Organisation statistique spatio-temporelle d'une collection d'images acquises d'un terminal mobile géolocalisé

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

A. Pigeau M. Gelgon
Laboratoire d'Informatique Nantes-Atlantique (LINA) / projet INRIA ATLAS

2, rue de la Houssinière - BP 92208, 44322 Nantes cedex 03 - France
Mail:nom@irin.univ-nantes.fr

14 novembre 2003

Résumé

Nous présentons une technique automatique d'organisation de collection d'images personnelles, pour répondre aux besoins particuliers émergents des téléphones portables équipés d'appareil photographique. Après avoir examiné ce qui fait la particularité de ce contexte, nous proposons une technique de structuration de collection d'image basée sur la date et le lieu de prise de vue des images. L'objectif est formalisé comme un problème de classification non-supervisée, temporelle et spatiale. Le critère statistique de vraisemblance complétée intégrée (ICL) est retenu, car il fournit une solution efficace pour déterminer la complexité du modèle et un bon niveau de séparabilité de ses composantes, tout en limitant le caractère arbitraire de la paramétrisation. La fiabilité des classifications obtenues est ensuite évaluée, afin d'en sélectionner la plus pertinente, pour fournir une structure utilisable avec une interface de type calendrier électronique permettant d'explorer la collection.

Mots Clefs

Recherche d'image, application sur mobiles, métadonnées temporelles et spatiales, classification statistique.

Abstract

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

Keywords

Image retrieval, mobile applications, spatio-temporal meta-data, statistical clustering.

1 problème et contexte

La recherche d'images par le contenu est un domaine de recherche très étudié ces dernières années, et est maintenant doté de récentes contributions [14] et de synthèses [24]. Il a été récemment souligné dans [1] que l'extension de ce domaine de recherche aux collections d'images personnelles était un enjeu important, et que peu de propositions dans ce domaine avaient été réalisées. Récemment, des solutions ont émergé des côtés industriel [10, 16, 18, 21, 22] et académique [13, 15]. On trouvera un état de l'art sur ce sujet dans [19].

Nous pensons que, avec les collections d'images personnelles, une niche applicative intéressante de l'indexation multimédia est en train d'émerger. Vis-à-vis des PC de bureau ou même des appareils photos numériques, les téléphones portables sont toujours à portée de main de l'utilisateur et permettent une transmission aisée de ces images (MMS). Ces propriétés les rendent intéressants pour l'ac-

quisition et la consultation d'images, mais ouvrent d'autres problèmes. Une autre caractéristique essentiel pour notre proposition est le très bon potentiel de géolocalisation de ce terminal. En effet, il peut combiner le GPS embarqué aux techniques de triangulation dans réseaux de mobiles - combinaison avantageuse puisque, dans l'ensemble, les deux dispositifs fonctionnent de manière satisfaisante dans des zones géogaphiques complémentaires.

Au fur et à mesure que les images sont collectées, elles constituent progressivement la mémoire d'une vie, qui peut être plus tard consultée, un peu ou beaucoup plus tard, dans un but pratique, émotionnel, ou pour le loisir. Les utilisateurs doivent être capables de retrouver un ensemble précis d'images dans leur collection, facilement et rapidement, ou de pouvoir la parcourir globalement pour se faire une idée générale de son contenu.

Un état de l'art des différentes techniques se rapportant aux téléphones portables munis de capteur d'images ("camera phones") est fourni dans [21], tandis qu'une présentation de leur usage peut être trouvée dans [26]. Une conclusion de [21] est que la gestion (recherche et stockage) de collection d'images dans ce contexte est un besoin industriel important et un problème ouvert. De plus, nous croyons que ce domaine de recherche est assez clairement défini (probablement mieux que la recherche d'images par le contenu) et en général son intérêt quotidien est avéré.

Les collections d'images personnelles acquises à partir d'un téléphone portable se distinguent des bibliothèques d'images numériques usuelles (exemple : web général) sur plusieurs points :

- le contenu (nature des scènes, ordonnancement de la collection, présence de méta-données) ;
- l'utilisateur a une connaissance partielle de la collection qui affecte la manière dont il parcourt l'ensemble des images. La recherche par exploration présente des avantages sur cet aspect ;
- les critères de recherche, détaillés ci-dessous.

En ce qui concerne le dernier point, les études sur les utilisateurs [23] montrent que les critères de recherche préférés sont la date, le lieu et les annotations/le contenu/le contexte. La présence de personnes sur l'image est aussi un critère apprécié. D'un autre côté, les critères classiques tel que la couleur, la forme, la disposition et la texture sont jugés peu pertinents.

La recherche d'image par le contenu a souvent été justifiée par le manque de méta-données. Dans le cadre des téléphones portables, on a accès à de nouvelles méta-données temporelles et spatiales [30] qui peuvent être utilisées pour classer les images et qui de plus correspondent aux critères de recherche appréciés par les utilisateurs.

Par ailleurs, les considérations d'interactions homme-machine sont importantes, puisqu'elles peuvent définir ou affecter l'analyse du contenu. Comme défendu dans [11], en considérant les contraintes d'interaction homme-machine (entrée/sortie) sur des mobiles, nous croyons qu'un point important est la capacité à produire des résumés selon les critères identifiés ci-dessus, pour satisfaire les besoins de visualisation et d'exploration. Nous proposons une contribution dans cette direction.

Et enfin, la structuration de collections d'images doit être accompagnée de systèmes de gestion complexe afin de permettre à un utilisateur d'y accéder facilement. En effet, il doit pouvoir être capable de consulter ses images à partir de son mobile ou, par exemple, de son ordinateur personnel. Afin de gérer des accès multiples et la sécurité des données, un système distribué est envisagé [20], avec une gestion spécifique des réplications, prenant en compte la variation des performances du réseau et le coût des transmissions (par exemple WLAN vs. GPRS/GSM). Pour ce dernier point, ne gérer que les classes les plus hautes hiérarchiquement dans la collection (celles-ci fournissant une vue globale de la collection) permet de réduire le trafic. D'autre part, des stratégies de prefetching et de cache peuvent être mises en place en conjonction avec des adaptateurs de contenu pour différents types de terminaux, comme décrit dans les travaux sur l'*Universal Multimedia Access* [25]. Ces aspects sont importants du point de vue applicatif de nos travaux mais ne seront pas traités dans ce papier.

Ce papier est organisé comme suit : la section 2 présente un état de l'art sur les systèmes existants qui exploitent la date ou le lieu comme critère de structuration dans des collections d'images. Dans la section 3, nous présentons une technique d'organisation temporelle et spatiale d'une collection d'images, tout d'abord succintement, puis plus en détail. La section 4 fournit des résultats expérimentaux, tandis que la section 5 présente un type d'interface utilisateur exploitant les résultats. Enfin, la section 6 est consacrée aux conclusions et perspectives.

2 Etat de l'art

2.1 Date

Structurer une collection d'images en se basant sur la date de prise de vue de chaque photographie est intuitivement une idée intéressante, pratique et fiable. Comme noté dans [13], le processus de génération d'une image (c'est à dire le comportement des utilisateurs) est susceptible de comporter des groupes temporels, souvent de manière hiérarchique. Globalement, deux techniques différentes peuvent être distinguées. La première, détection de changement d'événements, illustrée par eg.[13, 22], possède l'avantage de ne pas utiliser de modèle paramétré particulier sur la distribution temporelle intra-classe. Cela est le cas dans [16] qui se base sur l'algorithme k-means. Cependant,

les problèmes délicats liés à la paramétrisation de la technique ne sont pas détaillés. Afin de résoudre le problème de l'initialisation de l'intervalle de temps intra-classe, les deux systèmes [16, 22] calculent d'une manière dynamique un temps moyen intra-classe caractérisant l'ensemble de la collection. Enfin, une autre technique, n'utilisant pas directement la date pour regrouper des images, a été récemment proposée dans [10] et consiste à combiner la date avec les informations sur la prise de vue et le contenu pour définir une mesure de similarité.

2.2 Géolocalisation

Récemment présentées dans [30], les techniques de géolocalisation sont progressivement intégrées dans les téléphones portables et les réseaux. Si en pratique la géolocalisation est principalement motivée sur le marché par les services liés à la localisation [17], nous pouvons cependant en tirer grand bénéfice.

L'importance de la localisation pour l'organisation d'une collection d'images est évoquée dans [13] mais, à notre connaissance, il y a pour le moment peu de systèmes qui ont considéré le problème en détail. Indépendamment de la recherche d'images mais toujours dans le but de fournir une interface de type calendrier électronique, nous avons proposé dans [12] une technique d'apprentissage non-supervisée permettant de trouver les lieux significatifs fréquentés par un utilisateur. La géolocalisation est mesurée en continue dans le temps et des partitions selon la date et le lieu sont extraites à différentes échelles, en se basant sur un modèle paramétrique de trajectoire. On tente ainsi de récupérer des épisodes de temps et des lieux significatifs. Un travail dans le même esprit est présenté dans [2], bien que le formalisme du modèle diffère.

3 Organisation spatiale et temporelle

3.1 Vue générale de l'approche proposée

L'objectif est de créer une représentation structurée d'une collection d'images, permettant à l'utilisateur de l'explorer facilement selon la date et le lieu. Dans cet article, nous considérons seulement les méta-données temporelles et spatiales attachées à chaque image, le contenu des images étant ignoré. De plus, nous souhaitons réaliser une classification de la manière la moins supervisée possible, en déterminant en ligne les limites et dimensions temporelles et spatiales des classes d'images, et le nombre de ces classes.

Nous formulons notre objectif comme un problème de classification non-supervisée basée sur un modèle de mélange. Dans notre cas, la génération des données suit une loi Gaussienne et la densité de probabilité des données est ainsi définie par :

$$p(D) = \sum_{k=1}^{K} \alpha_k \cdot p(x|\Theta_k), \qquad (1)$$

où la probabilité α_k représente la probabilité *a priori* des données générées à partir de la composante k, θ_k les paramètres de la composante k et K le nombre de composantes. Les paramètres θ_k d'une composante k sont définis par sa covariance Σ_k et son centre μ_k. On note l'ensemble des paramètres d'un modèle $\Theta = \{\theta_1, \ldots, \theta_K\}$.

Les principales idées de la technique proposée sont les suivantes :

- deux classifications distinctes sont construites, l'une temporelle et l'autre spatiale, à partir des méta-données des images.
- nous avons recours au critère de vraisemblance complétée intégrée (ICL), initialement proposé dans [3]. En effet, ce critère nous fournit une solution efficace pour déterminer la complexité du modèle, i.e. le nombre approprié de classes. D'un autre côté, il permet de favoriser les modèles associés à des classes bien distinctes, fournissant ainsi des partitions plus facilement exploitables pour notre objectif. Un avantage pratique, important pour notre application, est la robustesse de ce critère face aux erreurs de classification dues à la forme (gaussiennes) des composantes du modèle probabiliste vis à vis des classes recherchées.
- L'optimisation du critère ICL est réalisée avec l'algorithme Expectation-Maximisation (EM), avec une procédure de recherche dédiée (impliquant des tests sur l'initialisation du modèle et sur des fusions/divisions de composantes) .
- Les classifications trouvées, temporelle et spatiale, sont évaluées, en comparant le niveau de séparabilité des composantes : la meilleure classification est celle qui présente les groupes de données les plus distincts. La partition retenue fournit ainsi une structure pouvant être adaptée à une interface de type calendrier électronique permettant d'explorer la collection d'images.

3.2 Critère d'optimisation

La classification via un modèle de mélange est un cadre de travail classique et performant pour identifier les groupes pertinents dans un ensemble de données [9].

En prenant un point de vue bayesien, on peut montrer qu'une manière efficace d'évaluer la pertinence d'une classification pour expliquer des données D, en prenant en compte le besoin de comparer les hypothèses H_i avec des nombres de classes éventuellement différents, est fournie par la vraisemblance marginalisée :

$$P(D|H_i) = \int P(D|w_i, H_i) P(w_i|H_i) dw_i \qquad (2)$$

où w_i représente les paramètres du modèle associé à l'hypothèse H_i.
L'objectif est ainsi de déterminer le modèle de mélange fournissant la vraisemblance marginalisée maximale. Pour évaluer en pratique ce critère, plusieurs méthodes de calculs et approximations existent, détaillées dans [6]. Nous optons ici pour le critère d'information bayesien (BIC) [9], défini par :

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

Où ML est le logarithme de la vraisemblance maximisée, $N(K)$ est le nombre de paramètres indépendants dans le modèle composé de K classes et n est le nombre de données. Cette approximation représente en fait un critère de vraisemblance pénalisé par la complexité du modèle, illustrant l'utilisation de la vraisemblance marginalisée comme l'implémentation bayesienne du critère d'Occam.
En général, le critère BIC sert à identifier les paramètres du modèle, et le nombre de classes. Néanmoins, la forme paramétrique imposée aux classes (en pratique, gaussienne) entraîne parfois de mauvaises segmentations et/ou un nombre surévalué de composantes, dans le cas où les données ne peuvent être correctement décrites par un mélange gaussien. Dans notre cas, l'hypothèse de gaussianité est souvent mise en défaut.
Afin d'améliorer ce point, nous optons pour l'utilisation du critère de vraisemblance complétée intégrée (ICL), proposé dans [3]. Ce critère est défini par :

$$ICL = BIC - \Phi(K) \quad (4)$$

où $\Phi(K)$ est l'entropie du modèle de mélange, défini par :

$$\Phi(K) = -\sum_{k=1}^{K}\sum_{i=1}^{n} t_{ik} \cdot log(t_{ik}) \geq 0 \quad (5)$$

où K est le nombre de composantes du modèle de mélange, n le nombre de données, et t_{ik} la probabilité à posteriori d'une observation i pour la composante k. Les valeurs t_{ik} sont calculées pendant la phase d'optimisation, qui est décrite dans la section 3.4.
Le critère ICL pénalise le critère BIC avec l'entropie calculée à partir de l'affectation des données au modèle dans le mélange, i.e. un mélange dont les données sont bien séparées a une entropie faible et est ainsi favorisé.

3.3 Traitement du problème des petits échantillons

Dans notre contexte, le nombre d'éléments dans une classe pouvant être faible, il est parfois difficile de réaliser une estimation fiable des matrices de covariance décrivant la dispersion des données intra-classes. On se restreint ici à la classification spatiale, le cas de la classification temporelle étant similaire.
Plusieurs techniques de régularisation pour l'estimation de matrices de covariance sont présentées dans [28], définissant la matrice régularisée comme une combinaison linéaire d'un échantillon de la matrice estimée et d'autres formes plus contraintes.
La solution que nous proposons consiste à estimer une matrice de covariance pleine en imposant deux contraintes :

- une "surface minimale" est imposée à chaque composante en fixant des valeurs propres minimales de la matrice de covariance. Ceci permet de régler efficacement le problème des composantes ne contenant qu'une seule observation dans la classification, en évitant qu'une composante ait une variance nulle. En outre cette technique permettrait de prendre en compte le bruit des données spatiales, si on en avait connaissance (la fourniture de cette information est prévue dans la standardisation de la géolocalisation sur mobiles).
- un rapport minimum est imposé entre les deux valeurs propres de la matrice de covariance. Une matrice de covariance pleine est plus flexible et préférable à une matrice diagonale, étant donné que les orientations des axes sont libres. Mais avec seulement deux observations (ou plus, si elles sont alignées), les deux valeurs propres présentent des grandeurs différentes et les observations ont des probabilités d'appartenance aux composantes excessives.

Soit S la matrice diagonale de covariance d'une composante telle que $S = U \cdot \Lambda \cdot U^{-1}$ avec $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, où $\lambda_1 > \lambda_2$ et U est la matrice composée des vecteurs propres de S.
La matrice de covariance de chaque composante est donc modifiée comme suit :

$$\Sigma = U \cdot \begin{bmatrix} max(\beta', \lambda_1) & 0 \\ 0 & max(\beta', \lambda_2, \beta\lambda_1) \end{bmatrix} \cdot U^{-1} \quad (6)$$

où β' et β sont des paramètres. α dépend du bruit des données (ici, les données spatiales), tandis que β est initialisé à 0.3.
Un autre problème, lié au manque de données dans une classe, est que l'approximation du critère BIC n'est plus valide. Plusieurs expériences ont été réalisées avec un autre critère adapté aux petits échantillons (AICc[27]), mais aucune différence significative n'a été trouvée.
Dans la section suivante, on présente une technique d'optimisation du critère ICL.

3.4 Recherche de la meilleure classification

L'algorithme Expectation-Maximization (EM) [7, 5] permet d'optimiser localement le critère défini ci-dessus. Il est très couramment utilisé en association

avec des modèles de mélange. Il se décompose en deux étapes, l'étape E, dans laquelle les probabilités d'affectation des données à la composante sont estimées conditionnellement aux paramètres des modèles, et l'étape M, dans laquelle les paramètres du modèle sont estimés en se basant sur l'estimation courante des affectations des données aux modèles. Cette approche est retenue, entre autres, parce qu'elle permettra, naturellement, de réaliser une extension incrémentale de la technique exposée dans cet article.

Nous décrivons ci-dessous plusieurs problèmes posés par l'algorithme EM, et les solutions proposées dans notre contribution.

Cet algorithme prenant en paramètre le nombre de composantes, on décide de tester plusieurs sous-espaces d'hypothèses associés à un nombre donné de composantes. Parmi toutes les solutions trouvées, on retient la solution qui maximise le critère ICL. Une amélioration de cette recherche exhaustive est proposée dans [8]. Cependant, désirant afficher la partition trouvée sur un écran de petite taille, le nombre maximum de classes pour segmenter les données est limité. Nous effectuons donc une recherche exhaustive comprise entre 2 et 15 composantes (la classification avec une seule composante est considérée mais ne nécessite pas d'optimisation, uniquement le calcul du critère ICL). Ce choix est aussi motivé par l'intérêt de pouvoir construire une classification multi-échelle, i.e. trouver plusieurs partitions plausibles avec des complexités différentes, afin de permettre une exploration hiérarchique. Ceci est l'objet de travaux en cours.

Le second problème est dû au caractère local de l'optimum trouvé par l'algorithme EM. Deux techniques complémentaires sont utilisées pour améliorer ce point :

- l'initialisation exploite une procédure efficace inspirée de la technique "em-EM" [4]. Cette procédure, décrite dans l'algorithme 1 ci-dessous, affecte à la fois les formes des modèles initiaux (CEM au lieu du k-means classique) et la stratégie de recherches aléatoires.
- un algorithme Split and Merge, issu de [29], qui est appliqué après une optimisation avec l'algorithme "em-EM". Cette technique permet d'évaluer et d'appliquer des divisions/fusions de composantes, et ainsi de rechercher un meilleur optimum local du critère ICL en faisant varier les paramètres du modèle. Il faut noter qu'à chaque fois qu'une division de composantes est réalisée, deux autres composantes sont fusionnées. Ainsi le nombre de composantes reste constant. Cet algorithme est décrit dans la suite.

L'algorithme SMEM [29] permet d'améliorer le résultat d'un modèle optimisé en tentant plusieurs successions de fusion/division de composantes. Cela peut aider à éviter un minimum local du critère optimisé. Cet algorithme est appliqué après l'optimisation du modèle avec la stratégie de recherche "em-EM". A chaque itération de l'algorithme SMEM, 3 composantes i, j et k sont choisies. Des "sauts" semi-locaux sont réalisés dans l'espace de recherche, en fusionnant i et j et divisant k. Le nouveau modèle est alors retenu si le critère ICL est amélioré.

Algorithm 1 Stratégie "em-EM"

for i=1 to NB_INITIALIZATIONS **do**
1. Affectation aléatoire des données au modèle.
2. Itération de l'algorithme Classification EM (CEM est une version de l'algorithme EM où l'affectation des données est réalisée en dur, et dont la convergence est plus rapide)
3. Exécution d'un petit nombre d'itérations de l'algorithme EM (max. 20 itérations)

end for

Itération de l'algorithme EM (plus de 500 itérations, afin de converger), initialisé à partir de la meilleure solution (meilleure optimisation du critère ICL) trouvée à l'étape précédente.

Critères de fusion et division :

Le critère de fusion que l'on propose se base sur la distance de Mahalanobis. On fusionne en priorité les composantes ayant une distance de Mahalanobis faible deux à deux :

$$J_{merge}(i,j,\Theta) = D(i,j), \tag{7}$$

où $D = min\{dist_{Mah}(\mu_i,\Sigma_i,\mu_j), dist_{Mah}(\mu_j,\Sigma_j,\mu_i)\}$, et $dist_{Mah}(\mu_j,\Sigma_j,\mu_i) = (\mu_i-\mu_j\}^T \cdot \Sigma_j^{-1} \cdot (\mu_i-\mu_j)$. Notre critère de division se base sur l'entropie de chaque composante. Une composante avec une forte entropie suggère qu'elle n'est pas pertinente par rapport à ses données associées. Nous divisons donc en priorité les composantes ayant une forte entropie :

$$J_{split}(k,\Theta) = \sum_{i=1}^{n} t_{ik} \cdot log(t_{ik}) \tag{8}$$

Initialisation des nouveaux paramètres :

L'initialisation des paramètres du nouveau modèle obtenu après une fusion et une division se base sur les paramètres courant Θ^*. Les paramètres initiaux pour deux composantes fusionnées sont définis par :

$$\alpha_{i'} = \alpha_i^* + \alpha_j^* \quad \text{and} \quad \theta_{i'} = \frac{\alpha_i^*\theta_i^* + \alpha_j^*\theta_j^*}{\alpha_i^* + \alpha_j^*} \tag{9}$$

Pour la division d'une composante k en deux composantes j' et k' :

$$\alpha_{j'} = \alpha_{k'} = \frac{\alpha_k^*}{2}, \quad \mu_{j'} = \mu_k^* + \epsilon, \quad \mu_{k'} = \mu_k^* + \epsilon' \quad et$$
$$\Sigma_{j'} = \Sigma_{k'} = det(\Sigma_k^*)^{(1/d)}/I_d, \tag{10}$$

Où ϵ et ϵ' représentent une faible variation, $det(\Sigma)$ est le déterminant de Σ et I_d est une matrice identitée de dimension d.
L'algorithme SMEM consiste ainsi à calculer une liste de candidats à la fusion et division et ensuite pour le premier candidat, d'initialiser les paramètres et d'optimiser le nouveau modèle avec l'algorithme EM. Si le critère ICL est amélioré, ce nouveau modèle est conservé. Dans le cas où il est rejeté, le candidat suivant dans la liste est testé. L'algorithme SMEM est résumé par l'algorithme 2. Il est appliqué pour chaque hypothèse H_i, après convergence de l'algorithme EM.

Algorithm 2 Algorithme Split and Merge

1. Itération de l'algorithme EM à partir de paramètres initiaux Θ jusqu'à convergence. On note Θ^* et Q^* respectivement les paramètres estimés et la valeur du critère optimisé après convergence de l'algorithme EM.
2. Classement des candidats à la fusion et division en calculant les critères de fusion et division en se basant sur Θ^*. On note $\{i, j, k\}_c$ le c-ème candidat.

for $c = 1, \ldots, C_{max}$ **do**
Après l'initialisation des nouveaux paramètres basé sur Θ^*, on applique l'algorithme EM jusqu'à convergence. Soit Θ^{**} les nouveaux paramètres obtenus et Q^{**} la nouvelle valeur du critère optimisé. Si $Q^{**} > Q^*$ alors $Q^* \leftarrow Q^{**}$, $\Theta^* \leftarrow \Theta^{**}$ et on retourne à 2.
end for

3.5 Comparaison des classifications temporelles et spatiales

Rappelons que l'on construit deux classifications indépendantes, l'une temporelle et l'autre spatiale. Il est possible qu'en raison d'un manque of structure nette dans les données ou à une insuffisance du critère d'optimalité (un minimum local médiocre), une ou l'ensemble des partitions obtenues puisse être occasionnellement de mauvaise qualité. Nous proposons ainsi de sélectionner la classification ayant une valeur d'entropie $\Phi(K)$ minimale, i.e. la classification présentant la meilleure séparabilité. Cela permet généralement de sélectionner la classification la plus pertinente et la plus fiable pour explorer la collection d'images.

4 Résultats expérimentaux

4.1 Voyage en Loire-Atlantique

Afin d'illustrer les variétés de situations que notre technique peut supporter, le résultat de deux expériences est présenté par la figure 1 (en fin d'article). On se focalise ici sur les classifications spatiales. Ces deux scénarios, situés en Loire-atlantique,

Dans le premier cas (fig. 1a), il existe un nombre important de zones pertinentes : 11 composantes ont été trouvées (incluant une composante très peu gaussienne sur la côte). Dans le second scénario, les lieux de prises de vue sont dispersées et présentent une hiérarchie à deux niveaux : sur l'ensemble de la région et plus localement aux alentours de Nantes. La classification obtenue identifie correctement cette hiérarchie, ce qui présente un avantage important pour parcourir cette collection d'images.

4.2 20 jours de vacances

Ces expériences correspondent à deux voyages A et B, durant lesquels l'utilisateur s'est promené dans différents lieux.

Pendant le voyage A, l'utilisateur a pris 300 images sur une période de dix jours. Les figures 2 et 3 représentent les résultats obtenus, respectivement pour la classification temporelle et la classification spatiale. Nous avons testé les modèles comprenant entre 1 et 15 composantes. La classification temporelle retenue est composée de 14 composantes et présente un résultat acceptable. Nos contraintes sur les matrices de covariances sont effectives puisque nous obtenons des composantes associées à une seule donnée. On note un problème de sur-segmentation dans l'intervalle $[9000, 11000]$, qui présente une mauvaise séparabilité : 3 composantes auraient été plus adéquates. Néanmoins, tous les épisodes temporelles pertinents sont identifiés. La classification spatiale est composée de 12 composantes et les différents lieux sont bien mis en valeurs. On peut juste noter des composantes un peu trop larges (les composantes 1, 2, 7 et 12), qui sont dues à un manque de données dans la région. Le niveau de séparabilité des données est évalué par leur entropie $\Phi_{temporelle} = 12.46$ et $\Phi_{spatiale} = 7.45$. C'est donc la classification spatiale qui est sélectionnée.

Pendant le voyage B d'une durée de 10 jours, l'utilisateur a pris 500 images. Nous avons testé les modèles comprenant entre 1 et 15 composantes. 14 et 8 composantes ont été respectivement trouvées pour la classification temporelle (fig. 4) et la classification spatiale (fig. 5). La classification temporelle obtenue est satisfaisante puisque tous les épisodes temporelles sont correctement déterminés. La partition est composée de nombreuses classes de petites tailles, présentant une bonne séparabilité. Pour la classification spatiale, les lieux pertinents sont identifiés de manière satisfaisante. Bien que la classification obtenue aux alentours des latitudes $[0, 4000]$ soit discutable, elle fournit un résultat utilisable. Cette division est certainement due au caractère très peu gaussien des données. Enfin, la séparabilité des données est évaluée par leur entropie $\Phi_{temporelle} = 31.45$ et $\Phi_{spatiale} = 34.12$. La forte valeur de l'entropie spatiale est due à la division "discutable" citée précédemment et favorise ainsi la classifi-

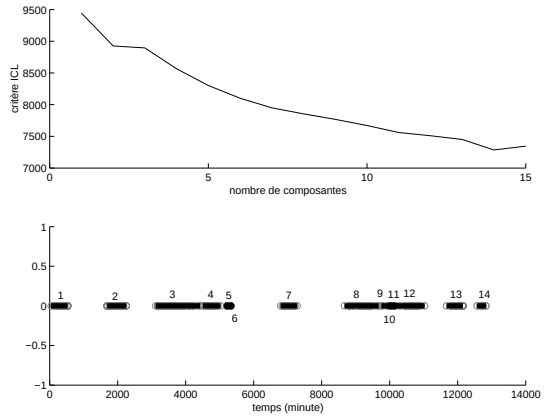

FIG. 2 – Classification temporelle du voyage A : au dessus, le critère ICL obtenu pour chaque complexité de modèle et en dessous, la classification présentant la meilleure optimisation du critère ICL. Les lignes noires représentent les classes et chaque composante est numérotée.

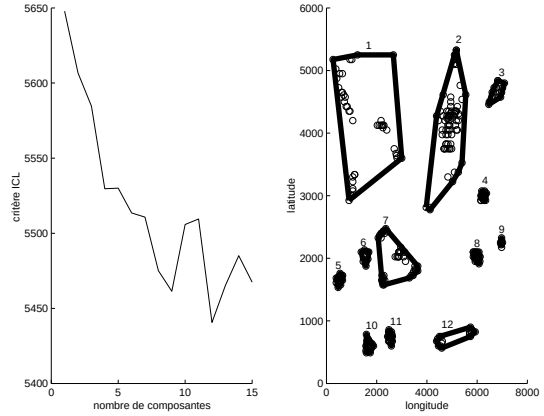

FIG. 3 – Classification spatiale du voyage A : sur la gauche, le critère d'optimisation ICL obtenu pour chaque complexité de modèle et sur la droite, la classification présentant la meilleure optimisation du critère ICL. Les classes sont numérotées et leurs enveloppes convexes sont dessinées en noires.

cation temporelle.
On peut déduire de ces expériences qu'il serait judicieux de prendre en compte l'aspect séquentiel des données géographiques.

4.3 Comparaison des critères BIC et ICL

Dans cette expérience, nous illustrons l'avantage du critère ICL par rapport au critère BIC. Pour cela, nous avons sélectionné des images prises lors d'une journée de balade. Les classifications obtenues avec les critères ICL et BIC sont représentées respectivement par les figures 6(a) et 6(b). Dans cette exemple, la classe E ne peut être approximée correctement par une distribu-

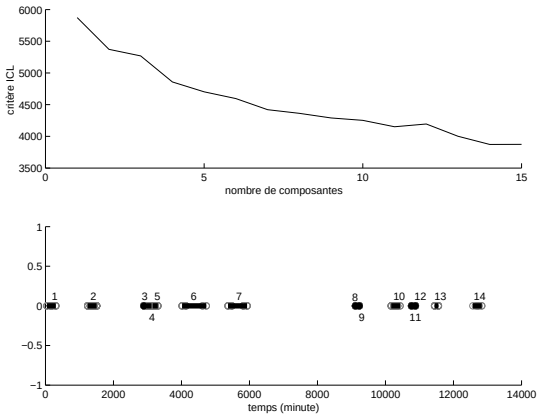

FIG. 4 – Classification temporelle du voyage B : au dessus, le critère ICL obtenu pour chaque complexité de modèle et en dessous, la classification présentant la meilleure optimisation du critère ICL. Les lignes noires représentent les classes et chaque composante est numérotée.

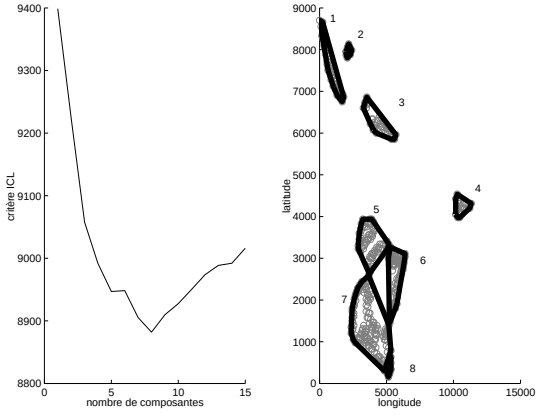

FIG. 5 – Classification spatiale du voyage B : sur la gauche, le critère d'optimisation ICL obtenu pour chaque complexité de modèle et sur la droite, la classification présentant la meilleure optimisation du critère ICL. Les classes sont numérotées et leurs enveloppes convexes sont dessinées en noires.

tion gaussienne. En conséquence, le critère BIC segmente ces données en deux classes alors que le critère ICL regroupe ces données en une seule classe grâce à la pénalisation entropique. Ce test met en valeur la plus grande robustesse du critère ICL face au problème du caractère non-gaussien des données.

Bien que ces expériences se basent sur un jeu de données correspondant à une dizaine de journées de prises de vues plutôt que plusieurs mois ou plusieurs semaines, elles fournissent une première validation de la technique que nous proposons.

5 Application à l'exploration d'une collection d'images

Dans ce paragraphe, nous illustrons l'aspect pratique de la technique proposée, en proposant un exemple d'interaction homme-machine permettant d'explorer une collection d'images sur un téléphone mobile ou un PDA.

Nous considérons ici une classification temporelle, identique à celles trouvées sur les calendriers électroniques disponibles sur les PDAs. Les divisions de temps proposées pour explorer les images sont soit réalisées selon la classification temporelles, soit spatiales, ceci dépendant de la classification choisie. Dans le cas où la classification spatiale est sélectionnée, les groupes sont temporellement déconnectés. La restriction à un critère de représentation assure un nombre limité d'épisodes et la cohérence du critère de division avec la représentation. Le choix des images représentatives d'un groupe n'est pas ici traité, quelques solutions sont proposées dans les travaux référencés dans la section 2.

On peut aussi exploiter simultanément les classifications temporelles et spatiales en les fusionnant sur un même axe temporelle. La figure 7 illustre un exemple d'interface d'un calendrier électronique. Des couleurs de fond ou des limites de séparation différentes entre les groupes d'images indiquent sur quels critères est basée la classification. Des boutons peuvent être associés aux fonctions suivantes "saut jusqu'aux prochaines images dans cette zone temporelle", "saut jusqu'aux prochaines images dans cette zone spatiale".

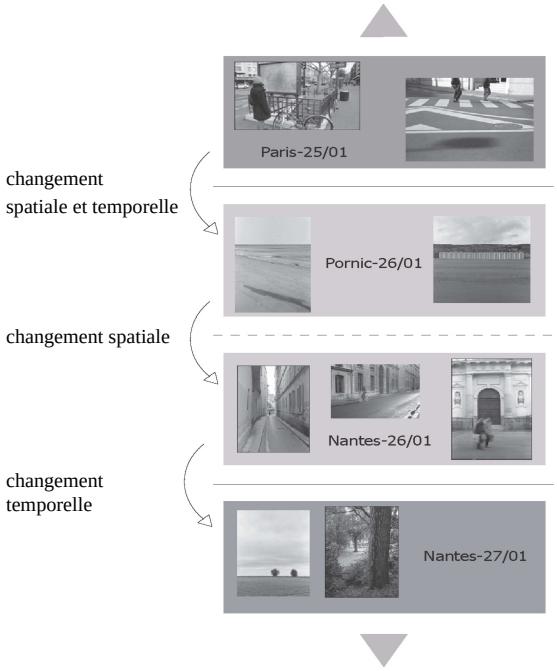

FIG. 7 – Exemple d'interface de calendrier électronique : les lignes en pointillées représentent un changement de lieu, et celles en continues un changement de date ou les deux. L'assignement des noms des classes est réalisé manuellement et n'est pas lié à la technique proposée.

6 Conclusion

Dans ce papier, nous nous sommes focalisés sur le problème d'organisation de collection d'images personnelles. L'intérêt d'une classification temporelle et spatiale basée sur les méta-données est mis en avant, notamment pour les appareils photographiques inclus dans les téléphones portables. Nous proposons alors une technique de classification non-supervisée basée sur le critère ICL, satisfaisant plusieurs conditions de l'application (nombre inconnu de composantes, données non gaussiennes), optimisée avec les algorithmes EM et SMEM. Le critère d'entropie est ensuite utilisé pour sélectionner la classification la plus pertinente entre la partition temporelle et spatiale. Parmi les qualités du formalisme, on peut citer l'absence de paramètres critiques et les perspectives de version incrémentale et de classification multi-échelle. En conclusion, cette technique semble efficace et présente une direction réaliste pour l'organisation de collection d'images, fournissant une structure pouvant permettre plusieurs types d'exploration.

Références

[1] P. Anandan. Personal digital media : It's about sharing experiences. In *MMCBIR 2001 - Multimedia Content-based Indexing and Retrieval*, INRIA Rocquencourt, France, Septembre 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, Octobre 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, Juillet 2000.

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

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

[6] D.M. Chickering and D. Heckerman. Efficient approximations for the marginal likelihood of Baye-

sian networks with hidden variables. Technical Report MSR-TR-96-08, Microsoft Research, Mars 1996.

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

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

[9] Chris Fraley and Adrian E. Raftery. How many clusters ? Which clustering method ? Answers via model-based cluster analysis. *The Computer Journal*, 41(8) :578–588, 1998.

[10] U. Gargi, Y. Deng, and D. R. Tretter. Managing and searching personal photo collections. Technical Report HPL-2002-67, HP Laboratories, Palo Alto, Mars 2002.

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

[12] 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, Août 2002.

[13] 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, Juin 2002.

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

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

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

[17] N. Marmasse and C. Schmandt. Location-aware information delivery with commotion. In *Handheld and Ubiquitous Computing HUC'2000, Second Int. Symp.*, pages 157–171, Bristol, UK, Septembre 2000.

[18] T. J. Mills, D. Pye, D. Sinclair, and K. R. Wood. Shoebox : A digital photo management system. Technical Report MSR 2000.10, AT&T Labs., Cambridge, England, 2000.

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

[20] A. Myka, J. Yrjänäinen, and M. Gelgon. Enhanced storing of personal content. Finnish Patent PCT/FI02/00277, Nokia Research Center, Nokia corp., Finland, Avril 2002.

[21] Y. Neuvo and J. Yrjanainen. Wireless meets multimedia. *Wireless communications and mobile computing*, 2 :553–562, Septembre 2002.

[22] J. C. Platt and B. A. Field M. Czerwinski. PhotoTOC : Automatic clustering for browsing personal photographs. Technical Report MSR-TR-2002-17, Microsoft Research, Février 2002.

[23] K. Rodden. How do people manage their digital photographs ? In *ACM Conference on Human Factors in Computing Systems*, pages 409 – 416, Fort Lauderdale, Avril 2003.

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

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

[26] A. Soronen and V. Tuomisto. Mobile image messaging - anticipating the outlines of the usage culture. In *4th Int. Symp. on Mobile HCI'2002*, pages 359–363, Pisa, Italy, Septembre 2002.

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

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

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

[30] Y. Zhao. Standardization of mobile phone positioning for 3G systems. *IEEE Communications Magazine*, pages 108–116, Juin 2002.

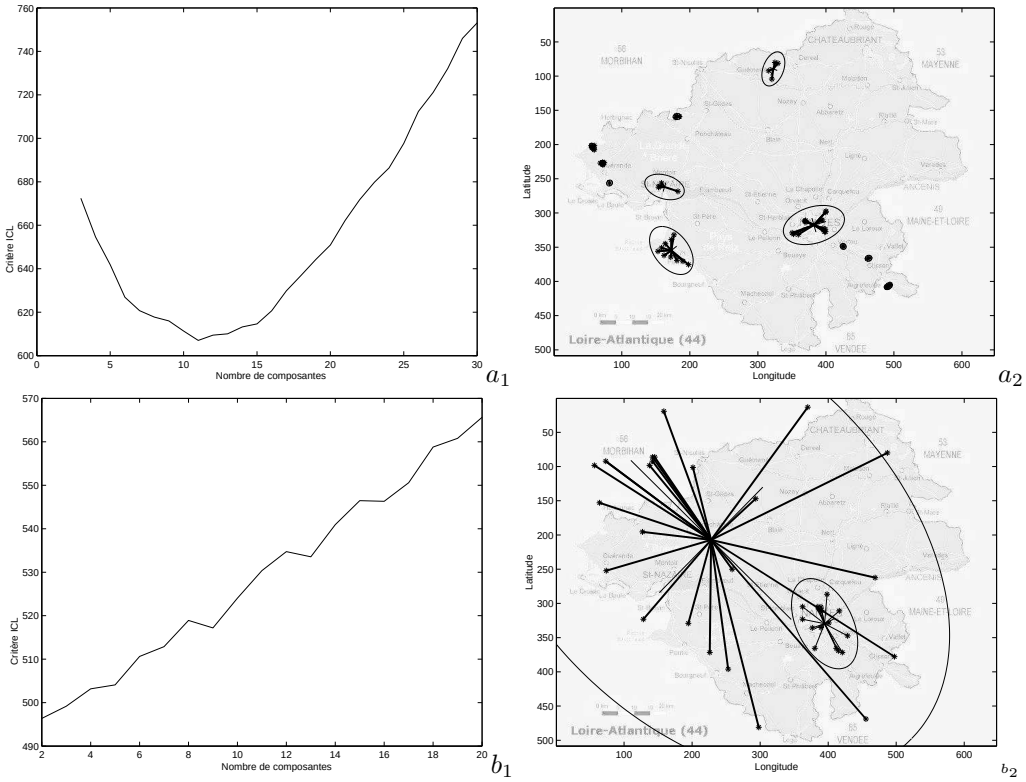

FIG. 1 – Classification de deux jeux de données aux propriétés particulières : dans (a), les images sont regroupées par petit groupe dans différents lieux tandis que dans (b), elles sont dispersées et présentent une hiérarchie à deux niveaux. Les ellipses représentent les variances des composantes et chaque donnée est reliée au centre de sa composante par une ligne. Les valeurs optimales du critère ICL sont respectivement présentées par les figures a_1 et b_1.

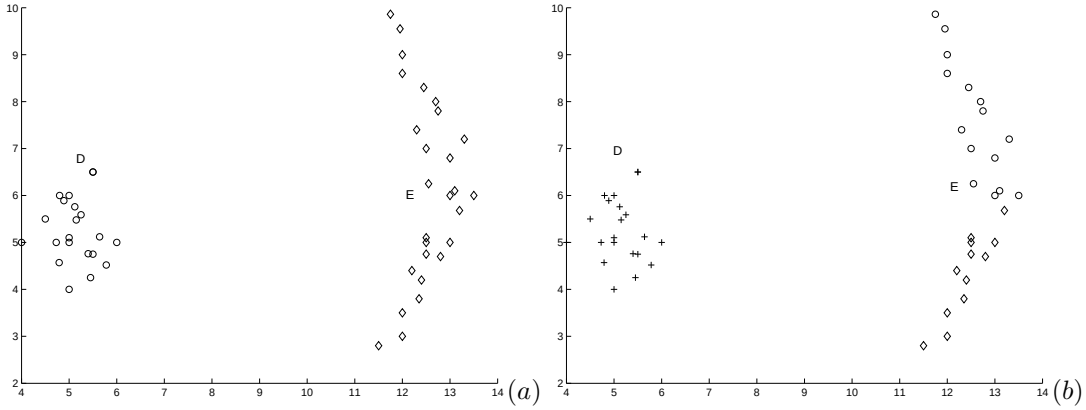

FIG. 6 – Classification en deux dimensions avec le critère BIC (a) et le critère ICL (b) : la classification (a) est composée de trois classes ($\circ$, $\diamond$ and $+$) , et la classification (b) de deux classes ($\circ$ et $\diamond$) .