



# Incremental statistical geo-temporal structuring of a personal camera phone image collection

Antoine Pigeau, Marc Gelgon

## ► To cite this version:

Antoine Pigeau, Marc Gelgon. Incremental statistical geo-temporal structuring of a personal camera phone image collection. International Conference on Pattern Recognition, Aug 2004, United Kingdom. pp.878-881. hal-00415976

**HAL Id: hal-00415976**

**<https://hal.science/hal-00415976>**

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.

# Incremental statistical geo-temporal structuring of a personal camera phone image collection

A. Pigeau, M. Gelgon

LINA (FRE 2927 CNRS) / INRIA ATLAS group, Nantes university  
2, rue de la Houssinière - BP 92208, 44322 Nantes cedex 03 - France  
email:{name}@lina.univ-nantes.fr

## Abstract

*This paper makes a proposal for automatically organizing the personal image collection that would be collected from a mobile phone equipped with a digital camera. Doing so, it attempts to address emerging needs from this rapidly developing device family. Having sketched user needs, we underline the interest of temporal and spatial meta-data in the chosen context. Collection organization is then formulated as an unsupervised classification problem, in both space and time. A criterion and an estimation procedure are proposed, based on the statistical integrated completed likelihood criterion, providing effective solutions to model complexity determination, non-Gaussianity of clusters, and incrementality. For further summarization, a technique for fusing the temporal and geolocation-based partitions is finally put forward, to ease browsing of the image collection along a single dimension.*

## 1 Introduction

Content-based image retrieval problems have been dealt with for the many years now but, as stressed in [1], extending the scope of this field to address personal image collections is an important stake. Recently, proposals have emerged from most leading industrials (Microsoft MyLifeBits [7], Nokia LifeBlog, ...) and academics in the field of multimedia (e.g. [5, 9]).

The present paper examines, within personal image collections, the following niche and its particularities. *Camera phones* are always carried by their user, and are hence advantageous both as an image acquisition device and as an image retrieval terminal. Indeed, despite the technical issues image retrieval sets on such devices, their permanent availability and the ability to easily share retrieved pictures (MMS) makes this context propitious for building large collections. A further quality of camera phones resides in their excellent geolocation potential, which founds our proposal. For

instance, they can switch between, or fuse, measurements from an embedded GPS receiver with eg. GSM position estimation tools, such as E-OTD.

As images are gathered, they progressively build up a valuable memory of one's life, which can be later searched for many purposes. The system should offer users the possibility of comfortably retrieving a well-defined piece of information in their potentially large collection, and also functions for browsing to get an overall idea of the content of the collection. Providing such overviews is in fact also beneficial, let alone necessary, for retrieving a well-defined piece of information.

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 (progressive recollection is favoured by browsing if the collection is suitably organized, which is the goal of our proposal). 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, rather than image content-based criteria (except person identity), as demonstrated by Rodden's study on users [10].

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.

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. Finally, section 5 is devoted to concluding remarks.

## 2 Related work in time-based and geolocation-based structuring

Structuring an image collection according to the time stamp of each picture is intuitively appealing, practically quite cheap and reliable. The generative process of pictures (i.e. behaviour of users) is likely to exhibit time clusters and, furthermore, often in a hierarchical fashion. Change detection techniques, as employed in [9], possess the advantage of not setting a particular parametric model on the intra-cluster time distribution. A combination of this with clustering is proposed in [8], in which (preset size)-gap detection leads to initial groups for clustering. However, these proposals appear to come short, regarding usual issues such as number of clusters and 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 [9]. Finally, besides direct use of time for image grouping, it was recently proposed in [5] to combine time linearly with camera settings features and image content information, within an 'image similarity' measure.

To our knowledge, there are currently few systems that seem to have considered the geolocation-based structuring closely, at least in the multimedia retrieval setting. Also with a view to providing structured calendar-type views, but in setting and solution differing from the present paper, we proposed in [6] a technique towards unsupervised learning of meaningful locations. Similar work was simultaneously proposed in [2].

## 3 Spatio-temporal organization

### 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. Probabilistic model-based clustering with statistical estimation form a favourite framework for identifying meaningful groups in data [4].

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

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

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

The main features of the proposed scheme are as follows:

1. distinct classifications are built for time and space;
2. we resort to a statistical optimality criterion, that exhibits several good properties (implements Occam's razor, robust to non-Gaussianity of clusters) with regard to our goal.
3. optimization of this criterion is conducted using a specific Expectation-Maximisation (EM) technique. Classifications are built in an incremental manner (i.e., on-line with regarding to arrival of data) using a dedicated search procedure based on merges and splits of components, and more classical EM runs.
4. the two partitions obtained are finally combined into a single, hybrid partition, occasionally switching from one criterion to the other, depending of the presence of local structure in the data in space or time.

### 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 :

$$P(D|H_K) = \int P(D|\Theta_K, H_K)P(\Theta_K|H_K)d\Theta_K \quad (2)$$

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

The Bayesian Information Criterion (BIC) is an effective approximation of this expression, that enables its interpretation as the data likelihood, penalized by model complexity. In our case, we opt for a slightly differing alternative the Integrated Completed Likelihood (ICL) criterion [3]. Overall, the ICL criterion is in principle able to identify both model parameters and the number of clusters, while attempting to cope with the non-Gaussianity of some clusters, i.e. limit poor clustering due to the Gaussianity assumption, which often plagues model-based clustering. This feature is important in our case, as non-Gaussianity characterizes many clusters. This property is visible from the definition of the ICL criterion:

$$ICL = -ML + \frac{1}{2} \cdot N_K \cdot \log(n) - \Phi_K, \quad (3)$$

where  $ML$  is the maximized mixture loglikelihood,  $N_K$  is the number of independent parameters in the model with  $K$  components,  $n$  is the number of data elements and  $\Phi_K$  is an entropy-based criterion, defined by:

$$\Phi_K = - \sum_{k=1}^K \sum_{i=1}^n t_{ik} \cdot \log(t_{ik}), \quad (4)$$

where  $t_{ik}$  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.

### 3.3 Optimization of the proposed criterion

Overall, the proposed technique for optimizing the clustering criterion consists in suitable variations on the Expectation-Maximization (EM) algorithm. This local optimization framework is inherently well-suited to incremental clustering, since one may use a initialization/update mechanism as new data arrives. Further, this principle favours stability of the partitions over time, which certainly suits the wishes of the user that would appreciate some stability in the graph he navigates through.

The needed ability to update the number of clusters over time suggests enabling semi-local jumps in the search space, by allowing splits & merges among clusters. We proposed a technique adapted from the SMEM algorithm [11], but adapt their procedure to the case of an incremental problem (Ueda et al. keep the total number of clusters constant).

Given new data, the principle of the update is to test several splits followed by several merges to update the classification. Because of the high number of split and merge possibilities, the candidates are ranked, so as to attempt first splitting components with the highest entropy, as such cases suggest that the component does not fit well its associated data, or that another existing model also somewhat fits this data. The Mahalanobis distance is employed for tentative merging of clustering. Let us point out that, in any case, splits and merges are only carried out if they improve the ICL criterion, and their use is combined with more usual (local) improvements of the ICL, with the EM procedure. That is, the target cost function remains constant. The steps of the algorithm may be outlined as follows :

1. Add new data and optimize locally using EM;
2. Rank candidates for splitting according to their entropy, as computed from  $\Phi_K$ ; attempt and carry out splitting of clusters, as long as ICL is improved; run EM after each split;
3. Rank candidate pairs for merging according to their Mahalanobis distance; attempt and carry out merging, as long as ICL is improved; run EM locally after each merge.

### 3.4 Addressing the small sample problem

In our context, as the number of data elements in a cluster is often small, inaccuracies can occur in the estimation of the covariance matrices describing the intra-cluster spread. We regularize these cases by estimating a full covariance matrix, yet imposing two constraints:

- *setting minimum covariance eigenvalues*. This enables effective handling of single-data clusters in the classification, through the limitation of the likelihood increase such zero-variance critical situations imply;
- *a maximum ratio between the two covariance matrix eigenvalues* to avoid degenerated elongated covariance matrices.

These modifications are inserted within the EM algorithm. Although a Bayesian treatement would in principle be more appropriate, the above technique has a low cost and alleviates the issue.

### 3.5 Building a single, hybrid spatio-temporal partition

Let us recall that classifications are carried out independently in time and space. Some sections of the partitions obtained may not reflect in a relevant manner the structure present in the data, due to insufficiencies of the clustering criterion, poor local minima configurations, or simply because of lack of clear structure in *some part* of the data.

We propose to build a single partition out of the two obtained, trying to discard unreliable sections. As sketched in the figure below, this is formulated as the search for the best path along the time axis, possibly switching from one partition to the other at boundaries that coincide for both partitions. The principle is to obtain the hybrid partition with the strongest entropy, in the spirit of eq.(4). This is obtained at low cost, as the problem can be split into local optimizations.

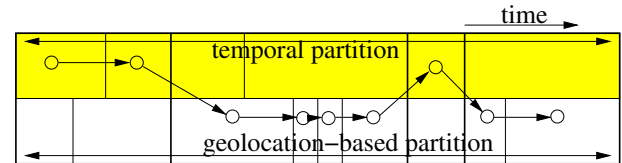
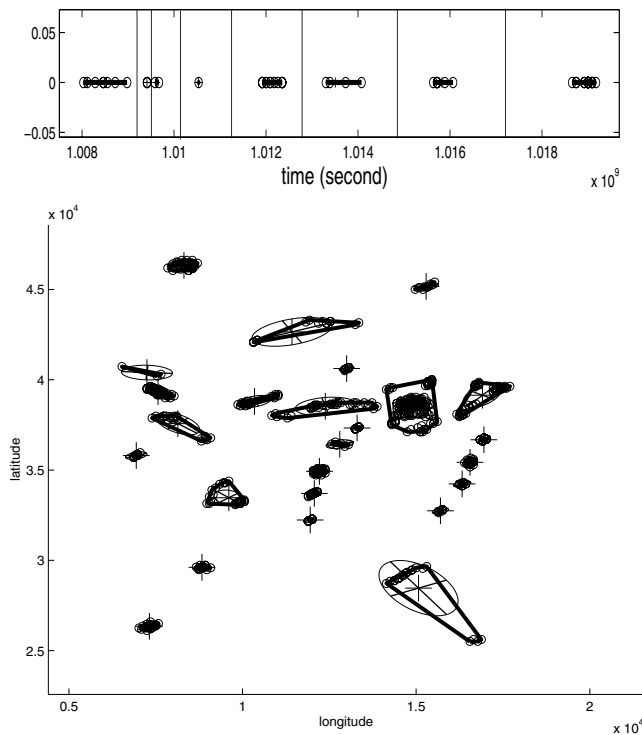


Figure 1. Example of hybrid partition

## 4 Experimental results

This experiment corresponds to a personal collection formed over several months and composed of 500 pictures. Time meta-data was included in each image (exif meta-data) and the location was added manually to each image based on the real location (longitude and latitude of each location was obtained from the <http://www.multimap.com> website). We built the temporal, spatial and hybrid classifications with the incremental algorithm described above and submitted the result to the user. The first two classifications are composed of 42 and 17 components respectively for the time and the



**Figure 2. Example of obtained classification: (top) a zoom on the temporal classification. Solid lines represent the limit of the components and 'o' the data. (bottom) A zoom on the spatial classification.**

location. Figure 2 shows zooms of the temporal (top) and the spatial classification (bottom). In both figures, we can noticed that the data is mostly correctly clustered.

The user was very satisfied of 95% of the components in both classifications. A few errors subsist when pictures are isolated (in time or in space): our technique tends to associate them to the nearest component instead of creating a new one ; the likelihood of isolated data is too low to counterbalance model complexity. Although this cannot be shown here, it was checked that the incremental determination of clusters remains satisfactory (providing a relevant and temporally stable partition ) as data is progressively made available. Besides, the ability of the proposed scheme to re-allocate data to clusters in a flexible way is seen to be beneficial.

From user feedback, the hybrid classification presents good properties and enables to improve the result obtained with the two separate classifications.

## 5 Conclusion

In this paper, we focus on the problem of organizing a personal digital image collection collected from a camera phone, with a view to image browsing and retrieval. Having stressed the interest of temporal and geolocation-based meta-data, we proposed a purely data-driven structuring technique for the collection, formulated as an incremental, unsupervised classification issue. Dedicated statistical criterion and optimization procedures are designed to this purposed, that avoid delicate parameter tuning. Overall, we believe the applicative need addressed is still rather novel and open, and the proposed technique provides a realistic (maybe overcomplex, but with good properties) direction towards organizing an image collection in a useful and effective manner.

## References

- [1] P. Anandan. Personal digital media: It's about sharing experiences. In *MMCBIR 2001 - Multimedia Content-based Indexing and Retrieval*, INRIA Rocquencourt, France, Sept. 2001.
- [2] D. Ashbrook and T. Starner. Learning significant locations and predicting user movement with GPS. In *IEEE Int. Symp. on Wearable Computing (ISWC'2002)*, Seattle, USA, pages 101–108, Oct. 2002.
- [3] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated classification likelihood. In *IEEE Transaction on pattern analysis and machine intelligence*, volume 22, pages 719–725, Jul. 2000.
- [4] C. Fraley and A. E. Raftery. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41(8):578–588, 1998.
- [5] U. Gargi, Y. Deng, and D. R. Tretter. Managing and searching personal photo collections. Technical Report HPL-2002-67, HP Laboratories, Palo Alto, Mar. 2002.
- [6] M. Gelgon and 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)*, pages 248–252, Lausanne, Switzerland, Aug. 2002.
- [7] J. Gemmel, R. Lueder, and G. Bell. Living with a lifetime store. In *ATR Workshop on ubiquitous experience media*, pages 69–76, Kyoto, Japan, Sept. 2003.
- [8] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In *ACM Joint Conference on Digital Libraries JCDL*, pages 326–335, Jun. 2002.
- [9] J. C. Platt and B. A. F. M. Czerwinski. PhotoTOC: Automatic clustering for browsing personal photographs. Technical Report MSR-TR-2002-17, Microsoft Research, Feb. 2002.
- [10] K. Rodden. How do people manage their digital photographs? In *ACM Conference on Human Factors in Computing Systems*, pages 409 – 416, Fort Lauderdale, Apr. 2003.
- [11] N. Ueda, R. Nakano, Z. Gharhamani, and G. Hinton. SMEM algorithm for mixture models. *Neural computation*, 12(9):2109–2128, 2000.