Geo-temporal Structuring of a Personal Image Database with Two-Level Variational-Bayes Mixture Estimation

Pierrick Bruneau[1,2], Antoine Pigeau[1],
Marc Gelgon[1,2], and Fabien Picarougne[1]

[1] Nantes university, LINA (UMR CNRS 6241), Polytech'Nantes
rue C.Pauc, La Chantrerie, 44306 Nantes cedex 3, France
[2] INRIA/IRISA Atlas project-team
firstname.surname@univ-nantes.fr

Abstract. This paper addresses unsupervised hierarchical classification of personal documents tagged with time and geolocation stamps. The target application is browsing among these documents. A first partition of the data is built, based on geo-temporal measurement. The events found are then grouped according to geolocation. This is carried out through fitting a two-level hierarchy of mixture models to the data. Both mixtures are estimated in a Bayesian setting, with a variational procedure: the classical VBEM algorithm is applied for the finer level, while a new variational-Bayes-EM algorithm is introduced to search for suitable groups of mixture components from the finer level. Experimental results are reported on artificial and real data.

1 Introduction

This paper addresses the goal of automated structuring a collection of geo-temporally annotated documents. A typical motivation arises from sets of images that can be captured and annotated automatically from mobile camera phones. Schemes for navigating efficiently into a shoebox of personal photographs have attracted growing attention in the past few years, as reviewed in the next section, using image content or meta-data provided by the camera, or a combination thereof. Our proposal does not consider the image content.

Geolocation and time are naturally essential criteria for this organization process [1], as they are both quite reliably measured and provide valuable navigation axes into one's memory and onto man-machine user interfaces. To address large amounts of data, we aim at determining data groups (i.e. clustering) into this geo-temporal space, in order to recover meaningful *events* and *places*. Let an *event* be a group of images that are close both in time and space, while a *place* is a group of data consistent from the sole geolocation criterion. The purpose of this task is to enable efficient browsing through time at a higher level of granularity than data, and hyperlinking co-located events that are distant in time.

M. Detyniecki, U. Leiner, and A. Nürnberger (Eds.): AMR 2008, LNCS 5811, pp. 127–139, 2010.

A main point of the paper is the way these two clustering operations are carried out and relate to one another. First, at the finer level, events are clusters directly identified from image meta-data in three-dimensional space (t, x, y). Then, at the coarser level, places are identified by forming groups of the previously found events, but considering only the location features (i.e. discarding time).

The technical framework for formalizing and solving this issue is that of Gaussian Mixture models (GMM) [2], as classically employed for numerous clustering tasks [3, 4, 5, 6, 7], including geo-temporal data [8]. For the problem at hand, where a hierarchy is sought with as little supervision as possible, we propose an original technique, based on the variational approach to Bayesian estimation at each level. The Variational Bayes framework has shown an efficient way of addressing mixture model estimation [9]: while Bayesian modelling enables regularization of estimates (avoiding degenerate situations for covariance matrices) and provides a principled manner of determining adequate model complexity (practically, number of clusters), the associated estimation algorithm often known as VBEM preserves the general form and the relatively low complexity of the Expectation-Maximization (EM) algorithm, but further handles the model complexity issue efficiently.

We propose to extend VBEM to handling a hierarchy of models: we describe in this paper how, also at the coarser level, mixture parameters for places may be inferred from a variational-Bayes EM algorithm applied to event-level components parameters. In other words, two slightly different variational-Bayes mixture estimation procedures are applied: the classical one, operating on punctual geo-temporal data, then the novel version, operating on component-level geolocation-only mixture parameters. The abovementionned advantages of VBEM algorithms are thus generalized to this coarser layer. Besides, because of the hierarchical constraint, the associations between events and places are explicit, in contrast with what would be obtained by determining both of them from the data. At the same time, thanks to the probabilistic framework, uncertainties on these associations are preserved, enabling various decision-making policies. Finally, because this coarse-level algorithm operates on mixture parameters rather than data, it only requires moderate computation cost.

Closely related work offering iterative component-grouping solution for building hierarchies on mixture models were proposed in [10] and [11]. However, they address a maximum likelihood estimate (or similarly, minimizing Kullback-Leibler loss) rather than a Bayesian setting, and hence leave open the issue of model complexity.

Let us mention that the nature of the scheme is also straightforwardly amenable to extensions: the data may be processed incrementally with an initialization/update mechanism in the probabilistic assignments at component-level, and events provided by multiple users may be handled, thus identifying common events, common places. We do not cover these perspectives herein.

From the user point of view, the main advantage of the obtained hierarchy is to facilitate the browsing task, a point emphasized on a mobile device. Indeed

such a device presents energy and interface constraints (for example a small screen and poor input keys) that raise the needs for adapted tools to browse the collection. Our hierarchical aspect improves this task since the coarse-level provides a summarization of the collection events. It enables then to decrease user interactions to finding a specific image and provides energy saving since each summary can be represented with a pertinent image subset.

The remainder of this paper is organized as follows. Section 2 reviews work pertaining to organization of personal images. We then disclose the proposal based on a hierarchy of mixture models (section 3). Experimental results are then provided in section 4, and conclusions and perspectives are drawn (section 5).

2 Related Work

Time stamp is the obvious criterion for ordering pictures, as well as for identifying groups of images close in time (i.e. assumed to have a common topic). Segmenting the sequence of time stamps has been viewed in [12,13] as the incremental detection of gaps, with the advantages that data may be processed as a flow, and each temporal unit is not assumed to be generated from a particular parametric law. However, there result a somewhat arbitrary definition of what is a significant "gap" in time.

Most works on personal image indexing now cope with image location. Practically, handy GPS systems are now largely popular to provide this information (we disregard herein all GPS measurement problems). Systems such as WWMX [14] or Flickr propose a map-based interface to browse the collection. The main problem of such approaches is that the map gets cluttered, when the number of images grow, especially on handheld devices. Alleviating this issues is indeed a main motivation for the work in the present paper. In this direction, building compact representations of the images set and easy navigation procedures has been proposed in [14], where images are aggregated in accordance with the map scale, while [15] selects relevant images from multi user collection based on their meta-data (see below).

Directly combination of the temporal and geographical meta-data, which is the focus of the present paper, was put forward in [16, 8], which also organize an image collection hierarchically, based on time and location clusters. A series of heuristic rules derived from user's expectations are implemented [16] to build a geo-temporal hierarchy of events. In [8], we proposed an incremental EM algorithm to carry out distinct temporal and spatial hierarchical classifications, significantly different from the present paper with regard to the relation between geolocation and time, the way the hierarchy is built, and the technique for conducting Bayesian estimation. Finally, time and geographical structuring can also be combined with image features [17], or with the camera settings [18]. Such criteria may indeed be of interest if applied on subset of images corresponding to an event.

Recent contributions extend these principles to the multi user context [15, 19, 20, 21]. Popular websites as Flicker or GoogleMap enable users to share

their personal collections, leading to potentially huge amounts of images. Experiments on user with the Zurfer system [19] show that favorite organization criteria differ from the single user context: users prefer to browse image sets according to social interactions (photos from friends and family members). Let us however notice the distinction between pictures authored by these relatives and personal pictures involving theses persons (such as present in the picture). This was for instance taken into account in [15], where the distance, in a social network, between the image authors and the query author, is taken into account to select a representative image. The importance of images and image cluster is assessed through a heuristic combination of textual tag originality (tf/idf), the diversity of image authors in the location of interest, and various other criteria. This work was extended in [19] to include image content. In [20], a temporal variable is studied, that reflects the density of pictures (from multiple users) on the temporal axis, enabling temporal determination of events. Interestingly, it compares to our proposal in the same way Parzen density estimation compares to mixture models density estimators. Finally, let us notice that, in such multi-user search scenarios, image browsing is generally restricted to a subset of the complete collective collection (generally a specific location).

3 Determinating Events and Places with Hierarchical Variational-Bayes

3.1 Clustering into Events with VBEM on Punctual Geo-temporal Data

In this section, we aim at clustering images based on their metadata in three-dimensional (x,y,t) space. Briefly stated, we model the data as sampled from a Gaussian mixture, and carry out a Bayesian estimation of model parameters by means of a variational approximation.

A Gaussian mixture is defined by the following probability distribution function (pdf):

$$p(x) = \sum_{k=1}^{K} \omega_k \mathcal{N}(x \mid \mu_k, \Lambda_k^{-1}) \tag{1}$$

where x is a d-dimensional feature vector and $\mathcal{N}(. \mid \mu_k, \Lambda_k^{-1})$ is a Gaussian pdf with mean vector μ_k and precision matrix Λ_k. In the remainder of this paper, we will designate $\mathcal{N}(. \mid \mu_k, \Lambda_k^{-1})$ as the k-th component of the GM. $\Omega = \{\omega_k\}$ is a weight vector associated to the components, following the constraint $\omega_k \geq 0 \; \forall k, \sum \omega_k = 1$. We introduce a lightweight notation for the GMM parameters: $\theta = \{\Omega, \mu, \Lambda\}$ where $\mu = \{\mu_k\}$ and $\Lambda = \{\Lambda_k\}$.

In our case, where K is unknown, Maximum Likelihood Estimation (MLE) is not applicable to determine jointly K with model parameters. Although this issue may be overcome using penalized likelihood criteria to compare models of various complexities (AIC [22], BIC [23]), this requires computation of each model separately before comparison. Instead of treating model parameters as

unknown scalars from which we seek MLE, we can define pdfs over these parameters, leading to a fully Bayesian approach. In this context, we can define a variational distribution that, in its optimal setting, will approximate the true posterior distribution in the KL sense [9, 2]. Furthermore, using exponential distributions as Gaussians or multinomials allows to define prior pdfs over parameters. Therefore, by choosing an appropriate prior, estimating a correct K will now be part of the global estimation process. Indeed, under this condition the estimation will typically lead to a model with some insignificant component weights that can be pruned ([2]). Hence, by choosing an initial K sufficiently large, we will automatically obtain an effective K (let us denote it K'). The restriction of variational distributions to factorized versions allows to decline coupled update equations over latent variables and parameters. As there is no closed-form solution for this equations system, we can iteratively find a local MAP (Maximum A Posteriori) pdf by using an EM-based algorithm [9,2], often known as VBEM. Besides handling model complexity, setting a spherical prior of covariance matrices helps avoid poor estimates, that classically plague clusters that are responsible for little data.

3.2 Grouping Events into Places with a Component-Level VBEM Algorithm

Once events are identified, we attempt to find groups of events, based on the mere parameters of the mixture that describe this set of events. We now describe the technique employed to this aim.

We follow notations used [2] for punctual data. Classically, mixture variational estimation considers a set of data $X = \begin{pmatrix} x_1^T \\ \dots \\ x_N^T \end{pmatrix}$ and $Z = \begin{pmatrix} z_1^T \\ \dots \\ z_N^T \end{pmatrix}$ that is assumed to be generated from the mixture. x_i is a d-dimensional feature vector and z_i the associated binary variable indicating from which component x_i was generated (e.g. from k-th component $\equiv z_{ik} = 1, z_{ij} = 0\ \forall j \neq k$). In the clustering context, Z is hidden, and the purpose of the procedure is to compute a joint estimate of θ and Z. The associated pdfs are:

$$p(Z \mid \Omega) = \prod_{n=1}^{N} \prod_{k=1}^{K} {\omega_k}^{z_{nk}} \tag{2}$$

$$p(X \mid Z, \mu, \Lambda) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{N}(x_n, \mu_k, \Lambda_k^{-1})^{z_{nk}} \tag{3}$$

Now consider an arbitrary mixture defining L components, with parameters $\theta' = \{\Omega', \mu', \Lambda'\}$. Typically this model might have redundant components, or might be a group of mixtures. We then assume that X and Z were i.i.d sampled from this distribution. It is therefore possible to regroup X by the component that originated its various items. It leads us to the following formalism: $X = \{\hat{x}_1, \dots, \hat{x}_L\}$ with $\text{card}(X) = N, \hat{x}_l = \{x_i \mid z_{il} = 1\}$ and $\text{card}(\hat{x}_l) = \omega_l' N$. We are

now going to express the distributions (2) and (3) w.r.t this formalism. To achieve tractability, we make the following assumption: $\forall x_i \in \hat{x}_l, z_{ik} = \text{const} = z_{lk}$. Thus we can rewrite the expression (3):

$$p(X \mid Z, \mu, \Lambda) = \prod_{k=1}^{K} \prod_{l=1}^{L} p(\hat{x}_l \mid Z, \mu_k, \Lambda_k)^{z_{lk}} \tag{4}$$

$$p(X \mid Z, \mu, \Lambda) = \prod_{k=1}^{K} \prod_{l=1}^{L} \left[\prod_{i=1}^{\omega_l' N} \mathcal{N}(x_{li} \mid \mu_k, \Lambda_k^{-1}) \right]^{z_{lk}} \tag{5}$$

$$\ln p(X \mid Z, \mu, \Lambda) = \sum_{k=1}^{K} \sum_{l=1}^{L} z_{lk} \left[\sum_{i=1}^{\omega_l' N} \ln \mathcal{N}(x_{li} \mid \mu_k, \Lambda_k^{-1}) \right] \tag{6}$$

For N sufficiently large, we can make the following approximation:

$$\sum_{i=1}^{\omega_l' N} \ln \mathcal{N}(x_{li} \mid \mu_k, \Lambda_k^{-1}) \simeq \omega_l' N E_{\mu_l', \Lambda_l'} \left[\ln \mathcal{N}(x \mid \mu_k, \Lambda_k^{-1}) \right] \tag{7}$$

This statement is known as *virtual sampling*, and was introduced in [24, 10] in the context of max. likelihood estimation.

We can also write:

$$E_{\mu_l', \Lambda_l'} \left[\ln \mathcal{N}(x \mid \mu_k, \Lambda_k^{-1}) \right] = \int \mathcal{N}(x \mid \mu_l', \Lambda_l^{'-1}) \ln \mathcal{N}(x \mid \mu_k, \Lambda_k^{-1}) \, \mathrm{dx} \tag{8}$$

$$\begin{aligned} E_{\mu_l', \Lambda_l'} \left[\ln \mathcal{N}(x \mid \mu_k, \Lambda_k^{-1}) \right] = -KL \left(\mathcal{N}(x \mid \mu_l', \Lambda_l^{'-1}) \parallel \mathcal{N}(x \mid \mu_k, \Lambda_k^{-1}) \right) \\ - H(\mathcal{N}(x \mid \mu_l', \Lambda_l^{'-1})) \end{aligned} \tag{9}$$

with $KL(q_0 \parallel q_1)$ the KL divergence of q_1 from q_0 and $H(q_0)$ the entropy of q_0. These two terms have closed-form expressions [25]. Thus by reinjecting (9) into (7), and then (7) into (6), we obtain the following expression:

$$\begin{aligned} &\ln p(X \mid Z, \mu, \Lambda) = N \sum_{k=1}^{K} \sum_{l=1}^{L} z_{lk} \omega_l' \\ &\left[-KL \left(\mathcal{N}(x \mid \mu_l', \Lambda_l^{'-1}) \parallel \mathcal{N}(x \mid \mu_k, \Lambda_k^{-1}) \right) - H(\mathcal{N}(x \mid \mu_l', \Lambda_l^{'-1})) \right] \end{aligned} \tag{10}$$

$$\begin{aligned} &\ln p(X \mid Z, \mu, \Lambda) = N \sum_{k=1}^{K} \sum_{l=1}^{L} z_{lk} \omega_l' \\ &\left[\frac{1}{2} \ln \det \Lambda_k - \frac{1}{2} \mathrm{Tr}(\Lambda_k \Lambda_l^{'-1}) - \frac{1}{2} (\mu_l' - \mu_k)^T \Lambda_k (\mu_l' - \mu_k) - \frac{d}{2} \ln(2\pi) \right] \end{aligned} \tag{11}$$

The formalism change we made also has consequences on (2): as we previously stated that $z_{lk} = z_{nk} \; \forall x_n \in \hat{x}_l$, we can write:

$$p(Z \mid \Omega) = \prod_{n=1}^{N} \prod_{k=1}^{K} \omega_k^{z_{nk}} = \prod_{l=1}^{L} \prod_{k=1}^{K} \omega_k^{N \omega_l' z_{lk}} \tag{12}$$

Variational update equations are partially based on moments evaluated w.r.t $p(Z)$ and $p(X)$. Therefore we can review the cascading consequences relatively to the method introduced in [2].

$q(Z)$ expression is based on the computation of the $\ln(\rho_{nk})$ terms. As a consequence of (11) and (12), it now reduces to the computation of:

$$\begin{aligned}\ln(\rho_{lk}) = \frac{N\omega_l'}{2}\left(2E[\ln \omega_k] + E[\ln \det \Lambda_k] - d\ln(2\pi)\right)\\ -\frac{N\omega_l'}{2}\left(E_{\mu_k,\Lambda_k}\left[\mathrm{Tr}(\Lambda_k \Lambda_l^{'-1}) + (\mu_l' - \mu_k)^T \Lambda_k (\mu_l' - \mu_k)\right]\right)\end{aligned} \tag{13}$$

This leads to the computation of the set $\{r_{lk}\}$ and defines $q(Z)$ in its optimal setting.

The moment w.r.t μ_k and Λ_k is easily evaluated to give

$$\frac{d}{\beta}_k + \nu_k\left[\mathrm{Tr}(W_k \Lambda_l^{'-1}) + (\mu_l' - m_k)^T W_k (\mu_l' - m_k)\right].$$

The pdfs $q(\Omega)$ and $q(\mu, \Lambda)$ are also modified. In their optimal setting, the update equations become:

$$\alpha_k = \alpha_0 + \sum_l N\omega_l' r_{lk} \tag{14}$$

$$\beta_k = \beta_0 + \sum_l N\beta_l' r_{lk} \tag{15}$$

$$m_k = \frac{1}{\beta_k}\left(\beta_0 m_0 + \sum_l N\omega_l' r_{lk}\mu_l'\right) \tag{16}$$

$$W_k^{-1} = W_0^{-1} + \beta_0 m_0 m_0^T - \beta_k m_k m_k^T + \sum_l N\omega_l' r_{lk}(\mu_l'\mu_l^{'T} + \Lambda_l^{'-1}) \tag{17}$$

$$\nu_k = \nu_0 + \sum_l N\omega_l' r_{lk} \tag{18}$$

Cycling through these update equations implements an EM-based algorithm, analogously as presented in [2, 9].

Also in [2], we learn that the previously described algorithm monotonically decreases the KL distance between the variational pdf and the true posterior. This is equivalent to maximising the lower bound of the complete likelihood. As we can compute this lower bound, and as this bound should never decrease, we can test for convergence by comparing two successive values of the bound. Only terms of the bound that depend on Z or X are impacted, we list these changes below:

$$\begin{aligned}E[\ln p(X \mid Z, \mu, \Lambda)] = \frac{1}{2}\sum_k \sum_l N\omega_l' r_{lk}\\ \left[\ln \tilde{\Lambda}_k - \frac{d}{\beta_k} - \nu_k[\mathrm{Tr}(W_k \Lambda_l^{'-1}) + (\mu_l' - m_k)^T W_k (\mu_l' - m_k)]\right]\end{aligned} \tag{19}$$

$$E[\ln p(Z \mid \Omega)] = \sum_l \sum_k N\omega_l' r_{lk} \ln \tilde{\omega}_k \tag{20}$$

Choosing appropriate priors can be problematic. If we have expert knowledge on the clustering structure to build, this can be used to build it, but we must be able to carry out the problem in a totally unsupervised fashion. Fortunately, it is possible to define uninformative but efficient priors. We set:

1. Uniformly chosen prior means (on the observed data space),
2. Isotropic prior covariances (diagonal values scaled to half of the observed variance)

This approach is valuable if we choose an enough big number of prior components. Doing this enables a sufficiently large search through parameter space, therefore avoiding the worst local optima. Experimentally, we have found that for low dimensional spaces (2 or 3), using 30 to 50 components is a rather good compromise to capture most of the clustering structure while keeping a low computational cost.

4 Experimental Results

We present here results obtained on the proposed method (let's name it VBMerge), first on an artificial, then on a real data set.

We want to evaluate the overall quality of estimations made with VBMerge. For this, we will compare the relative quality of models estimated with VBMerge with respect to a batch VBEM approach. Indeed, the reduction of a too rich model should not degenerate the estimation, in other words it should converge to an estimation close to the true distribution, while reducing significantly the number of components. For the comparison we will use 2 synthetic data sets (sampled from a 4-component 2-dimensional GM with randomly placed means, various covariance matrices and various populations per component) and the *Glass Identification* data set [26]. This data set describes 214 samples of glass with 9 physical (numerical) characteristics. Each sample is associated with a true label from 7 possible. We consider here that the partition of items w.r.t. labels constitutes the distribution we try to discover.

Then we will illustrate the joint usage of VBEM and VBMerge on the suggested geo-temporal context. This data set was introduced in [8] and describes 721 images over spatial and temporal attributes. It was obtained from a personal collection taken in several countries over 3 years. In [8] this data set was used to build a Gaussian mixture hierarchy designed to navigate conveniently in the obtained events. We shall see that VBMerge can be a promising building block in such a context.

4.1 VBMerge vs. VBEM

For the quality assessment, we used the following settings:

– For the synthetic data sets, VBEM is initialised randomly with $K = 30$ (a usually good value for 2D spaces). To generate a redundant GM we use a classic EM algorithm estimation ([27], [2] chap. 9) with $K = 8$.

Indeed, classic EM usually tends to overfit data, so this will be a good point if VBMerge is able to compensate this drawback while giving a good estimate.
- For the *glass* data set, as its dimensionality is much higher we use $K = 300$ for VBEM, as this is necessary to cover the data space properly. For classic EM we use $K = 30$,

For all data sets, and for each setting (VBEM or classic EM-VBMerge), we will measure:

- the empirical cross-entropy (the data set is considered as a sample from the true distribution) of the estimated model w.r.t the true (unknown) distribution.
- the KL divergence of the estimated model w.r.t the true distribution. Of course this calculation is possible only in the case of a synthetic data set, where the true distribution is known. An approximate value of this measure is obtained through a simple sampling scheme (see [2] chap. 11) (100000 draws)
- the number of components in the estimated model
- the couple error [28] measured between the inferred labels (i.e. which component is associated with each datum) and the true labels. This error measures if two items that are in the same (respectively a different) class w.r.t the original distribution also are in the same (resp. different) class in the estimated distribution.
- We made 20 measurements per study case.

Fig 1 displays the mean of these results, with associated standard deviation in brackets. As the variational estimation integrates important characteristics such as a low and appropriate number of components and a good separation between low-entropic components (for a discussion about these aspects, see [8]), we believe that the cross-entropic and KL values are significant quality measurements.

Obtained values on synthetic data set show that there is no significant difference between our VBMerge and the VBEM. Therefore our proposed algorithm applied on a redundant model can lead to a result as good as obtained with an

data \measure	VBEM cross-ent.	VBEM KL div.	VBEM nb. comp.	VBEM couple
synth1	790.3 [2.2]	0.207 [0.011]	5.4 [0.5]	0.19 [0.01]
synth2	781.1 [0.49]	0.054 [0.003]	3 [0]	0.14 [0.06]
GLASS	-366.7 [203.5]		4.0 [0.67]	0.38 [0.08]
data \measure	VBMerge cross-e.	VBMerge KL	VBMerge nb. comp.	VBMerge couple
synth1	817.9 [5.7]	0.118 [0.013]	4.5 [0.97]	0.15 [0.04]
synth2	794.0 [2.1]	0.095 [0.008]	3.7 [0.82]	0.16 [0.02]
GLASS	578.4 [41.2]		2 [0]	0.69 [0.08]

Fig. 1. Benchmarking results

usual variational estimate. On the other hand, the results are much worse on the *glass* data set. This must be due to a bad initial estimate by the classic EM algorithm. Indeed, this data set is not designed for Gaussian mixture modelling at all (e.g. contains some very low entropic variables). A classic EM approach is very sensitive to this problem, and will typically lead to a very degenerated optimum, as VBEM alone is much more robust to such issues [2]).

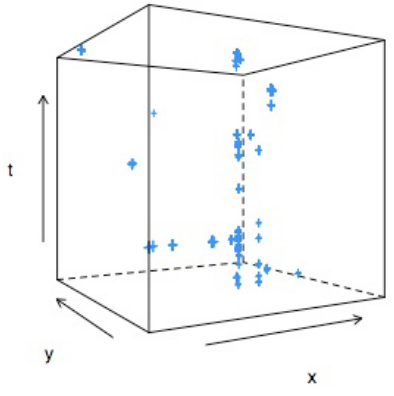

Fig. 2. 3D representation of the image collection (spatial and temporal coordinates). Each + represents a coordinate (t, x, y) of one image. Here Z is the temporal axis.

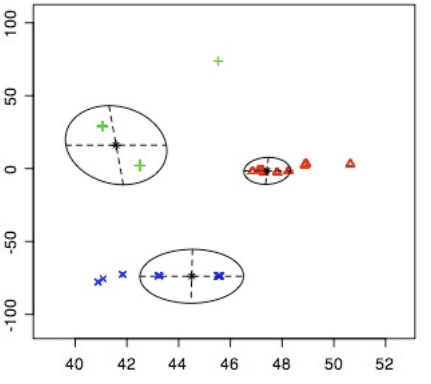

Fig. 3. Obtained 2D Reduced model. The ellipses are the variance of the components. The different symbols represent the data-to-class assignments. The 8 obtained clusters in the *finer* level are here summarized with 3 clusters.

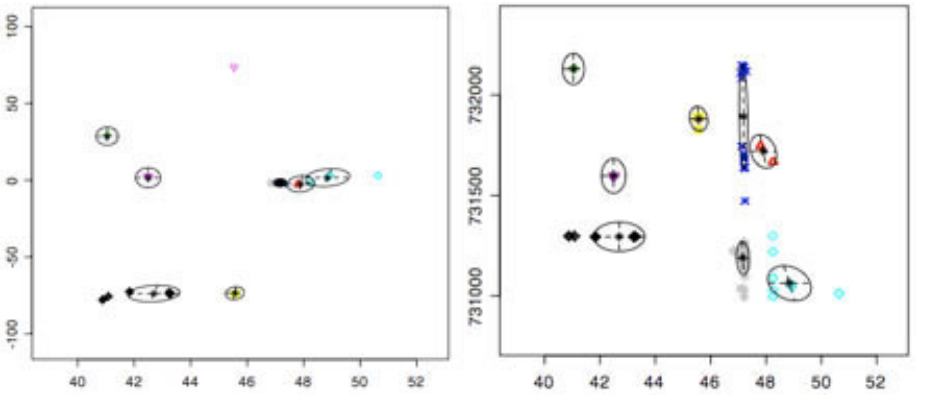

Fig. 4. 3D VBEM inferred model (xy and xz projections). The ellipses are the variance of the components. The different symbols represent the data-to-class assignments. We clearly identify some captured *events* and *places*.

4.2 Building a Mixture Hierarchy on Geo-temporal Data

The data set used for our experiment is represented in figure 2. We will follow these steps to build a simple hierarchy:

1. Estimate a model on geo-temporal data using VBEM (see figure 4),
2. Project this model on the (x, y) sub-space,
3. Reduce this model using VBMerge.

When merging components with VBMerge, we define implicitly a probabilistic mapping of the original components with the merged components. This allows to define a 2-level GM hierarchy.

The second level is presented in Figure4(a). We obtained 8 compact and well defined classes. All the main location of the user collection was correctly retrieved. The first level on Figure 3 displays the reduced model defined only on (x, y) space (clusters of the *finer* also appears). This can be an entry point for user navigation since the obtained summary seems pertinent: it is composed of just 3 clusters, with well defined boundaries. Nevertheless, the image group situated at the coordinate $[45, 75]$ is isolated: as in [8], small image group tends to be grouped with the nearest clusters.

A user can then browse its collection switching between the different obtained classifications: for example, he can first start by browsing our geographical hierarchy to select a specific location, and then switch to the initial 3-D partition, therefore offering a view of closely related places at a same time.

5 Conclusion

This paper discloses a technique for building a hierarchy of mixture models, where the main contribution is a variational approach to Bayesian clustering of components. The main features of the approach is that it possesses advantages of Bayesian modelling, while being computationally very tractable. The technique was motivated by an applicative need arising from indexing image collections, based on geo-temporal metadata, enriching the browsing possibilities within the collection.

Besides considering multi-users and integrating information from geographical information systems, such as work quoted in Section 2, we believe that structure from time-continuous capture of geolocation, rather than from the mere time instants of pictures, should be investigated to gain valuable insight.

References

1. Rodden, K.: How do people manage their digital photographs? In: ACM Conference on Human Factors in Computing Systems, Fort Lauderdale, pp. 409–416 (2003)
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)

3. Blekas, K., Lagaris, I.E.: Split-merge incremental learning (smile) of mixture models. In: de Sá, J.M., Alexandre, L.A., Duch, W., Mandic, D.P. (eds.) ICANN 2007, Part II. LNCS, vol. 4669, pp. 291–300. Springer, Heidelberg (2007)
4. Celeux, G., Govaert, G.: A classification em algorithm for clustering and two stochastic versions. Computational Statistics and Data Analysis (1992)
5. Celeux, G., Govaert, G.: Gaussian parcimonious clustering models. Pattern Recognition 28, 781–793 (1995)
6. Fraley, C., Raftery, A.E.: Mclust: Software for model-based clustering, density estimation and discriminant analysis. Technical report 415, Department of Statistics - University of Washington (2002)
7. Fraley, C., Raftery, A.E.: Model-based clustering, discriminant analysis and density estimation. Journal of the American Statistical Association (2002)
8. Pigeau, A., Gelgon, M.: Building and tracking hierarchical geographical & temporal partitions for image collection management on mobile devices. In: Proceedings of International Conference of ACM Multimedia, Singapore, pp. 141–150 (2005)
9. Attias, H.: A variational bayesian framework for graphical models. In: Advances in Neural Information Processing Systems (2000)
10. Vasconcelos, N.: Image indexing with mixture hierarchies. In: Proceedings of IEEE Conference in Computer Vision and Pattern Recognition (2001)
11. Goldberger, J., Roweis, S.: Hierarchical clustering of a mixture model. In: NIPS (2004)
12. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proceedings of the ACM Joint Conference on Digital Libraries JCDL, pp. 326–335 (2002)
13. Platt, J.C., Czerwinski, M., Field, B.A.: PhotoTOC: Automatic clustering for browsing personal photographs. Technical Report MSR-TR-2002-17, Microsoft Research (2002)
14. Toyama, K., Logan, R., Roseway, A., Anandan, P.: Geographic location tags on digital images. In: Proceedings of the eleventh ACM international conference on Multimedia, Berkeley, CA, USA, pp. 156–166 (2003)
15. Jaffe, A., Naaman, M., Tassa, T., Davis, M.: Generating summaries and visualization for large collections of geo-referenced photographs. In: Proceedings of the 8th ACM SIGMM International Workshop on Multimedia Information Retrieval, pp. 853–854 (2006)
16. Naaman, M., Song, Y.J., Paepcke, A., Garcia-Molina, H.: Automatic organization for digital photographs with geographic coordinates. In: Proceedings of the ACM/IEEE Conference on Digital libraries (JCDL 2004), pp. 53–62 (2004)
17. Cooper, M., Foote, J., Girgensohn, A., Wilcox, L.: Temporal event clustering for digital photo collections. In: Proceedings of the ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), vol. 1, pp. 269–288 (2005)
18. Gargi, U., Deng, Y., Tretter, D.R.: Managing and searching personal photo collections. Technical Report HPL-2002-67, HP Laboratories, Palo Alto (2002)
19. Kennedy, L., Naaman, M.: Generating diverse and representative image search results for landmarks. In: Proceedings of The Seventeenth International World Wide Web Conference, WWW 2008 (2008)
20. Nair, R., Reid, N., Davis, M.: Photo loi: Browsing multi-user photo collections. In: Proceedings of International Conference of ACM Multimedia, pp. 222–223 (2005)

21. O'Hare, N., Gurrin, C., Jones, G., Smeaton, A.F.: Combination of content analysis and context features for digital photograph retrieval. In: Proceedings of of the 2nd IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, pp. 323–328 (2005)
22. Akaike, H.: A new look at the statistical model identification. IEEE Trans. on Automatic Control AC-19(6) (1974)
23. Schwarz, G.: Estimating the dimension of a model. The Annals of Statistics 6, 461–464 (1978)
24. Vasconcelos, N., Lippman, A.: Learning mixture hierarchies. In: Neural Information Processing Systems (1998)
25. Blahut, R.E.: Principles and Practice of Information Theory. Addison-Wesley, Reading (1987)
26. Evett, I.W., Spiehler, E.J.: Rule induction in forensic science. Ellis Horwood in Expert Systems, Knowledge Based Systems, 152–160 (1989)
27. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. J. Royal Stat. Soc. B(39), 1–38 (1977)
28. Azzag, H.: Classification hiérarchique par des fourmis artificielles: application à la fouille de données et de textes pour le Web. PhD thesis, Ecole Doctorale Santé Sciences et Technologies, Université François Rabelais Tours (2005)