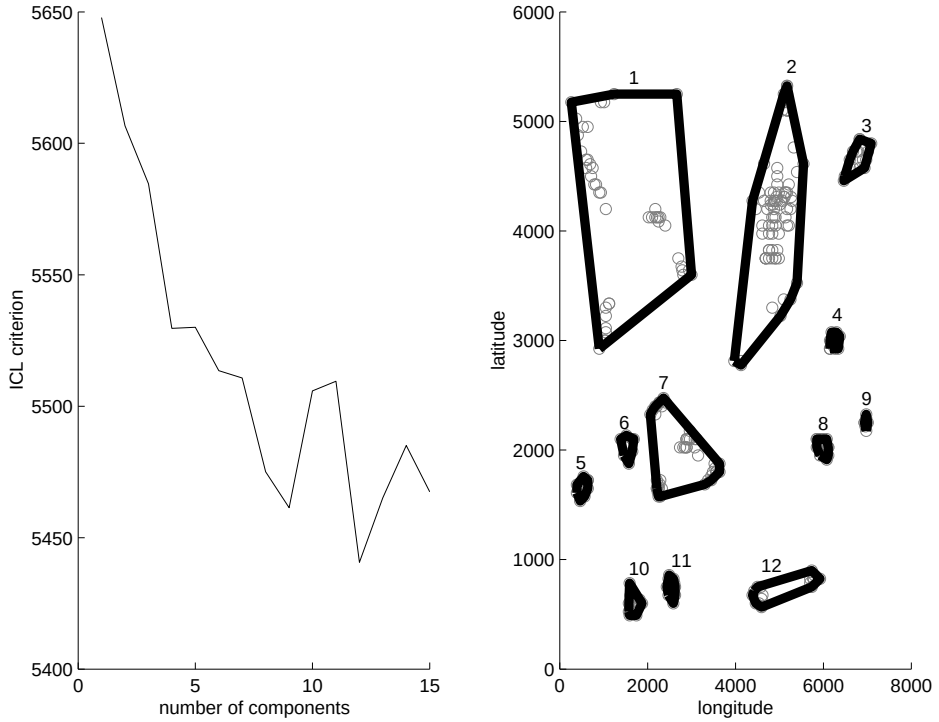

Fig. 8. Spatial classification of trip A: on the left, optimal ICL criterion obtained for each tentative model complexity. On the right, the obtained classification. Classes are numbered and their convex hulls are drawn.

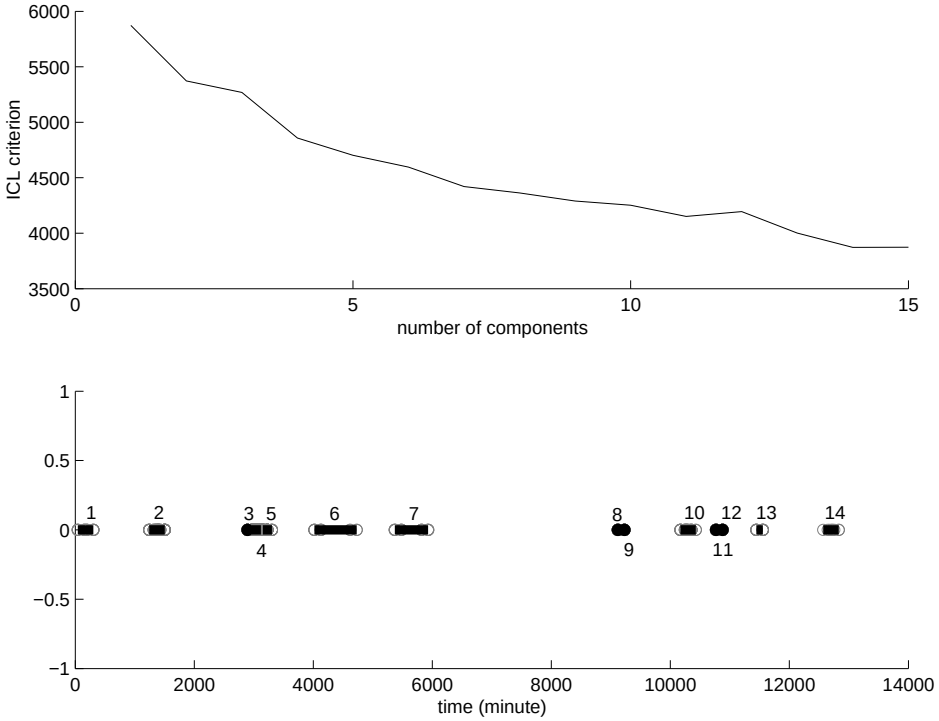

Fig. 9. Temporal classification of trip B: on the top, optimal ICL criterion obtained for each tentative model complexity. On the bottom, the obtained classification in the best configuration among all model complexities. Black lines represent the classes and each component is numbered.

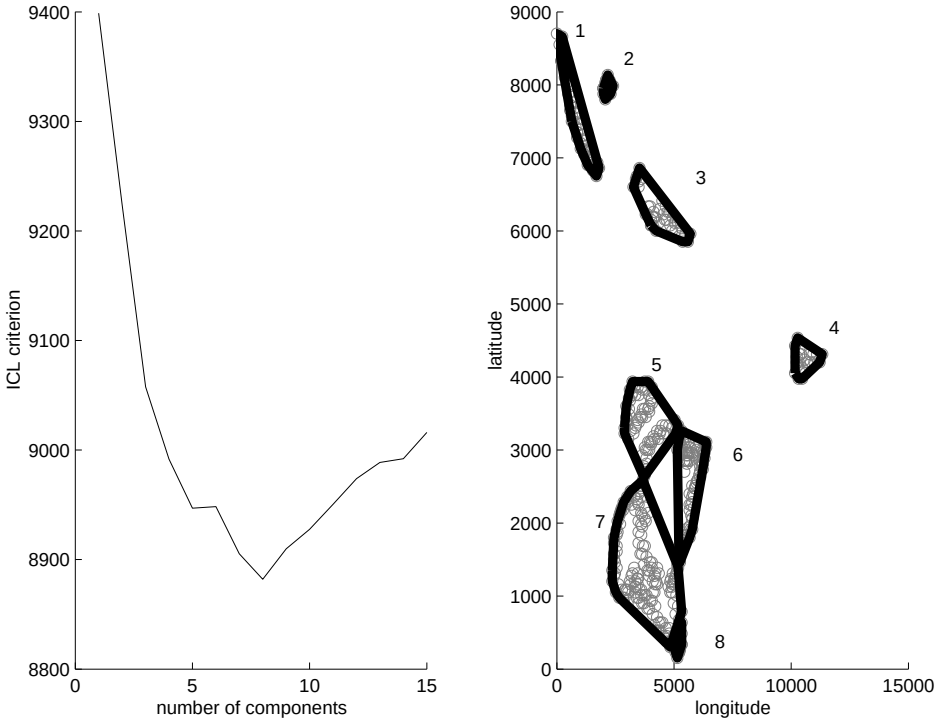

Fig. 10. Spatial classification of trip B: on the left, optimal ICL criterion obtained for each tentative model complexity. On the right, the obtained classification. Classes are numbered and their convex hulls are drawn.

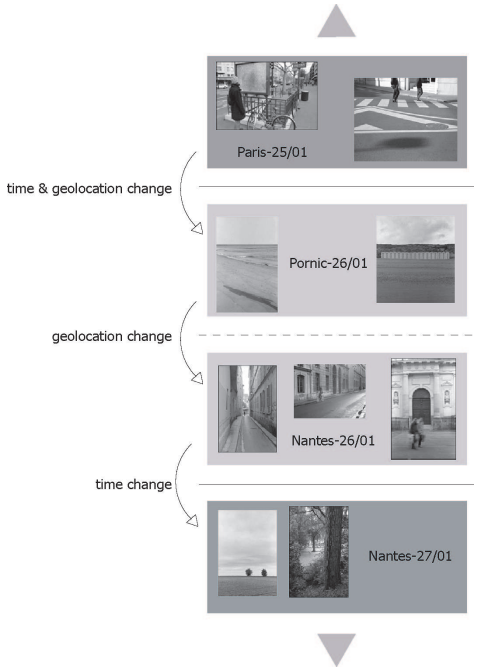

Fig. 11. Example of electronic calendar: dashed lines represent a change of geolocation, and continuous lines a time change or both. Assignement of names to clusters is carried out manually and does not relate to the proposed technique.