Incremental and hierarchical classification of a personal image collection on mobile devices

Antoine Pigeau

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

MyOwnLife: incremental and hierarchical classification of a personal image collection on mobile devices

Antoine Pigeau

Received: date / Accepted: date

Abstract Browsing multimedia collection on mobile devices raises the needs for new multimedia indexing solutions. In this paper, we focus on the management of personal image collections. We propose a method to simplify the browsing task on such a collection. The contributions reside in an incremental hierarchical algorithm, a method to provide a textual representation of the groups obtained and an algorithm to build a geo-temporal view of the collection.

The proposed incremental hierarchical algorithm builds a temporal tree from the time stamp of each image. We opt here for a combination of a supervised clustering and an incremental algorithm based on mixture model. Good properties of the hierarchy are determined automatically thanks to the Integrated Likelihood Criterion (ICL). Based on the events obtained, a textual representation is proposed and then used to improve our temporal classification, combining geographical and temporal information. Results are validated on several real user collections with our prototype *MyOwnLife*.

1 Introduction

Sharing personal images in a multi user collection is now widespread thanks to web sites as GoogleMap or Flickr. Everyone has now the possibility to publish pictures and to browse different user collections from all over the world. Finding specific images in such an aggregate collection is then a complex task, leading to a need for new tools of multimedia management and adapted human machine interaction. This latter point is stressed in the context of mobile devices. Indeed map-based interface and keyword

LINA (CNRS UMR 6241)
2, rue de la Houssinière 44322 Nantes cedex 03 - France
antoine.pigeau@univ-nantes.fr

queries are not adapted for this kind of device. Summarization of image sets seems then pertinent to ease the browsing task in both multi or single user context: displaying a limited set of representative images of each event's collection is of high interest to simplify the browsing task.

In this paper, we organize temporally and geographically a personal image collection. Several hierarchical classifications are carried out from the time-stamp and geographical meta-data, obtained for example from a GPS system integrated in the mobile device. Our objective is to favour the browsing task rather than querying, a point motivated by the partial memory that the user has of the collection, taking into account various contexts of visualization. Therefore, we provide several views of image collections: a temporal, a geographical and a geo-temporal view. Each one is represented by a hierarchical structure composed of events of the image collection. Users have then the possibility to switch between those views to browse/search images at different scales. This hierarchical aspect is salient since a user must now have an access to its collection from various terminals with various display capabilities. Our hierarchies supply adapted image summaries in accordance with the device constraints.

The main steps of our approach are:

1. **incremental building of a hierarchical temporal classification**: we propose an improvement of our incremental and hierarchical algorithm [1], which provides temporal and spatial classifications. The overall principle is that a hierarchy of Gaussian mixture models is progressively built, as new data enters the system. This algorithm is here improved to support the direct add of image group (contrary to one by one), a point that better takes into account the way users build their collection [2] since images are generally taken in bursts. A new incremental process is then proposed to update the hierarchy. Note that our temporal classification, contrary to a direct geo-temporal clustering, is motivated by the fact that the temporal criterion is more pertinent to emphasize the links between images. Indeed, an event can be composed of several distinct locations (for example a wedding);
2. **textual representation of classes**: each class at the previous step is represented with a textual set of labels. The goal is to provide a succinct representation of each event. Those summaries are built with new meta-data obtained from the initial time stamps and a GIS. The advantages of such an approach are the energy saving, since image displaying is costly in term of energy, and the possibility of keyword queries directly on events of the collection.
3. **combination of temporal and geographical information**: to improve our hierarchical temporal classification, we propose a method to re-structure classes based on their geographical information. Our approach consists in merging succes-

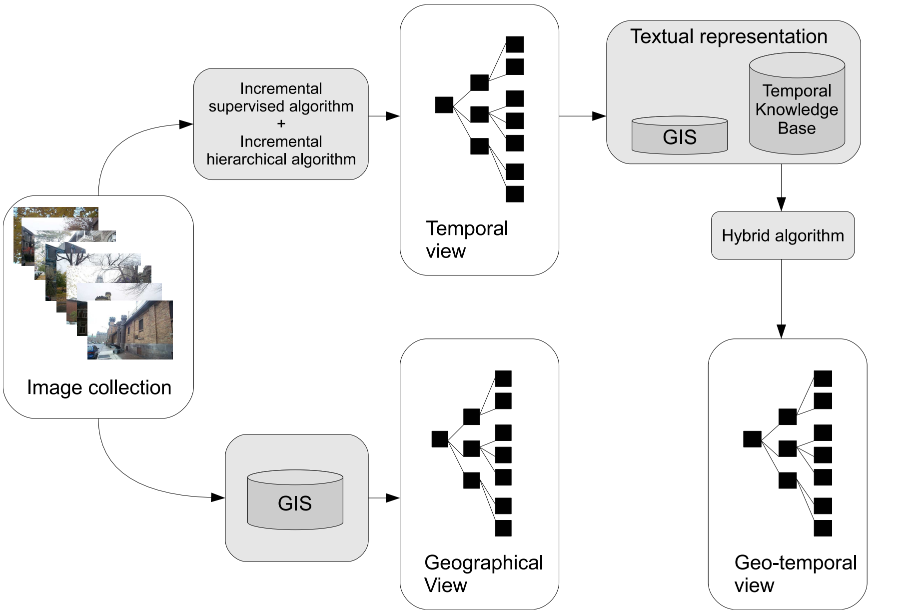

Fig. 1 Building of the geographical, temporal and geo-temporal views : the geographical view is built from meta-data obtained from a GIS. It is is then similar to a map-based classification. The temporal view is built with our proposed incremental algorithm based on the time stamp. Each temporal class is then represented with a textual summary, composed of a temporal part and a geographical part. Finally, the temporal partition is re-structured based on the textual geographical information to provide the geo-temporal view.

sive temporal nodes with similar geographical summary (similar location based on a geographical map).

4. **building of a hierarchical geographical classification**: we classify the images in accordance with geographical meta-data obtained from a Geographical Information System (GIS). The classification amounts to a map-based classification;

Figure 1 presents the different steps to provide our three distinct classifications of the image collection.

The classifications obtained enable then the user to navigate into a personal multimedia diary/photo album without having a particular target-picture in mind (as a passtime or to get an overview). Our summarization process provides a compact visual representation of each event, a pertinent point if the collection is browsed on mobile devices. Indeed, several events can be represented with a limited set of images thanks to the hierarchical aspect.

The remainder of this paper is organized as follows. Section 2 surveys work related to temporal and/or geo-location based structuring for the application at hand. Section 3 discloses the process to track a temporal hierarchy. Section 4 described how a geo-temporal hierarchy may be built, based on the geographical textual representation of each temporal class. Section 5 provides experimental results. Finally, the work is summarized and perspectives are sketched in section 6.

2 Related Work

Time stamp uses to be a favorite criterion to index a personal image collection. Its availability leads to several works [3–5] on the incremental segmentation of a time-stamp sequence. Detection of meaningful gaps enables to find events or hierarchy of events [5] in the collection. Such a boundary could be arbitrary defined or obtained from the average gap over a temporal window. In our previous work [1], we propose a method to compute it automatically, avoiding arbitrary parameters.

Most recent works on personal image indexation now deal with the image localization [6–9]. Indeed such a meta-data is now easily available thanks to GPS system, and helpful to automatically annotate the image [7, 10]. Several software, as for example the WWMX system [6] or Flickr, propose a map-based interface to browse the collection. The main problem of such an approach is that the map gets cluttered with a huge number of images. Several solutions were proposed to summarize an image sets: the work [6] proposes to aggregate images in accordance with the map scale, while [9] selects representative images from multi user collection based on their meta-data. Our proposal also seems relevant to solve this problem since our classification provides temporal and geographical summaries of the collection.

Combinations of the temporal and geographical meta-data are also proposed in [1, 2, 8, 11]. Similarly to this paper, [2] proposes to hierarchically organize an image collection, while [8] provides one level partition. To our understanding, [2] incorporates a series of rules derived from user's expectations to build a geo-temporal hierarchy of events. In [1], we proposed an incremental EM algorithm to carry out hierarchical temporal and spatial classifications. Here we propose an improvement of our clustering algorithm and an extension work to combine our temporal and spatial partitions.

Most recent contributions on image collection organization reside now in the multi user context. Popular websites as Flickr or GoogleMap enable users to share their own collection leading to a huge mass of image to deal with. Experiments on user with the Zurfer system [12] show that favorite organization criteria differ from the single user context: users prefer here to browse image sets according to social interactions (photos from friends and family members) and personal interests, associated for a large part to a specific location. Owner and location criteria seem here more pertinent than the temporal aspect.

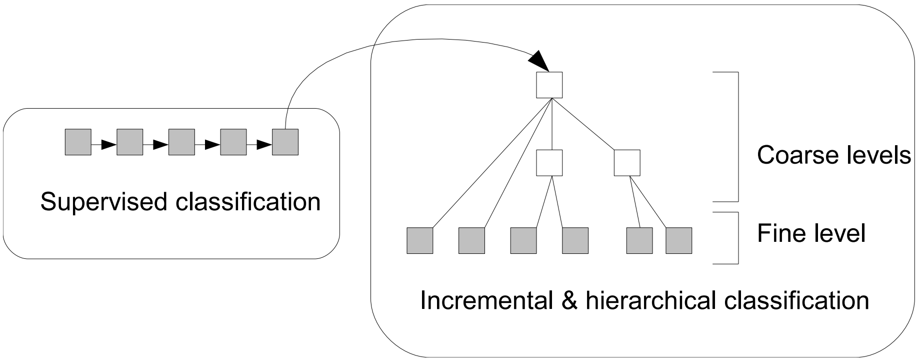

Fig. 2 Temporal hierarchical classification : the events obtained with our supervised algorithm are then added one by one in our hierarchical classification. The update of the tree is incremental.

Several works were proposed on organization of multi user collections [9, 12–14]. We notice that it generally focuses on the browsing task facilities of a pre-selected subset of images (generally a specific location) and not of the whole collection. The work [9] proposes a method to select representative photographs from a particular spatial region. The images are classified hierarchically based on their location and then ranked based on an empiric heuristic, emphasizing their originality. First images provide a summary of the selected place. A similar work, including content based criterion, is proposed in [12]. Another content based approach [14] combines content similarity and context features to retrieve photographs of monuments in a specific location. Finally, [13] proposes to emphasize salient events in several collections of users sharing social context. A visualization on a temporal axis enables detection of local optimum density of several collections.

Adapted to a multi user context, hierarchical clustering of a personal collection can provide several advantages. In addition with the fact that event sharing is facilitated by the summarization process, a new possibility is to find automatically similarities between various collections. Indeed, it could be interesting for users to directly share images of common events that they participated. Comparison of the different views of our partitions could emphasize those similarities.

3 Tracking temporal hierarchical partitions

This section describes how our temporal hierarchy is obtained. The proposed tree structure is built incrementally and the number of levels evolves as data streaming in provides evidence for this need. We opt here for a combination of a supervised clustering, for the *fine* partition (the set of leaves), and the mixture model framework, for the *coarse* partitions (providing several summaries of the *fine* partition). Figure 2 presents an overview of our method. The new images are first regrouped in accordance with their temporal proximity and each group obtained is then added one by one in our hierarchy.

As in [2], we set manually the precision degree of the *fine* partition fixing manually a boundary between events (for example, two successive images separated with more than 3 hours involve a new event/class). Such a solution seems meaningful since it corresponds to the building process of a personal image collection [2], and provides a clear and robust initial partition to build the summaries. Each leave of our temporal tree is then an event with temporally connected images.

Based on the *finer* partition obtained, we then build a hierarchical classification to provide several *coarse* summaries. To provide such summaries, we make the assumption that time stamps in an event follow a Gaussian distribution. Experiments in [1,15] support this hypothesis (more discussion on the time stamp distribution is available in [16]). The advantages of this modelling are the availability of (i) an agglomerative algorithm [17] providing a binary tree of Gaussian components and (ii) probabilistic criteria to select model complexities. Here, we propose to use the Integrated Likelihood Criterion (ICL)[18] to select Gaussian models in a binary tree obtained with [17].

The incremental update of our hierarchy with new images consists in the following steps:

1. build a *fine* partition of the new images and add each group obtained by the top of the hierarchy;
2. detect the correct subtree s to add the new image group;
3. build a new binary tree t from the leaves of s with the agglomerative algorithm [17];
4. select levels of t thanks to the ICL criterion, to avoid uninteresting and strongly redundant partitions. The ICL criterion provides a consistent solution to the issue.

In the following, we first present our modelling, the ICL criterion and the way we select levels in a binary tree, and then our incremental and hierarchical algorithm.

3.1 Gaussian model and ICL selection of relevant levels

To represent our hierarchy of events, we opt for the mixture model framework in which probabilistic models are associated both to classes and data-to-class assignments. This very latter point makes it attractive for the incremental nature of the task, since data-to-class assignments may be updated in a flexible way as new data streams into the system.

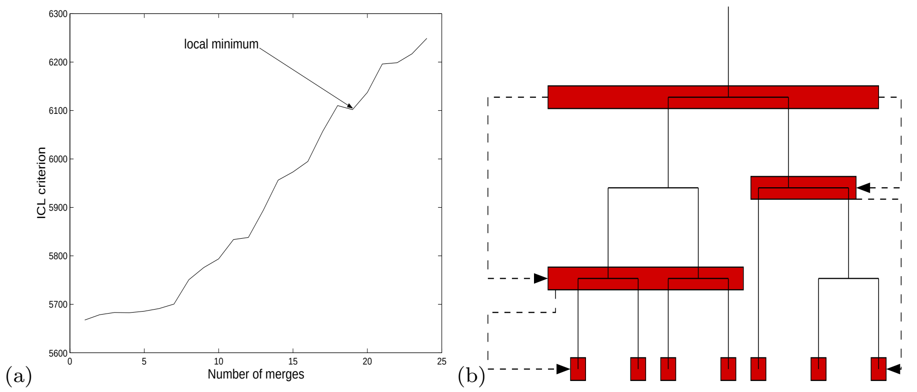

Fig. 3 Selection of levels corresponding to local optima of the ICL criterion: (a) the optimal ICL criterion found at each level of the binary tree represented on (b) is plotted. The rectangles indicate the corresponding selection of partitions. Once an optimum is found at a level l, we search for another local optima in each subtree from l. 'Local' minima here is to be interpreted as follows: both slightly coarser and slightly finer partitions are worse, in the ICL sense.

The data D (set of time stamps t) are assumed to be drawn from a random Gaussian mixture process with probability density:

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

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

In our tree, each group of the *fine* partition is then considered as a Gaussian component retrieved from its contained data. Each node i is defined with a center μ_i, a mixing proportion ω_i and a covariance Σ_i ($\theta_i = (\omega_i, \mu_i, \Sigma_i)$). Therefore, a level with K nodes can be interpreted as a Gaussian mixture $\Theta_K = (\theta_1, \theta_2, \dots, \theta_K)$. Each node i contains the image set of its children. Parameters θ_i and μ_i are computed based on i's data, and ω_i is retrieved from i's sibling nodes.

With mixture model modelling, a criterion for fair comparison between clustering hypotheses that might have different number of classes consists in comparing their integrated completed likelihoods (ICL) [18]. It is a variation of the BIC criterion which add an entropic term based on the data-to-class assignment to favor well-separated clusters. This criterion is defined as follows:

$$ICL = -ML + \frac{\nu_K log(n)}{2} - \sum_{k=1}^{K} \sum_{i=1}^{n} t_{ik} \cdot log(t_{ik}) \tag{2}$$

where $ML = \sum_{i=1}^{n} ln\Big(\sum_{k=1}^{K} \omega_k \cdot \mathcal{N}(x_i|\mu_k, \Sigma_k) \Big)$ is the maximized mixture log-likelihood, ν_K is the number of independent parameters in the model with K components, n is the number of data elements and t_{ik} is the posterior probability for an observation i originating from cluster k. These t_{ik} are in fact expectation values of the binary data-to-model assignment. In practice, they are supplied by the E step of the EM algorithm, generally used to determine the model parameters of a Gaussian model.

Roughly, the ICL criterion find the model with the best compromise between the model likelihood (first term of equation 2) and its complexity (second term). Indeed, the higher the number of components, the higher both the likelihood and the model complexity. The last term of equation 2 favors models where each data element is clearly assigned to one cluster (and so the values of the terms t_{ik} are close to 0 or 1): selection of models with distinct clusters is then favoured.

Here, we use the ICL criterion to compare different levels of a hierarchy of Gaussian components and select the ones with good properties. Apply on a binary tree obtained from our *fine* partition, our goal is to obtain a hierarchical partition which properties (width, height and number of children) are determined automatically by the data structure.

The different steps of our algorithm are:

1. from a Gaussian model, we build a tree of Gaussian components with the agglomerative algorithm [17];
2. from the top of the tree, we compute the ICL criterion of each level and select the first one with a local minimum. If no level is found, we stop at the leaves and obtain one single level, i.e., the root and all the leaves;
3. once an optimum is found at a level l, we search for another local optima in each subtree from l. 'Local' minima here is to be interpreted as follows: both slightly coarser and slightly finer partitions are worse, in the ICL sense.

Figure 3 shows an example of level selections. Figure 3(a) is the ICL value of the binary tree represented in 3(b). The first local minimum is associated with the first selected level (the first square from the root). Then we search for another local minimum in each subtree from the selected level. The final tree is then represented by the squares.

3.2 Our incremental and hierarchical algorithm

To add new images in our classification, we first build a *fine* partition. We group together new images with less than T_{diff} hours of time differences. This parameter represents the temporal resolution of the event clustering. Our summaries are then built from these leaves.

The algorithm is detailed in Algorithm 1. In a nutshell, it proceeds in 4 steps:

Algorithm 1 Incremental update of our temporal hierarchy.

Building the *coarse* classification is an incremental process. Each group is added by the root of our hierarchical classification and a node q is updated as follows:

1.Initialisation: if q is the root with no child, the group is added as a new child of q;

2.Update:

if q is a leaf **then**

2.1.add the new group in this node and parameters θ_i of the tree are updated (from q to the root).

else

2.2.detect the change due to the new image group: we retrieve the model composed of q's children and compute the likelihood l of the new group, based on the parameters of the Gaussian model obtained.

if $l > \beta$ **then**

3.1.it means that the new group is associated with an existing cluster, selected with the maximum a posteriori (MAP) criterion; we update this node and re-iterate step 2;

else

3.2.add a new component q_{new}, as a new child of q, associated with the new image group and go to the next step;

3.3.search for summaries: we build a new binary tree from the **leaves** of the N_{near} nearest nodes of q_{new} and its other sibling nodes with algorithm [17]. This agglomerative algorithm provides a binary tree from a Gaussian mixture;

3.4.select the summaries in the binary tree. This selection is carried out with the ICL criterion (fig.3);

3.5.merge the subtree updated with the initial one.

end if

end if

1. detect the correct level to add the new images: the likelihood criterion, compared to a threshold β, enables to find the subtree s to update;
2. re-build a local binary tree t: to improve the quality of summaries, parameter N_{near} enables to select the nodes from which the search for new summaries is carried out from the leaves;
3. select levels of t with our ICL selection;
4. update the main tree.

Our incremental algorithm depends on parameters β and N_{near}:

– parameter β is a threshold used to check if the initial model parameters of a level updated fits well the new image group (see step 2.2). The decision to add the new group in the current level or to propagate the update depends on this threshold. A high likelihood for the new images indicates that they can be included in an existing node. The value of β is set manually: the higher its value, the more distinct the classes in the summaries, and the lower the height of the tree (leading to a small

id	decade	year	season	month	π
1	90's	1999	winter	December	1.0
2	00's	2000	winter	January	0.9
3	00's	2000	winter	January	1.0

id	continent	country	state	city
1	North America	USA	New York	Rochester
2	North America	USA	New York	NY city
3	North America	USA	New York	NY city

Table 1 Example of attributes and values obtained respectively from the time stamp and the location. The column π represents the image-to-class assignment used to compute the score α.

number of summaries). If the parameters fit well the new group, the MAP criterion is used to select the node to propagate the update;

– The selection of the N_{near} sibling nodes at step 3.3 attempts to avoid poor summaries, due to the incremental property of our algorithm. Indeed, a new group of images can have an influence on the whole tree, meaning it could lead to a new *coarse* summary. Then, parameter N_{near} sets the number of sibling nodes from which the search for new summaries is carried out from the leaves. Others nodes on the same level updated are also used but not from their leaves. The higher N_{near}, the higher the summary quality and the calculation complexity. Generally, setting N_{near} to a high value would consist in rebuilding the subtree from scratch (from the leaves).

Figure 4 presents an update example with our incremental algorithm.

4 Geo-temporal partition

We present here a method to summarize a class with textual labels and then a technique to improve our temporal classification. The approach consists in combining the classification with the geographical textual summaries.

Let us recall that, for each image, time stamp and location are the initial meta-data recorded with a mobile device. Then, a knowledge base can be used to provide information meaningful for the user from the initial meta-data. For example, given a GPS coordinate, a GIS could provide the continent, the country and so forth. A temporal knowledge base can provide the season, the day of the week or the time of the day (e.g. “afternoon”). Of course, diverse meta-data could be added to each image such as the altitude, the weather or the time zone as showed in [19, 20].

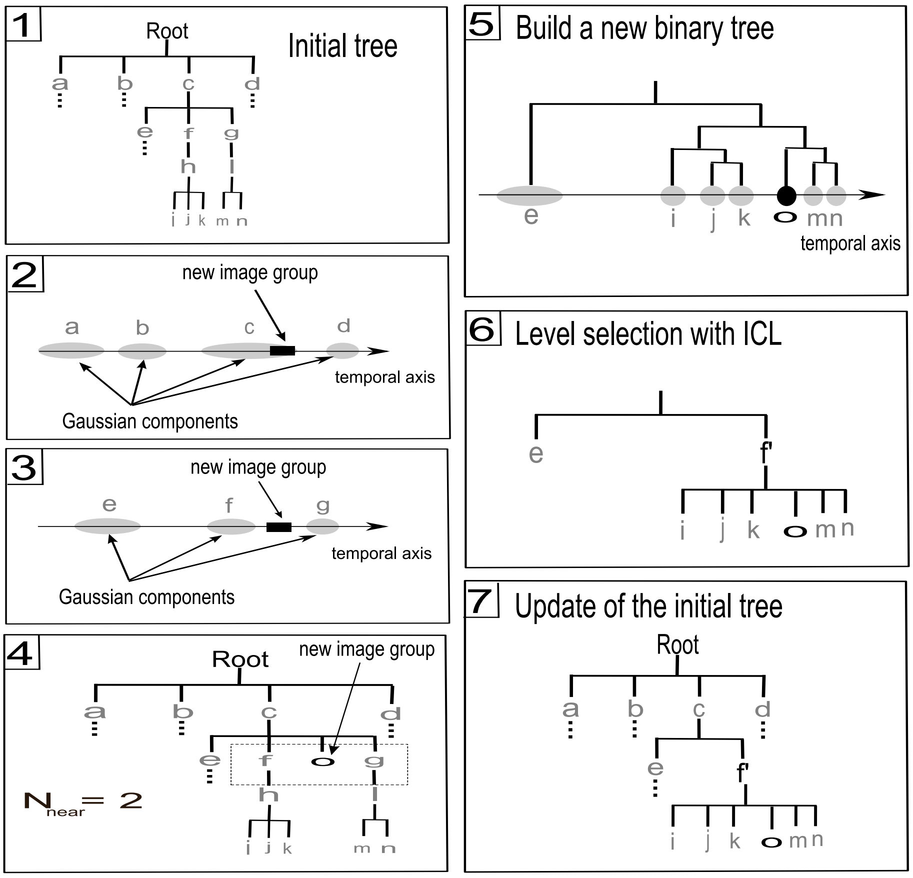

Fig. 4 Update of a subtree with a new image group: first, we add the new data group by the root of the initial tree (fig.4.1), retrieve the Gaussian model associated to its children and compute the likelihood of the new image group based on the initial parameters (fig.4.2). In this example, the new group is assigned to the component c: the likelihood is higher than β and the choice of the node c is based on the MAP criterion. The parameters of the components c clearly fit well the new image group: the update is then propagated to the node c. We then retrieve the model associated to its children and compute the likelihood of the new group (fig.4.3). Here the likelihood is lower than β since the new image group is not contained in any existing group. Now that we find the right position of the new image group, we need to update the hierarchical classification. This update is local since only the subtree of root c is involved. Fig. 4.4 presents the new node (node 'o') and its nearest sibling nodes (the square drawn with dashed lines). Here, the parameter N_{near} is equal to 2. We then re-build a binary tree from the node e, the leaves of the nearest node (i, j, k, m and n) and the new node o (fig. 4.5), and select the relevant levels based on the ICL criterion. Fig. 4.6 presents the new subtree obtained after the ICL selection. Finally the initial tree is updated with the subtree (fig. 4.7). Note that the children of the node e are kept.

Table 1 presents temporal and geographical meta-data for 3 images defined on 4 attributes in both cases. Notice that the attributes are organized in a hierarchical way (from left to right). Column π represents the image-to-class obtained from our temporal hierarchy. We use this score to emphasize the meta-data values of images strongly associated to their class.

First, we first define an *image summary*, *class summary* and *level summary*, and then we describe our geo-temporal approach.

4.1 The image summary

First, an *image summary* i is defined by 2 sets of meta-data, namely temporal and geographical:

$$\begin{aligned} m_i = \{ & \langle t_1, t_2, \ldots, t_L | t_l \in M_t \rangle, \\ & \langle s_1, s_2, \ldots, s_{L'} | s_l \in M_s \rangle \} \end{aligned}$$

where t_l (resp. s_l) is a textual label defined for attribute l, L (resp. L') is the number of temporal (resp. geographical) attributes to represent an image according to the knowledge base, and M_t and M_s are respectively the sets of label values defined on the temporal and geographical attributes. For example, image 3 in Table 1 is defined as:

$$\begin{aligned} M_3 &= \{m_3(t), m_3(s)\}, \\ M_3 &= \{\langle \mathbf{00's}, \mathbf{2000}, \mathbf{winter}, \mathbf{January} \rangle, \\ &\quad \langle \mathbf{North\ America}, \mathbf{USA}, \mathbf{NY}, \mathbf{NY\ city} \rangle \} \end{aligned}$$

4.2 The class summary

We build the *class summary* based on the *image summaries* associated to the images it contains. Let k be a class, its summary is defined as:

$$\begin{aligned} c_k &= \{c(t), c(s)\}, \\ c_k &= \{\langle t_1, \ldots, t_{l-1}, \{\alpha_1/t_l^1, \ldots, \alpha_r/t_l^r\} \rangle, \\ &\quad \langle s_1, \ldots, s_{l'-1}, \{\alpha_{1'}/s_{l'}^1, \ldots, \alpha_{r'}/s_{l'}^{r'}\} \rangle \}, \end{aligned}$$

where $t_l \in M_t$, $s_{l'} \in M_s$, l and l' are the first attributes from which label values present differences. For attribute l (or l'), the summary is represented by r (or r') different values associated to the images contained: α is the average weight of each textual label.

This score, for both the temporal and spatial case, is the average of the image-to-class assignment π (see the temporal Table 1) for each textual value:

$$\alpha_{t_l^r} = \frac{\sum_{i|t_i=t_r^l} \pi_i}{\sum_{i=1}^{n_k} \pi_i}$$

where π_i is the assignment probability of image i to its class and n_k is the number of images in the class.

The use of a weighted score, that depends on the data-to-model assignment, emphasizes labels strongly associated to its class, which are likely to better represent its content. For example, if a class contains the images of Table 1, its *class summary* is defined as:

$$c_1 = \{\langle\{\frac{1}{2.9} = 0.35/\textbf{90's}, \frac{1.9}{2.9} = 0.65/\textbf{00's}\}\rangle,$$
$$\langle\textbf{North A.}, \textbf{USA}, \textbf{NY}, \{0.35/\textbf{Rochester}, 0.65/\textbf{NY city}\}\rangle\}$$

Here we have $l = 1$ and $l' = 3$. Indeed, the attributes *Years* and *Cities* present different values. Stopping the textual summary at attributes with different values simplifies their visual representation and avoids building sets of labels that do not belong to the same time interval or location.

4.3 The level summary

The same principle is applied to obtain the textual summaries of a node in our hierarchy, called *level summaries*. We build the *class summary* of each child and then find the first attribute presenting different values. Each child is then only represented by the attributes with equal values and the first one with differences. For example, let c_2 be a *class summary* defined as:

$$c_2 = \{\langle\textbf{00's}, \{0.42/\textbf{2005}, 0.58/\textbf{2006}\}\rangle$$
$$\langle\textbf{Europe}, \textbf{France}, \textbf{Île de F.}, \{0.33/\textbf{Paris}, 0.66/\textbf{NY city}\}\rangle\}$$

And let c_1 and c_2 be two children of a node c, their *level summaries* are then defined as:

$$c_{children(c_1)} = \{\langle\{0.33/\textbf{90's}, 0.66/\textbf{00's}\}\rangle, \langle\textbf{North America}\rangle\}$$
$$c_{children(c_2)} = \{\langle\textbf{00's}\rangle, \langle\textbf{Europe}\rangle\}$$

The summaries are limited to the *Decade* and *Continent* attributes. This emphasizes the main differences between events on a same level.

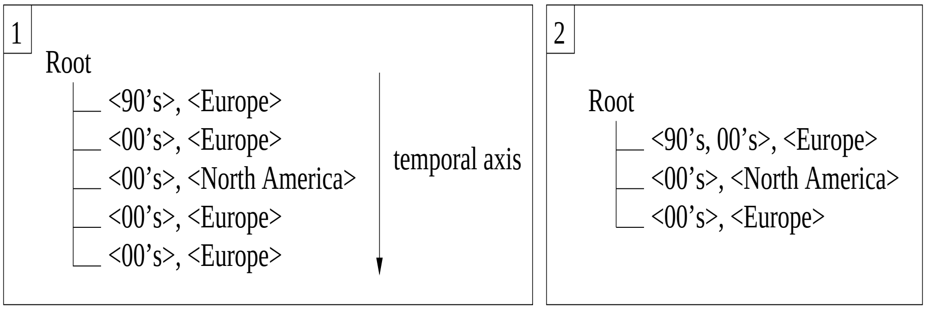

Fig. 5 Building our geo-temporal partition: (1) represents the initial temporal classification where each node is defined by its *level summary*. Geo-temporal partition is then obtained by merging temporal continuous nodes with similar geographical summary, as showed on (2). Note that the merge step includes the update of the temporal summary of each hybrid node (now, the first node in (2) contains values <90's,00's>).

4.4 Our geo-temporal approach

Finally, we propose a method to provide a hybrid partition of the image collection from the hierarchical temporal classification. Our method follows the assumption that successive temporal events in a same location are generally connected. In a nutshell, the approach consists in merging continuous temporal nodes with similar geographical meta-data. Let q be the root of the tree, our algorithm proceeds as follows:

1. get the geographical *level summaries* of q;
2. for each continuous temporal classes i and j (children of q) with a similar geographical summary: if i is a leaf and j a node (resp. j is a leaf and i a node) then move i as a new child of j (resp. j as a new child of i) else merge i and j (a new node containing the children of i and j).
3. apply step 2 to each child of q.

Practically, the hybrid tree presents the property that each node on a same level is temporally and geographically disconnected: a gap between two nodes is due to both a temporal and a location change. Figure 5 presents a combination example of temporal and geographical views.

5 Experiments

Experiments were carried out on three real user collections: G.B. (721 images taken over 4 years), C.C. (1731 images taken over 3 years) and S.P. (706 images taken over 4 years). These collections contain images taken on several continents, i.e. Asia, North America and Europe. For each collection, we built a hierarchical temporal classification

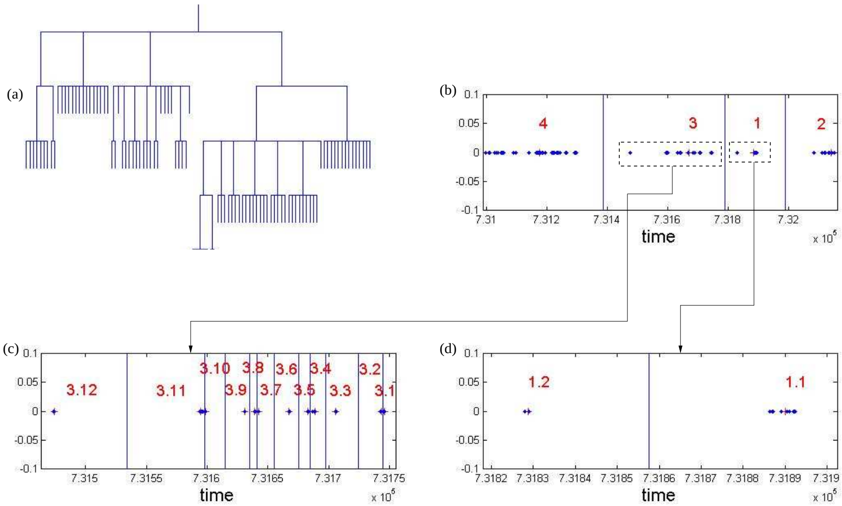

Fig. 6 Hierarchical temporal tree of G.B. collection. Figure (a) presents the tree structure while fig. (b), (c) and (d) show classification examples. Solid lines represent the boundaries between components and the dots indicate the temporal data. Figure (b) represents the coarser level of our temporal tree, while fig. (c) and (d) respectively show the children of components 3 and 1. Partitions generally present distinct clusters with visually justified boundaries, although over-segmentation occasionally occurs (for example components 3.10 and 3.11 of fig. (c)).

and asked the users to annotate the temporal events obtained to build a geographical decision tree. Finally, we applied our geo-temporal algorithm to provide a hybrid view. Parameters T_{diff} and N_{near} were set respectively to 3 hours and 2 nodes.

5.1 classifications obtained

The temporal tree obtained for the collection of G.B. is presented on figure 6(a). The tree is composed of 5 levels and is well-balanced (clearly, this depends both on the data and the classification technique, but is anyway a good property for browsing). The number of children per node varies from 2 to 15. Note that our classification extends in depth and width as new data is added. Only a minority of image groups implies serious restructuring of the tree, and hence, the overall computational cost grows almost linearly with the number of images.

Figures 6(b), (c) and (d) present partitions obtained at various levels of our tree. Figure 6(b) shows the coarsest level. All components are well delimited, providing

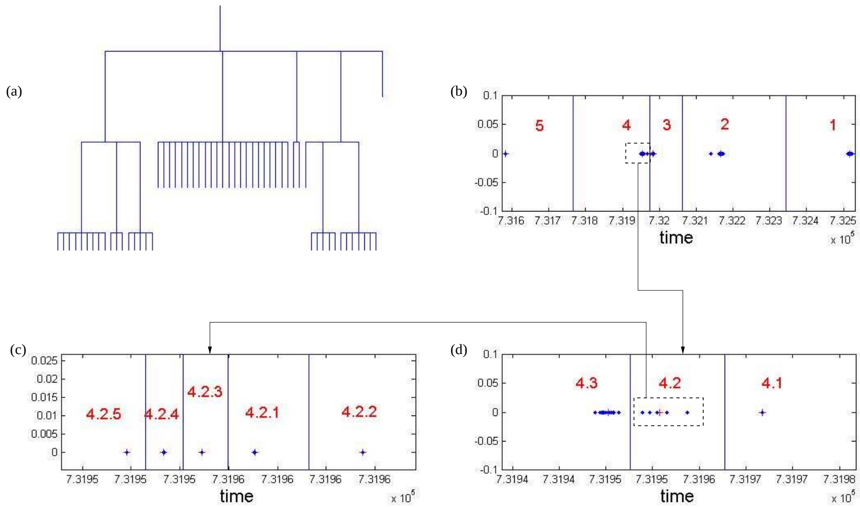

Fig. 7 Hierarchical temporal tree of C.C. collection. Figure (a) presents the tree structure while fig. (b), (c) and (d) show classification examples. Solid lines represent the boundaries between components and the dots indicate the temporal data. Figure (b) represents the coarser level of our temporal tree, fig. (d) shows the children of component 4 and fig. (c) presents the children of component 4.2. As in previous experiment, partitions present distinct clusters with visually justified boundaries and over-segmentation occasionally occurs (here, components 3 and 4 of fig. (b)).

relevant summaries of the collection. Components 1 and 3 on figure 6(b) are respectively detailed in figure 6(d) and (c). Children of the components 1 and 3 provide well-defined partitions since all the temporal gap are correctly emphasized. For component 3, meaningful temporal episodes are found but we notice over-segmentation, such as groups 3.10 and 3.11, certainly due to a larger evidence of small samples associated to one component.

The trade-off between temporal structural flexibility and computational load can be evaluated as follows. During the process of our algorithm on the temporal data, the agglomerative algorithm regenerating a binary tree has been called for all the iteration (here 98 leaves/events), and for each call, concerned a re-building from the leaves on average 22% of the data. It means that a large part of our tree remains stable all along the incremental process.

The tree obtained for collection C.C. in figure 7(a) is composed of 3 levels and is well-balanced. The number of children per node varies from 3 to 23. Figures 7(b), (c) and (d) present partitions at various levels of our tree. Figure 7(b) shows the coarsest level. Again all components seem well delimited, providing relevant summaries of the collection. Component 4 on figure 7(b) is detailed in figure 7(d), and component 4.2 on figure 7(d) is showed in figure 7(c). All the boundaries seem visually justified. We can just notice an over-segmentation for the components 4 and 3 on figure 7(b).

56 iterations were necessary to build the temporal tree, and at each iteration, the algorithm processed 21% of the data on average. Again the tree was stable during the incremental update.

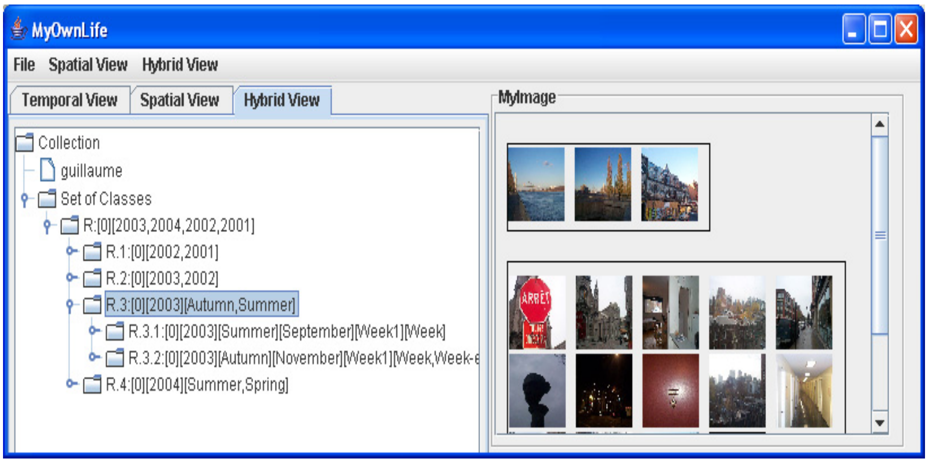

Fig. 8 Screenshot of our prototype *MyOwnLife*. Here, the hybrid view is selected and the summary of the selected node in the tree is displayed: this node contains 2 events, each one represented by 2 images per each leave included in their subtree. Thus, each event of the *fine* partition appears. Each node of the temporal tree is here represented by its textual summary.

5.2 User ratings

Our prototype *MyOwnLife* (see figure 8) is then used to evaluate the partitions. The prototype enables users to browse the different views (temporal, geographical or geo-temporal). The panel that contains images has a size similar to an Iphone's screen. Figure 8 is then an example of how our proposal could enhance the browsing task on a mobile device: the events are clearly depicted and summarized with a limited number of images. A touch interface could then be proposed to browse the hierarchies. Another solution could consist in mixing images and textual summaries.

To assess the improvement of the geo-temporal classification, we asked users to mark events and summaries, i.e., respectively, the leaves and the nodes of the tree.

	Bad	Average	Good
S.P. Collection			
Events/Leaves	0%	8%	92%
Temporal Summaries	15%	24%	61%
Geo-temporal summaries	0%	12.5%	87%
C.C. Collection			
Events/Leaves	5%	9%	86%
Temporal Summaries	10%	40%	50%
Geo-temporal summaries	6%	41%	53%
G.B. Collection			
Events/Leaves	9%	10%	81%
Temporal Summaries	5%	17%	78%
Geo-temporal summaries	3%	8%	89%

Table 2 Assessment of the events and summaries for collections S.P., C.C. and G.B. respectively.

Results are reported in Table 2. The different marks and their significations are as follows:

– *Good*: for a leaf, all its images belong to the same event. For a node, all its children represent a detail of a same event;
– *Average*: for a leaf, the set of image set represents several events that are temporally close, however some images are not related. For a node, its children represent well-defined events, but some events are not related (generally due to an isolated event, temporally close of its siblings). Users expect a different segmentation, even if the images are temporally close.
– *Bad*: for a leaf, a high proportion of images is not related. For a node, a high proportion of events is not related. The regrouped events do not make sense.

For each collection, we obtain a high score for the events, thanks to the small value of T_{diff}: $81\% - 92\%$ of the leaves are considered as *Good*. The comparison between the temporal and geo-temporal tree shows that our combination of the temporal partition with the geographical textual summaries improves significantly the summaries. We obtain an average increase of 13.5% for the *Good* summaries. The *Bad* summaries are due to over-segmentations of connected events (connected events of some summaries are not regrouped in a same node) or under-segmentations (disconnected events in a same node).

6 Conclusion

This paper deals with the issue of managing personal image collection. Our proposal is an incremental algorithm that provides a hierarchical temporal classification. This classification is then combined with geographical information to provide a geo-temporal partition. Our contributions focus on:

- the summarization process of the collection, based on supervised structuring and Gaussian mixture for the summarization aspect. The summary selection are guided by the data structure;
- the textual representation of each event, where main textual keywords are emphasized based on their data-to-class assignment;
- the building of a geo-temporal classification, providing a pertinent way to browse the image collection.

Our experiments are carried out on real user collections and present encouraging results.

Now, we are examining solutions to provide a semi-supervised algorithm, taking into account user constraints during the structuring process. The objective is to give users the possibility to modify or create their own events. An adaptation of our clustering algorithm is then necessary to handle these constraints. Another perspective work is the detection of event similarities between various personal image collection. Sharing or merging events of various collections is a domain of high interest.

References

1. A. Pigeau and M. Gelgon, “Building and tracking hierarchical geographical & temporal partitions for image collection management on mobile devices,” in *Proceedings of International Conference of ACM Multimedia*, Singapore, Singapore, Nov. 2005, pp. 141–150.
2. M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina, “Automatic organization for digital photographs with geographic coordinates,” in *Proceedings of The ACM/IEEE Conference on Digital libraries (JCDL’2004)*, Jun. 2004, pp. 53–62.
3. A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd, “Time as essence for photo browsing through personal digital libraries,” in *Proceedings of The ACM Joint Conference on Digital Libraries JCDL*, Jun. 2002, pp. 326–335.
4. J. C. Platt and B. A. F. M. Czerwinski, “PhotoTOC: Automatic clustering for browsing personal photographs,” Microsoft Research, Tech. Rep. MSR-TR-2002-17, Feb. 2002.
5. M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox, “Temporal event clustering for digital photo collections,” in *Proceedings of The ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 1, Aug. 2005, pp. 269–288.
6. K. Toyama, R. Logan, A. Roseway, and P. Anandan, “Geographic location tags on digital images,” in *Proceedings of The eleventh ACM international conference on Multimedia*, Berkeley, CA, USA, Nov. 2003, pp. 156–166.

7. W. Viana, J. B. Filho, J. Gensel, M. V. Oliver, and H. Martin, “Photomap - automatic spatiotemporal annotation for mobile photos,” *Web and Wireless Geographical Information Systems, Lecture Notes in Computer Science*, vol. 4857/2007, pp. 187–201, Apr. 2008.
8. Y. A. Lacerda, H. F. de Figueir, C. de Souza Baptista, and M. C. Sampaio, “Photogeo: A self-organizing system for personal photo collections,” in *Tenth IEEE International Symposium on Multimedia*, 2008, pp. 258–265.
9. A. Jaffe, M. Naaman, T. Tassa, and M. Davis, “Generating summaries and visualization for large collections of geo-referenced photographs,” in *Proceedings of The 8th ACM SIGMM International Workshop on Multimedia Information Retrieval*, Oct. 2006, pp. 853–854.
10. C. Chen, M. Oakes, and J. Tait, “A location data annotation system for personal photograph collections: Evaluation of a searching and browsing tool,” in *International Workshop on Content-Based Multimedia Indexing (CBMI)*, Jun. 2008, pp. 534–541.
11. ——, “Browsing personal images using episodic memory (time + location),” *Advances in Information Retrieval, Lecture Notes in Computer Science*, vol. 3936/2006, pp. 362–372, Mar. 2006.
12. L. Kennedy and M. Naaman, “Generating diverse and representative image search results for landmarks,” in *Proceedings of The Seventeenth International World Wide Web Conference (WWW 2008), To Appear*, Apr. 2008.
13. R. Nair, N. Reid, and M. Davis, “Photo loi: Browsing multi-user photo collections,” in *Proceedings of International Conference of ACM Multimedia*, Nov. 2005, pp. 222–223.
14. N. O’Hare, C. Gurrin, G. Jones, and A. F. Smeaton., “Combination of content analysis and context features for digital photograph retrieval,” in *Proceedings of The 2nd IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, 2005, pp. 323–328.
15. M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox, “Temporal event clustering for digital photo collections,” in *Proceedings of The ACM Multimedia*, Nov. 2003, pp. 364–373.
16. U. Gargi, “Modeling and clustering of photo capture streams,” in *Proceedings of the 5th ACM SIGMM workshop on Multimedia Information Retrieval*, Toronto, Canada, Aug. 2003, pp. 47–54.
17. C. Fraley, “Algorithms for model-based Gaussian hierarchical clustering,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 270–281, 1999.
18. 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*, vol. 22, Jul. 2000, pp. 719–725.
19. M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina, “Automatically generating meta-data for digital photographs with geographic coordinates,” in *International World Wide Web Conference archive, Alternate track papers & posters of the 13th international conference on World Wide Web*, May 2004, pp. 244–245.
20. R. Sarvas, E. Herrarte, A. Wilhelm, and M. Davis, “Metadata Creation System for Mobile Images,” in *Mobile Systems, Application, And Services (Mobysis)*, Jun. 2004.