

Models and inference for structured stochastic systems

Florence Forbes

▶ **To cite this version:**

Florence Forbes. Models and inference for structured stochastic systems. Modeling and Simulation. Université de Grenoble, 2010. tel-00578938

HAL Id: tel-00578938
https://theses.hal.science/tel-00578938

Submitted on 22 Mar 2011

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

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

Université Joseph Fourier

Ecole Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique

Modèles et inférence pour des systèmes stochastiques structurés

Models and inference for structured stochastic systems

Mémoire d'habilitation présenté par

Florence Forbes

pour l'obtention du diplôme

d'habilitation à diriger des recherches

Spécialité Mathématiques Appliquées

présenté et soutenu publiquement le 7 décembre 2010 devant

Le JURY :

Rapporteurs :	**Geoff. MacLachlan**	Prof. University of Queensland, Brisbane, Australia
	Eric Moulines	Prof. Institut Télécom / Télécom ParisTech, Paris, France
	Fionn Murtagh	Prof. University of London, London, England
Examinateurs :	**Olivier François**	Prof. Institut National Polytechnique de Grenoble, France
	Fabrice Heitz	Prof. Université de Strasbourg, Ecole Nationale Supérieure de Physique de Strasbourg, France
	Jerôme Idier	Directeur de Recherche, CNRS, Institut de Recherche en Communications et Cybernétique de Nantes, France
Président :	**Anestis Antoniadis**	Prof. Université Joseph Fourier, Grenoble, France

Résumé

Le contexte de mon travail est la mise au point d'outils statistiques pour le dévelopement et l'analyse de modèles stochastiques structurés. L'idée sous-jacente à la notion de structure est qu'il est souvent possible à l'aide d'hypothèses locales simples combinées de manière cohérente de rendre compte de phénomènes globaux potentiellement complexes. Cette idée de construction du local vers le global guide ainsi la modélisation, l'estimation et l'interprétation. Cette approche se révèle utile dans des domaines variés tels que le traitement du signal et de l'image, les neurosciences, la génomique, l'épidémiologie, etc. Inversement les besoins de ces domaines ont pu susciter en retour des dévelopements théoriques importants.

Par ailleurs, beaucoup de techniques statistiques sont encore limitées par des d'hypothèses restrictives pouvant conduire à des analyses imprécises voire erronées. Différentes sources de complexité peuvent mettre en défaut les approches classiques. Souvent les données exhibent une structure de dépendance non triviale, due par exemple à des répétitions, des groupements, des méthodes d'échantillonage particulières, des associations dans l'espace ou le temps. Une seconde source de complexité est liée au processus de mesure qui peut impliquer l'utilisation d'instruments physiquement très différents, qui produisent des données hétérogènes, en grandes dimensions et potentiellement de manière défaillante de sorte qu'une partie des données peut être manquante.

La plupart de mes objectifs de recherche sont centrés sur la mise au point de modèles et d'outils d'inférence pouvant faire face à ce genre de complications fréquentes dans les données modernes et contribuer ainsi au dévelopement de nouvelles méthodes statistiques.

En ce qui concerne la notion de dépendance et de localité, un concept central est celui d'indépendance conditionnelle. Les propriétés de Markov et les modéles markoviens permettent d'énoncer de telles indépendances conditionnelles et ce thème est central dans ma recherche.

Pour ce qui est des données manquantes ou incomplètes, les modèles de mélanges sont une approche classique. Ces modèles conduisent plus généralement à la notion de modèles à structure manquantes. Ces derniers sont également utiles pour rendre compte d'hétérogénités dans les données. Ils trouvent de nombreux échos en statistique : modéles de mélanges finis, modèles de Markov cachés, modèles à effet aléatoire, etc. La présence de données incomplètes induit cependant généralement des difficultés pour ce qui est de l'estimation des paramètres et de l'évaluation des performances.

Modèles markoviens et modèles de mélanges sont mes deux principaux thèmes de recherche avec cette idée unificatrice de structure dans les modèles mais aussi dans les données. J'ai pu montrer que ces deux thèmes pouvaient être reliés utilement en traitant des problèmes difficiles dans diverses applications.

Plus précisément, j'ai developé des modèles à structure cachée essentiellement dans le but de résoudre des problèmes de classifications inhérents à certaines questions. J'ai souvent abordé le problème de l'estimation de ces modèles à partir de l'algorithme EM et developpé des variantes permettant d'apporter des solutions satisfaisantes lorsque les outils classiques faisaient défaut. J'ai tenté également d'apporter des résultats sur

les propriétés théoriques, e.g. convergence et vitesse, de ces algorithmes. Enfin, j'ai abordé la question de la sélection de modèles essentiellement en cherchant à proposer des critères de sélection dans les cas où les critères classiques n'étaient pas calculables.

Table des matières

Research Activities **7**
1 Introduction . . . 9
2 Scientific foundations . . . 11
2.1 Mixture models . . . 11
2.2 Markov models . . . 12
3 Domains of research . . . 14
3.1 Clustering with hidden structure models . . . 15
3.2 The design and study of EM-like algorithms . . . 37
3.3 Hidden structure model selection . . . 45
4 Application domains . . . 49
4.1 Image Analysis . . . 49
4.2 Biology and Computational Neuroscience . . . 63

Technology development **73**
1 Software . . . 75
2 Industrial transfer of research result . . . 76

Summary of main contributions **81**
1 Variational approximations for inference in hidden Markov random fields 83
2 Models for complex and structured data . . . 85
3 Model-based clustering for the analysis of MR brain scans . . . 88

Research Program **91**
1 Bayesian conditional modeling for multimodal data analysis . . . 93
2 Extreme value theory and structured data . . . 94
3 Hybrid inference algorithms . . . 95
4 Statistical image processing and cognition . . . 95
5 Application domains and technology transfer . . . 96
6 International collaborations . . . 96

Detailed Curriculum Vitae **97**
1 Diplomas . . . 99
2 Professional history . . . 99
3 Prizes and awards . . . 100
4 Supervision of research activities . . . 100
5 Responsibilities . . . 104
6 Management . . . 105
7 Collaborations, visits . . . 109
8 Teaching . . . 111
9 Dissemination of scientific knowledge . . . 112
10 Other relevant information . . . 112

Publications **113**
1 Articles in international refereed journals and book chapters . . . 115
2 Articles in national refereed journals . . . 116
3 Publications in international Conferences and Workshops . . . 116
4 Publications in national Conferences and Workshops . . . 118
5 Internal Reports . . . 119
6 Others . . . 120

Research Activities

1 Introduction

The context of my work is the analysis of structured stochastic models with statistical tools. The idea underlying such models is that stochastic systems that exhibit great complexity can be accounted for by combining simple local assumptions in a coherent way. This provides a key to modelling, computation, inference and interpretation. Such stochastic models have found applications in areas as diverse as signal and image processing, neuroscience, genetics and epidemiology. The needs of these areas have in turn stimulated important theoretical developments.

However, these powerful and flexible techniques can still be restricted by necessary simplifying assumptions, such as precise measurement and independence between observations, and it long ago became clear that in many areas such assumptions can be both influential and misleading. Also there are several generic sources of complexity in data that require methods beyond the commonly-understood tools in mainstream statistical packages.

Often data exhibit complex dependence structures, having to do for example with repeated measurements on individual items, or natural grouping of individual observations due to the method of sampling, spatial or temporal association, family relationship, and so on.

Other sources of complexity are connected with the measurement process, such as having multiple measuring instruments or simulations generating high dimensional and heterogeneous data or such that data are dropped out or missing.

Such complications in data-generating processes raise a number of challenges when dealing with modern data. My goal is to contribute to statistical modelling by offering theoretical concepts and computational tools to handle properly some of these issues.

As regards dependencies and locality, a central part is played by the concept of conditional independence. It provides a precise description of the information conveyed by the value of one variable about others in a statistical model. Markov properties are statements about conditional independence assumptions and *Markov models* are the central subject of my research. The concept of conditional independence, whereby each variable is related locally (conditionally) to only a few other variables, is the key to both the construction and analysis of such models.

When dealing with missing data, *mixture models* are a central starting point. They lead naturally to more general hidden structure models. Hidden structure models are also useful for taking into account heterogeneity in data. They concern many areas of statistical methodology (finite mixture analysis, hidden Markov models, random effect models, ...). Due to their missing data structure, they induce specific difficulties for both estimating the model parameters and assessing performance.

My two main domains of research are Markov models and mixture models. The main particularity of my work is the focus on the key idea of structure in models and data. This focus is unifying and promising in its ability to generalize the ingredients of the models, to broaden the scope of applications and to allow cross-fertilization between different areas. Besides, various successful applications illustrate how I managed to

combine my two main domains of expertise.

More specifically, I designed and studied a number of models with hidden data structure. The main goal is usually to address clustering tasks that appear in various challenging practical issues. For inference, I have mainly focused on the EM algorithm principle and provided appropriate variants able to capture the main features of the designed models and to solve the targeted tasks satisfyingly. When accessible, I have provided insights on their theoretical properties such as their convergence and speed. I also investigated the issue of model selection. More specifically, I am interested in providing, for the models I consider, ways to use *standard* selection criteria when their computation is not straightforward.

My research activity is structured around the three following lines :

1. Clustering with hidden structure models : Section 3.1.
2. The design and study of EM-like algorithms : Section 3.2
3. Hidden structure model selection : Section 3.3.

Current application domains are mentioned in Section 4.

2 Scientific foundations

A general presentation of my main domains of expertise, in this section, is followed by more specific research topics in Section 3.

2.1 Mixture models

Keywords : *missing data, mixture of distributions, EM algorithm, statistical pattern recognition, clustering, unsupervised and partially supervised learning.*

Mixture models and more specifically Gaussian mixture models are among the most statistically mature methods for clustering and are also used intensively for density estimation. The success of mixture models lies partly in the fact that clustering can be seen as a labelling problem and therefore corresponds to many problems in practice.

A labelling problem is specified in terms of a set of sites S and a set of labels $\mathcal{L}$. A site often represents an item, a point or a region such as an image pixel or an image feature such as a corner, a line segment or a surface patch. A set of sites may be categorized in terms of their regularity. Sites on a lattice are considered as spatially regular (*e.g.* the pixels of a 2D image). Sites which do not present spatial regularity are considered as irregular. This is a common case when sites represent geographic locations, features extracted from images at a more abstract level, such as the detected corners and lines and more generally *interest points*[1]. It can also be that the sites correspond to items (*e.g.* genes) that are related to each other through a distance or dissimilarity measure or simply to a collection of independent items.

A label is an event that may happen to a site. I have been working mainly with discrete label sets. In this case, a label assumes a discrete value in a set $\mathcal{L}$ of K labels. In edge detection, for example, the label set is the two component set $\{edge, non-edge\}$.

The labelling problem is to assign a label from a label set $\mathcal{L}$ to each of the sites. If there are N sites, the set $\mathbf{z} = \{z_1, \ldots, z_N\}$ with $z_i \in \mathcal{L}$ for all $i \in S$, is called a labelling of the sites in S in terms of the labels in $\mathcal{L}$. When each site is assigned a unique label, a labelling can be regarded as a function with domain S and image $\mathcal{L}$. In mathematical programming a labelling is also called a coloring, in the terminology of random fields it is called a configuration. In vision, it can corresponds to an edge map, an interpretation of image features in terms of object features, or a pose transformation and so on.

The approach of the labelling problem considered in my work is based on mixture models and more generally on hidden structure models. They correspond to a form of clustering in which we assume that individual data points are generated by first choosing one of a set of multivariate distributions (typically Gaussian) and then sampling from it. Data is assumed to naturally divide into observed data $\mathbf{y} = \{y_1, \ldots, y_N\}$ and unobserved or missing data $\mathbf{z} = \{z_1, \ldots, z_N\}$. The missing data z_i represents the memberships to one of a set of K alternative categories, *i.e.* the labels. The distribution of

[1] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2) :91–110, 2004.

an observed y_i can be written as a finite mixture of distributions,

$$f(y_i \; ; \; \theta) = \sum_{k=1}^{K} P(z_i = k \; ; \; \theta) \; f(y_i \mid z_i; \theta) \; . \tag{1}$$

I have considered statistical parametric models, θ being the parameter, possibly multi-dimensional, usually unknown and to be estimated.

Besides their usefulness for clustering, mixture models are a very flexible method of modelling. As any continuous distribution can be approximated arbitrarily well by a finite mixture of normal densities, mixture models provide a convenient semi-parametric framework in which to model unknown distributional shapes, whatever the objective.

These models are also interesting in that they may point out a hidden variable responsible for most of the observed variability and so that the observed variables are *conditionally* independent. Their estimation is often difficult due to the missing data. The Expectation-Maximization (EM) algorithm[2,3] is a general and now standard approach to maximization of the likelihood in missing data problems. The algorithm iteratively maximizes the expected complete-data log-likelihood over values of the unknown parameters, conditional on the observed data and the current values of those parameters. In the clustering context, it provides parameter estimation but also values for missing data by providing membership probabilities to each group.

Standard mixture models correspond to independent z_i's. They are more and more used in statistical pattern recognition. They allow a formal (model-based) approach to (unsupervised) clustering and learning.

When the z_i's are not independent, the inter-relationship between sites can be maintained by a so-called neighborhood system usually defined through a graph. In this case, we will rather consider Hidden Markov Models (HMM) as presented in the following section.

2.2 Markov models

Keywords : *graphical models, Markov properties, conditional independence, hidden Markov field, EM algorithm, stochastic algorithms, selection and combination of models, image analysis.*

Graphical modelling provides a diagrammatic representation of a joint probability distribution, in the form of a network or graph depicting the local relations among variables. The graph can have directed or undirected edges between the nodes, which represent the individual variables. Associated with the graph are various Markov properties that specify how the graph encodes conditional independence assumptions.

It is the conditional independence assumptions that give the graphical models their fundamental modular structure, enabling computation of globally interesting quanti-

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B*, 39(1) :1–38, 1977.

[3] T. Krishnam G. McLachlan. *The EM algorithm and extensions.* John Wiley, New York, 1997.

ties from local specifications. Hence, graphical models form an essential basis for our methodologies based on structures.

The graphs can be either directed, *e.g.* Bayesian Networks, or undirected, *e.g.* Markov Random Fields (MRF). The specificity of Markovian models is that the dependencies between the nodes are limited to the nearest neighbor nodes. The neighborhood definition can vary and be adapted to the problem of interest. When parts of the variables (nodes) are not observed or missing, we refer to these models as Hidden Markov Models (HMM). Hidden Markov chains or hidden Markov fields correspond to cases where the z_i's in (1) are distributed according to a Markov chain or a Markov field. They are natural extension of mixture models. Although a great deal of work has been done on clustering and in particular on *feature-based clustering* in the image community, surprisingly not so much work has gone into how to make the resulting clusters spatially coherent. As an example, a very popular algorithm is the *Mean Shift* algorithm of Comaniciu and Meer[4]. It is a non parametric method and has many advantages and interesting features but as many other algorithms it does not offer the same capabilities as mixture models when trying to account for spatial dependencies. Hidden Markov models are widely used in signal processing (speech recognition, genome sequence analysis) and in image processing (remote sensing, MRI, *etc.*). Such models are very flexible in practice and can naturally account for the phenomena to be studied. They are very useful in modeling spatial dependencies but these dependencies and the possible existence of hidden variables are also responsible for a typically large amount of computation. It follows that the statistical analysis may not be straightforward. The following Section 3 presents the approaches I propose to use to address a number of typical issues.

[4] D. Comaniciu and P. Meer. Mean Shift : A robust approach toward feature space analysis. *IEEE Trans. PAMI*, 24(5) :603–619, 2002.

3 Domains of research

In the following developments, regular citations correspond to my publications listed at the end of the manuscript. Other bibliographical references quoted as footnotes appear at the bottom of each page.

In the last few years, mixture models have became a widely used approach to statistical pattern recognition. Efficient applications rely on the development of models and techniques allowing learning of the patterns of interest at a reasonable cost. In this context, I have focused on clustering tasks underlying a number of challenging practical issues. I have designed and studied a number of models with missing or hidden data structure (**Section 3.1**). I have provided appropriate inference frameworks based on the EM algorithm principle and have given, when accessible, insights on their theoretical properties such as their convergence and speed (**Section 3.2**). I have also considered the issue of model selection, which can be critical in applications (**Section 3.3**).
The various challenging practical issues we had to face are presented in **Section 4**.

3.1 Clustering with hidden structure models

One source of complexity is related to the measurement process. I addressed issues where in addition to the structure, some observations are missing [29, 47]. Section 3.1.1 summarizes our investigations to deal with high-dimensional observations (y_i's) while Section 3.1.2, describes a model we proposed to handle data from heterogeneous modalities. I also developed with collaborators a clustering approach for the issue of rigid and articulated Point Registration in computer vision. This is reported in Section 3.1.3.

A second source of complexity comes from possible dependencies between observed and hidden variables. In Sections 3.1.4 to 3.1.7, I propose ways to extend the standard hidden Markov field model to increase its modeling power. Typical issues are related to the neighborhood structure on the $z_i's$ to be chosen when not dictated by the context. This also requires a good understanding of the role of each parameter and methods to tune them depending on the goal in mind (Section 3.1.4). In Section 3.1.5, we then describe a model able to handle the need for cooperation between two related clustering processes ($z_i = (s_i, t_i)$ with the s_i's and t_i's being two distinct labels sets) of the same observed variables. In Section 3.1.6, we adopt a Bayesian approach. In addition to the z_i's, other missing variables are introduced but they represent unknown parameters. These parameters are used to weight differently each observation components in a multivariate observations case. In Section 3.1.7 the Markov field structure of the z_i's is enriched by introducing a set of sub-class variables. They are also missing but they allow to account for more complex class-dependent distributions (noise model) using standard hidden Markov models tools. Eventually in Section 3.1.8, we report some preliminary attempts to investigate possible relationships between our discrete graphical Markov models and models from geostatistics.

3.1.1 Taking into account the curse of dimensionality.

Related publications : 1 journal paper [8], 1 international conference [40].

This work has been done in the context of the ACI project Movistar (Section 6 of the Detailed Curriculum Vitae part) which I coordinated for team MISTIS*. My participation was more specifically related to Juliette Blanchet's PhD work that I advised (2004-2007).*

The mixture ingredient under consideration here is the choice of an appropriate class-dependent distribution $f(y_i|z_i;\theta)$ in equation (1) when y_i is high dimensional. There are not so many one dimensional distributions for continuous variables that generalize to multidimensional ones except when considering product of one dimensional independent components. The multivariate Gaussian distribution is the most commonly used but it suffers from significant limitations when it comes to modeling real data sets. For very high dimensional data, the general covariance matrix model involves the estimation of too many parameters, leading to intractable computations or singularity issues. Solutions have been proposed based on so-called *parsimonious models*[5] but

[5] G. Govaert G. Celeux. Gaussian parsimonious clustering models. *Pattern Recognition*, 28(5) :pages

they are not specifically designed for high dimensional data. They do not take into account the fact that real data points are often confined to a region of the space having lower effective dimensionality, so that the data actually live on a smaller dimensional manifold embedded within the high dimensional space. Other approaches consider reducing the dimensionality of the data as a pre-processing step possibly using principal component analysis or variable selection methods. In a classification context, this may not be satisfactory as relevant information may be lost that can help separating the classes. For these reasons, we rather considered a more recent approach developed for independent Gaussian mixtures. In the PhD work of Charles Bouveyron[6](co-advised by Stéphane Girard from team MISTIS and Cordelia Schmid from team LEAR, from 2003 to 2006 as part of the Movistar project), new Gaussian models of high dimensional data for classification purposes were proposed. It is assumed that the data live in several groups located in subspaces of lower dimensions. Such models are original in that a dimension reduction constraint can be introduced for each group. They are also parsimonious models in the sense that different groups can share the same values of some parameters. This modeling yields new supervised classification methods called HDDA for High Dimensional Discriminant Analysis [7,8]. This approach has been adapted to the unsupervised classification framework, and the related method is named HDDC for High Dimensional Data Clustering[9]. The approach actually includes as a particular case the *mixtures of probabilistic principal component analyzers*[10] which have been proposed only in a non supervised case. Link with the Discrimination Subspace Model (DSM) of Flury et al.[11] is also detailed in Charles Bouveyron's thesis[6].

In the context of J. Blanchet's PhD thesis, we combined the method to our Markov-model based approach of learning and clustering and reported at that time successful results as described in Section 3.1.4 and 3.1.7. More generally, this combination allows us to maintain also the efficiency and tractability of the Markov models as proposed in Sections 3.1.2, 3.1.5 and 3.1.6 for high dimensional data. Both dependencies between sites and dependencies between components of the multidimensional observations are modelled while the number of parameters to be estimated remains tractable. We obtained significant improvement in applications such as texture recognition [8, 40], where

781–793, 1995.

[6] C. Bouveyron. *Modélisation et classification des données de grande dimension. Application à l'analyse d'images.* PhD thesis, Université Grenoble 1, septembre 2006. `http ://tel.archives-ouvertes.fr/tel-00109047`.

[7] C. Bouveyron, S. Girard, and C. Schmid. Class-specific subspace discriminant analysis for high-dimensional data. In C. Saunder et al., editor, *Lecture Notes in Computer Science*, volume 3940, pages 139–150. Springer-Verlag, Berlin Heidelberg, 2006.

[8] C. Bouveyron, S. Girard, and C. Schmid. High Dimensional Discriminant analysis. *Communications in Statistics, Theory and Methods*, 36(14) :2607–2623, 2007.

[9] C. Bouveyron, S. Girard, and C. Schmid. High dimensional data clustering. *Computational Statistics and Data Analysis*, 52 :502–519, 2007.

[10] M. Tipping and C. Bishop. Mixture of probabilistic principal component analyzers. *Neural Computation*, 11(2) :443–482, 1999.

[11] L. Flury, B. Boukai, and B. Flury. The discrimination subspace model. *Journal of the American Statistical Association*, 438(92) :758–766, 1997.

the observations are high-dimensional (See Figure 1 and Section 4.1.1).

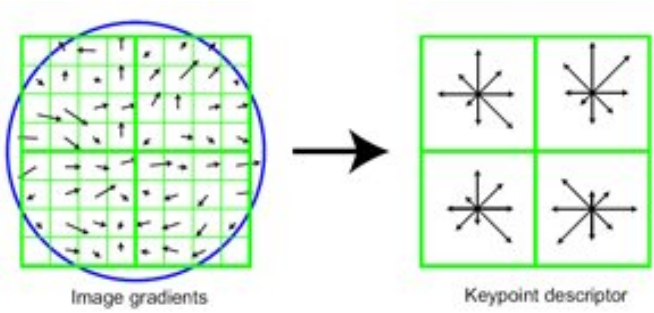


FIGURE 1 – SIFT (Scale Invariant Feature Transformation) descriptors used in computer vision are potentially high dimensional : when feature points have been identified, a descriptor is generated for each point. Left image : gradients in a neighborhood of a feature point. Right image : resulting quantification of gradient in 8 main directions for each of the 2×2 zones of the neighborhood. In practice we used 4×4 zones resulting in SIFT descriptors of size $4 \times 4 \times 8 = 128$.

3.1.2 Conjugate mixture model for clustering multimodal data

Related publications : 1 journal paper [1], 2 international conferences [32, 33].

This was the main topic of Vasil Khalidov's PhD thesis that I co-advised from 2007 to 2010. This is joint work with Radu Horaud from INRIA team PERCEPTION. *This work was initiated in the European STREP POP (Perception On Purpose 2006-2008) coordinated by Radu Horaud (see Section 6 in the Detailed Curriculum Vitae part). We derived benefit from the various collaborations generated by the project (see Sections 4.1.2 and 4.2.3 for more details).*

In some situations, it may not be possible to account for groups in data with a single mixture model such as (1). In particular, we addressed the issue of clustering observations that are gathered using multiple measuring instruments, *e.g.* using several physically different sensors. A typical such issue is audio-visual speaker detection (Section 4.2.3).

When the data originates from a single object (*e.g.* speaker), finding the best estimates for the object's characteristics is usually referred to as a pure fusion task and it reduces to combining multisensor observations in some optimal way. The problem is much more complex when several objects are present and when the task implies their detection, identification, and localization. In this case one has to consider two processes simultaneously : (i) segregation[12] which assigns each observation either to an object or

[12] J. W. Fisher III, T. Darrell, W. T. Freeman, and P. Viola. Learning joint statistical models

to an outlier category and (ii) estimation which computes the parameters of each object based on the group of observations that have been assigned to that object. In other words, in addition to fusing observations from different sensors, multimodal analysis requires the assignment of each observation to one of the objects.

This observation-to-object association problem can be cast into a probabilistic framework. In the case of unimodal data (possibly multidimensional), the problems of grouping observations and of associating groups with objects can be cast into the framework of standard data clustering. The problem of clustering multimodal data raises the difficult question of how to group together observations that belong to different physical spaces with different dimensionalities, *e.g.* how to group visual data (dense 2D or 3D data) with auditory data (sparse 1D data) ? When the observations from two different modalities can be aligned pairwise, a natural solution is to consider the Cartesian product of two unimodal spaces. Unfortunately, such an alignment is not possible in most practical cases. Different sensors operate at different frequency rates and hence the number of observations gathered with one sensor can be quite different from the number of observations gathered with another sensor. Consequently, there is no obvious way to align the observations pairwise. Considering all possible pairs would result in a combinatorial blow-up and typically create abundance of erroneous observations corresponding to inconsistent solutions. Alternatively, one may consider several unimodal clusterings, provided that the relationships between a common object space and several observation spaces can be explicitly specified. Multimodal clustering then results in a number of unimodal clusterings that are jointly governed by the same unknown parameters characterizing the object space.
In [1] we showed how the problem of clustering multimodal data could be addressed within the framework of mixture models. The proposed model is composed of a number of modality-specific mixtures. These mixtures are jointly governed by a set of K common object-space parameters (which are referred to as the tying parameters), thus ensuring consistency between the sensory data and the object space being sensed (see Figure 2). This is done using explicit transformations from the unobserved parameter space (object space) to each of the observed spaces (sensor spaces). Hence, the proposed model is able to deal with observations that live in spaces with different physical properties such as dimensionality, space metric, sensor sampling rate, *etc.* We believe that linking the object space with the sensor spaces based on object-space-to-sensor-space transformations has more discriminative power than existing multisensor fusion techniques and hence performs better in terms of multiple object identification and localization. To the best of our knowledge, there has been no attempt to use a generative model, such as ours, for the task of multimodal data interpretation. The concept of conjugate mixture models is described in more detail in our paper [1]. Standard Gaussian mixture models are used to model the unimodal data. The parameters of these Gaussian mixtures are governed by the object parameters $\{o_1, \dots o_K\}$ through a number of object-space-to-sensor-space transformations (one transformation for each sensing modality, *e.g.* $\mathcal{F}$

for audio-visual fusion segregation. In *Proceedings of Annual Conference on Advances in Neural Information Processing Systems*, Vancouver, BC, Canada, 2001.

and $\mathcal{G}$ if two modalities). A very general class of transformations, namely non-linear Lipschitz continuous functions is assumed. Figure 3 shows a graphical representation of our conjugate mixture models. Typically, for two modalities generating observations $\{f_1, \dots f_M\}$ and $\{g_1, \dots, g_N\}$, we considered a log-likelihood of the form :

$$\sum_{j=1}^{M} \log \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(f_j; \mathcal{F}(o_k), \Sigma_k) + \pi_{K+1} \mathcal{U}(f_j; \mathbb{F}) \right) + \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \lambda_k \mathcal{N}(g_i; \mathcal{G}(o_k), \Gamma_k) + \lambda_{K+1} \mathcal{U}(g_i; \mathbb{G}) \right) \quad (2)$$

where $\mathcal{N}(f_j; \mathcal{F}(o_k), \Sigma_k)$ (resp. $\mathcal{N}(g_i; \mathcal{G}(o_k), \Gamma_k)$) denotes the Gaussian distribution with mean $\mathcal{F}(o_k)$ (resp. $\mathcal{G}(o_k)$) and variance Σ_k (resp. Γ_k), $\mathbb{F}$ and $\mathbb{G}$ represent the observation spaces and additional uniform components on $\mathbb{F}$ and $\mathbb{G}$ have been added to account for outliers.

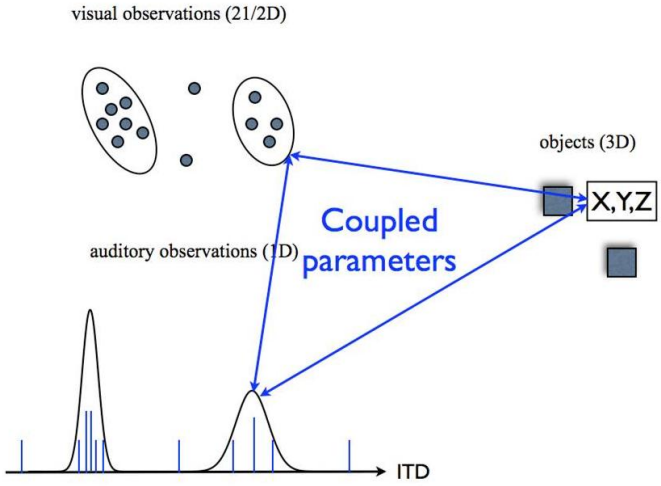


FIGURE 2 – Conjugate model fitting in the two modality case : observed audio data $\{G_1 \dots G_N\}$ are 1D (ITD) while video data $\{F_1 \dots F_M\}$ are 3D (stereo camera). The objects to be localized $o_1, \dots o_K$ are in 3D space.

3.1.3 Rigid and Articulated Point Registration with Expectation Conditional Maximization

Related publications : 1 journal paper [3], 1 international conference [39].

This is joint work with Radu Horaud, Guillaume Dewaele from the INRIA team PERCEPTION *and Manuel Iguel from team* EMOTION.

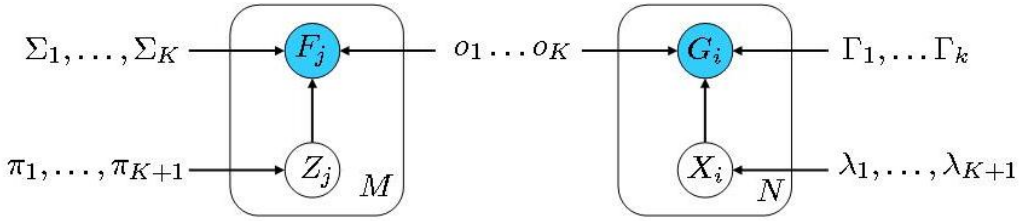
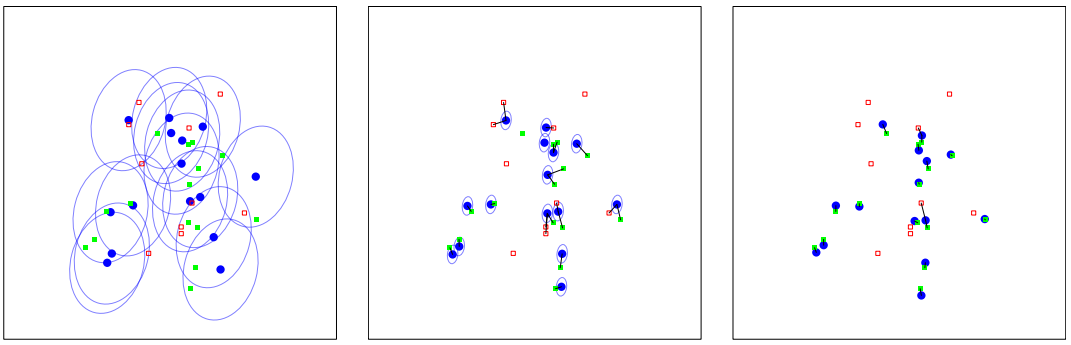


FIGURE 3 – Graphical representation of the conjugate mixture model in the case of two modalities. The observed data $\{F_1 \ldots F_M\}$ and $\{G_1 \ldots G_N\}$ consists of two sets of respectively M and N observations. We use the graphical notation introduced by C. Bishop[13]. Circles denote random variables, plates (rectangles) around them represent multiple similar nodes, their number being given in the plates. The representation inside each plate corresponds to a Gaussian mixture model (the observed data are in grey and additional hidden variables Z_j and X_i are introduced). The two modalities are then linked by the *tying parameters* $o_1, \ldots o_K$ shown in between the two plates.

We investigated a topic where mixture model clustering may not appear naturally at first due to the existence of a lot of classes compared to the number of observations to be clustered. This topic is that of finding an optimal alignment between two sets of points (see Figure 4). In image analysis and computer vision there is a long tradition of such algorithms. The issue is referred to as the *point registration* (PR) problem, which is twofold : (i) find point-to-point correspondences and (ii) estimate the transformation allowing the alignment of the two sets. Existing PR methods can be roughly divided into three categories : the Iterative Closest Point (ICP) algorithm and its numerous extensions, soft assignment methods and probabilistic methods. Probabilistic point registration uses, in general, Gaussian mixture models. Indeed, one may reasonably assume that points from the first set (the data) are normally distributed around points belonging to the second set (the model). Therefore, the point-to-point assignment problem can be recast into that of estimating the parameters of a mixture. One has to estimate the mixture parameters as well as the point-to-cluster assignments, *i.e.*, the missing data. In this case the algorithm of choice is the expectation-maximization (EM) algorithm. As explained in detail in [3], there are intrinsic difficulties when one wants to cast the PR problem in the EM framework. The main topic and contribution of our work was to propose an elegant and efficient way to do that. We introduced an innovative EM-like algorithm, namely the *Expectation Conditional Maximization for Point Registration* (ECMPR) algorithm. The algorithm allows the use of general covariance matrices for the mixture model components and improves over the isotropic covariance case. We analyzed in detail the associated consequences in terms of estimation of the registration parameters, and we proposed an optimal method for estimating the rotational and translational parameters based on *semi-definite positive relaxation.* We extended rigid registration to articulated registration. Robustness is ensured by

detecting and rejecting outliers through the addition of a uniform component to the Gaussian mixture model at hand. We have provided an in-depth analysis of our method and we have compared it both theoretically and experimentally with other robust methods for point registration. In general, ECMPR performed better than ICP. In particular it is less sensitive to initialization and it is more robust to outliers.



2nd iteration 6th iteration 35th iteration

FIGURE 4 – The point registration problem : Illustration of our ECMPR-rigid algorithm. The goal is to find an optimal alignement between model and data points. In this example, there are 15 model points (filled circles) and 25 data points, 15 inliers (filled squares) that correspond to model points that were rotated, translated and corrupted by anisotropic Gaussian noise, and 10 outliers (empty squares) drawn from a uniform distribution. Here, we modelled all the components of the mixture model with a common covariance matrix (shown with ellipses). The lines correspond to current data-to-model assignments. The algorithm stopped at the 35th iteration. There are 12 data-to-model assignments and 7 data-to-outlier assignments corresponding to the ground truth.

3.1.4 Constrained clustering

Related publications : 1 journal paper [5], 2 international conferences [42, 43], 6 national conferences [46, 47, 50, 51, 55, 56].

This part gathers essentially questions addressed first with Nathalie Peyrard during her PhD (1998-2001) and then with Matthieu Vignes and Juliette Blanchet (2004-07).

By constrained clustering model, we mean specific instance and usage of hidden Markov models that we developed to combine various sources of interactions and information. More specifically, as regards classification issues, the goal was to take into account simultaneously data from individual objects, *i.e.* data that make sense and exist for each objects, and data from pairs of objects reflecting for instance some distance or some similarity measure defined on the objects. In practice such data can be missing and EM offers a good framework to deal with this case (see *e.g.* [47]). Approaches to analyze such data fall mainly in two categories. Some focus on individual data and as a consequence assume that they are independent. Others use information on pairs in

the form of networks or graphs but do not directly use individual data associated with objects in the networks. Sequential procedures, clustering first individual data alone and incorporating additional information only after the clusters are determined, are necessarily suboptimal. Our aim is to take into account both types of information in a single procedure. We proposed a hidden Markov random field model in which parametric probability distributions account for the distribution of individual data for each object. Data on pairs are then included through a graph where the nodes represent the objects and the edges are weighted according to pair data reflecting distance or similarity measures between objects. There exist many ways to do that and it is not clear whether they are equivalent in terms of the amount of information taken into account and in terms of clustering results. An illustration is given in Section 4.2.1 with an application to genetic data analysis [5] carried out during the PhD thesis of Matthieu Vignes that I advised (2003-2007).

One of the difficulties is to choose how the various information can be incorporated in the model depending on the goal in mind. This requires a good understanding of the role of each parameter in a hidden Markov random field model. With this in mind, in [56], we investigated the role of *singleton potentials* which are parameters often ignored in standard Markov model-based segmentation. We then used (*e.g.* in [6]) these potentials to take into account cooperatively two sources of information so that two segmentation processes could refine mutually and lead to better segmentation results (see Section 4.1.3 for an application to MRI analysis). This later idea of cooperative clustering processes is further developed in the next Section 3.1.5.

Then an additional difficulty is how much or with which weight each source of information has to be incorporated. For instance, when considering multi channel MRI segmentation, taking advantage of the various protocols that acquire images using multiple modalities is a current issue (typically T1-weighted, T2-weighted, PD-weighted, Diffusion-weighted or Flair sequences in MR neuroimaging). Their unique and complementary information should be merged together before a segmentation to get rid of the inconsistencies one can encounter when segmenting on each modality separately. For single MRI volumes, intensity distributions are usually modelled as one dimensional Gaussian distributions, which have the advantage to generalize easily to higher dimensions. However, the presence of different MR sequences rises additional questions : Do all the sequences have the same importance in term of information content or discriminative power ? If not as we suspect, how to deal with this and combine these multiple channels in an optimal way ? To weight the different channels and address the issue of how to find the optimal weights, we proposed [27, 26] a weighted clustering approach further described in Section 3.1.6. Another question is how to encode expert knowledge on the temporal evolution of lesions ? A first natural idea is to segment the multichannel data available at some current date and use the result as an a priori when dealing with the data at the following date. This could be formalized like in Markov chains by defining conditional models to capture relationships between two successive dates either via the external field or the class distributions like in [31].

When dealing with Markov models on irregular graphs, choosing the neighborhood structure can then be an additional issue. This choice may depend on the application.

As regards Markov models, the automatic neighborhood selection has not been really addressed in the literature except more recently by Le Hegarat-Mascle et al.[14]. In our experiments described in Section 4.1.1, it appeared that graphs with similar numbers of neighbors for each site give more satisfying results. Directions of research for neighborhood selection can be found in the Le Hegarat-Mascle et al. paper, which contains in particular ideas related to *Triplet Markov Fields* (Section 3.1.7).

3.1.5 Bayesian Markov model for cooperative clustering

Related publications : 5 international conferences [36, 31, 28, 35, 25].

I initiated this work to address a number of issues raised in Benoit Scherrer's PhD work (2005-08) in collaboration with Michel Dojat from Grenoble Institute of Neuroscience and Catherine Garbay from Laboratoire d'Informatique de Grenoble. This framework appeared to be useful in Ramya Narasimha's PhD work too (2006-10).

In the previous section, we mentioned the need for approaches that couple different levels of information. When clustering or segmenting data, attention has been focused on clustering various data types, regular vector data, curve data or more heterogeneous data. But in these cases, the goal is to produce a single partitioning of the individuals that account for the groups existing in the observed data. The task we consider now is that of producing more than one partitioning using the same data. We refer to this task as a *cooperative clustering* task. Examples of applications in which this is relevant include tissue and structure segmentation in MR brain scan analysis [4, 6, 7] and Section 4.1.3, simultaneous estimation of motion discontinuities and optical flow in motion analysis[15], consistent depth estimation and boundary ([35] and Section 4.1.2) or depth discontinuity[16] detection, consistent depth and surface normal estimation ([25] and Section 4.1.2), *etc.*

Our so-called cooperative approach consists then in assuming the existence of two (or more) sets of missing variables, representing two (or more) sets of labels which are usually not independent but related in the sense that information on one of them can help in finding the other. Besides the need for modeling cooperation, in many applications, data points are not independent and require models that account for dependencies. For this purpose, we used Markov random field (MRF) models to further specify our missing data model. Faced with processes interacting on a wide variety of scales, modeling can be done from either a joint or conditional viewpoint. Although it may be intuitive to consider processes from a joint perspective, it can present serious challenges to statistical modeling. Even when it is possible to specify a joint multivariate

[14] S. Le Hegarat-Mascle, A. Kallel, and X. Descombes. Ant colony optimization for image regularization based on a nonstationary Markov modeling. *IEEE Trans. Image Processing*, 16(3) :865–879, 2007.

[15] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Trans. Pat. Anal. Mach. Intell.*, 15(12) :1217–1232, 1993.

[16] J. Sun, N.N. Zheng, and H.Y. Shum. Stereo matching using belief propagation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(7) :787–800, July 2003.

distribution, inference may be possible only through equivalent conditional models. Hierarchical approaches[17] deal with joint modeling by writing joint distributions as products of a series of conditional models, *e.g.* $p(\mathbf{y}, \mathbf{z}, \theta) = p(\mathbf{y}|\mathbf{z}, \theta)\ p(\mathbf{z}|\theta)\ p(\theta)$, with $\mathbf{y}$, $\mathbf{z}$, θ denoting respectively the data, the labels and the parameters. Such models can be considered from either a classical or Bayesian perspective. We proposed to use the latter one for its additional modeling capabilities in a number of applications. We used the term Bayesian which means that all the unknown quantities are treated identically, be they the segmentation labels or the model parameters.

In our cooperative setting, we focus more particularly on situations where the *process stage*, *i.e.* the specification of $p(\mathbf{z}|\theta)$, is made of different sub-processes which are linked and provide complementary information. We have proposed an approach different from standard hierarchical modeling. It is based on the decomposition of joint models into a series of conditional models but which do not necessarily correspond to the factors in a standard factorized decomposition. We refer to this alternative decomposition as the *cooperative approach* because the focus is on capturing interactions (cooperations) between the unknown quantities namely, sub-processes and parameters. We have derived then a class of joint Markov models based on the specification of a system of coherently linked conditional models that capture several level of interactions. They incorporate 1) dependencies between individuals within each label set, which is usually referred to as spatial interactions in spatial data analysis; 2) relationships between label sets for cooperative aspects and 3) *a priori* information for consistency with expert knowledge. Besides, the Bayesian setting and more specifically the addition of a conditional model on parameters allows to encode additional information and constraints. The decomposition we proposed is particularly well adapted to such inference techniques which are based on alternating optimization procedures in which variables of interest are examined in turn and that conditionally on the other variables. More specifically, we have shown that inference could be described in terms of the conditional models $p(\mathbf{z}|\mathbf{y}, \theta)$ and $p(\theta|\mathbf{y}, \mathbf{z})$. Defining these conditional models is equivalent to defining the conditional distribution $p(\mathbf{z}, \theta|\mathbf{y})$. It follows that for segmentation purpose, there is no need to define a joint model $p(\mathbf{y}, \mathbf{z}, \theta)$, the conditional distribution $p(\mathbf{z}, \theta|\mathbf{y})$ contains all useful information. Equivalently there is no need to specify $p(\mathbf{y})$. This point of view is also the one adopted in Conditional Random Fields (CRF)[18] which have been widely and successfully used in applications including text processing, bioinformatics and computer vision.

In the case of two label sets, denoting $\mathbf{z} = (\mathbf{t}, \mathbf{s})$ the two label sets, it appears that the specification of the three conditional distributions $p(\mathbf{t}|\mathbf{s}, \mathbf{y}, \theta)$, $p(\mathbf{s}|\mathbf{t}, \mathbf{y}, \theta)$ and $p(\theta|\mathbf{t}, \mathbf{s}, \mathbf{y})$ is necessary and sufficient. In practice, the advantage of writing things in terms of the conditional distributions $p(\mathbf{t}|\mathbf{s}, \mathbf{y}, \theta)$ and $p(\mathbf{s}|\mathbf{t}, \mathbf{y}, \theta)$ is that it allows to capture cooperations between $\mathbf{t}$ and $\mathbf{s}$. We have proposed then an appropriate way to build these conditional distributions for the model inference. We assume that $\mathbf{y}, \mathbf{z}$ and θ are all

[17] C. K. Wikle. Hierarchical models in environmental science. *Int. Stat. Review*, 71 :81199, 2003.

[18] S. Kumar and M. Hebert. Discriminative random fields. *Int. J. Comput. Vision*, 68(2) :179–201, 2006.

defined on S (more general cases are easy to derive) and that the conditional distribution $p(\mathbf{z},\theta|\mathbf{y})$ is a Markov random field with energy function $H(\mathbf{z},\theta|\mathbf{y})$, *i.e.* $p(\mathbf{z},\theta|\mathbf{y}) \propto \exp(H(\mathbf{z},\theta|\mathbf{y}))$, with $H(\mathbf{z},\theta|\mathbf{y}) = \sum_{c\in\mathcal{C}} \left(U^c_{\mathbf{Z},\Theta}(\mathbf{z}_c,\theta_c|\mathbf{y}) + U^c_{\mathbf{Z}}(\mathbf{z}_c|\mathbf{y}) + U^c_{\Theta}(\theta_c|\mathbf{y}) \right)$, where the sum is over the set of cliques $\mathcal{C}$ and $\mathbf{z}_c$ and θ_c denote realizations restricted to clique c. The U^c's are the clique potentials that may depend on additional parameters, not specified in the notation. In addition, in the formula above, terms that depend only on $\mathbf{z}$, resp. θ, are written explicitly and are distinguished from the first term in the sum in which $\mathbf{z}$ and θ cannot be separated. Conditions ensuring the existence of such a distribution can be found in H-O. Georgii's book[19].

From the Markovianity of the joint distribution it follows that any conditional distribution is also Markovian. Note that this is not true for marginals of a joint Markov field which are not necessarily Markovian. Let us then consider MRF $p(\mathbf{z},\theta|\mathbf{y})$ whose energy function can be written as

$$H(\mathbf{z},\theta|\mathbf{y}) \;=\; H_{\mathbf{Z}}(\mathbf{z}) + H_{\Theta}(\theta) + H_{\mathbf{Z},\Theta}(\mathbf{z},\theta) + \sum_{i\in S} \log f(y_i|z_i,\theta_i) \;.$$

The Markovian energy is separated into terms $H_{\mathbf{Z}}$, H_{Θ}, $H_{\mathbf{Z},\Theta}$ that involve respectively only $\mathbf{z}$, only θ and interactions between θ and $\mathbf{z}$. With $\mathbf{z} = (\mathbf{t},\mathbf{s})$, we can further specify

$$\begin{aligned} H_{\mathbf{Z}}(\mathbf{z}) &= H_{\mathbf{T}}(\mathbf{t}) + H_{\mathbf{S}}(\mathbf{s}) + \tilde{H}_{\mathbf{T},\mathbf{S}}(\mathbf{t},\mathbf{s}) \\ \text{and} \quad H_{\mathbf{Z},\Theta}(\mathbf{z},\theta) &= H_{\mathbf{T},\Theta}(\mathbf{t},\theta) + H_{\mathbf{S},\Theta}(\mathbf{s},\theta) + \tilde{H}_{\mathbf{T},\mathbf{S},\Theta}(\mathbf{t},\mathbf{s},\theta) \;, \end{aligned}$$

where we used a different notation $\tilde{H}$ to make clearer the difference between the energy terms involving interactions only (resp. $\tilde{H}_{\mathbf{T},\mathbf{S}}$ and $\tilde{H}_{\mathbf{T},\mathbf{S},\Theta}$) and the global energy terms (resp. $H_{\mathbf{Z}}$ and $H_{\mathbf{Z},\Theta}$). $H_{\Theta}(\theta)$ and $H_{\mathbf{Z}}(\mathbf{z})$ can be interpreted as priors resp. on Θ and $\mathbf{Z}$. In a cooperative framework, the prior on $\mathbf{Z}$ can be itself decomposed into an *a priori* cooperation term $\tilde{H}_{\mathbf{T},\mathbf{S}}(\mathbf{t},\mathbf{s})$ and individual terms which represent *a priori* information on $\mathbf{T}$ and $\mathbf{S}$ separately. $H_{\mathbf{T},\mathbf{S},\Theta}(\mathbf{t},\mathbf{s},\theta)$ specifies the *process*, *i.e.* the underlying model, that can also be decomposed into parts involving $\mathbf{t}$ and $\mathbf{s}$ separately or together. In what follows, we also assume that $\mathbf{t}$ and $\mathbf{s}$ are both defined on the set of sites S so that writing $z_i = (t_i, s_i)$ makes sense. With additional care, a more general situation could be considered if necessary. Eventually $\sum_{i\in S} \log f(y_i|t_i,s_i,\theta_i)$ corresponds to the data-term. From such a definition of $p(\mathbf{z},\theta|\mathbf{y})$, it follows then expressions of the conditional distributions required for inference.

As regards applications, the modeling capabilities of this new approach was first illustrated in [31] on brain MRI analysis. Benoit Scherrer the student I worked with, received the *Young Investigator Award* in segmentation at the MICCAI conference in 2008. Another paper along this line, with the important addition of a third registration process was accepted to the MICCAI conference in 2009 [28]. Therein, we considered a finite set S of N voxels on a regular 3D grid. We denote by $\mathbf{y} = \{y_1,\ldots,y_N\}$ the intensity values observed respectively at each voxel and by $\mathbf{t} = \{t_1,\ldots,t_N\}$ the hidden

[19] H-O. Georgii. *Gibbs measures and phase transition.* de Gruyter studies in Mathematics, 1988.

tissue classes. The t_i's take their values in $\{1, 2, 3\}$ that represents the three tissues cephalo-spinal-fluid, grey matter and white matter. In addition, we consider L subcortical structures and denote by $\mathbf{s} = \{s_1, \ldots, s_N\}$ the hidden structure classes at each voxel. Similarly, the s_i's take their values in $\{1, \ldots, L, L+1\}$ where $L+1$ corresponds to an additional background class. Relationships exist between tissues and structures. In particular, a structure is composed of a single known tissue. As parameters θ, we consider $\theta = \{\psi, R\}$ where ψ are the parameters describing the intensity distributions for the $K = 3$ tissue classes and R denotes registration parameters. Intensity distribution parameters are more specifically denoted by $\psi = \{\psi_i^k, i \in S, k = 1 \ldots K\}$. Standard approaches usually consider that intensity distributions are Gaussian distributions for which the parameters depend only on the tissue class. Here we consider that they could also depend on i and in the 1D Gaussian case, we would have $\psi_i^k = \{\mu_i^k, \sigma_i^k\}$ denoting respectively the mean and variance. *A priori* knowledge is incorporated through fields ϕ_T and ϕ_S representing *a priori* information respectively on tissues and on structures. In our study, these fields correspond to prior probabilities provided by a probabilistic atlas on structures that has to be registered on the observed brain image. The fields depend then on the registration parameters R. In our approach we perform registration and segmentation simultaneously by considering that the information provided by the atlas depends on the registration parameters R that have to be estimated as well as other model parameters and whose successive values will adaptively modified the registration. Within each label set $\mathbf{t}$ and $\mathbf{s}$, interactions between neighboring voxels are captured, for instance through two Potts fields involving interaction parameters respectively η_T and η_S. Interaction between label sets can be captured in a third term. Another term to be specified is that involving θ. The intensity distribution parameters and the registration parameters are supposed to be independent. The specific form of the prior on ψ is usually guided by the inference procedure as conjugate priors are easier to deal with. Figure 5 shows the graphical model representation of the model developed in [28] where $m, \lambda, \alpha, \beta$ denote hyperparameters.

3.1.6 Bayesian weighted Markov model for clustering multivariate data

Related publications : 2 international conferences [26, 27].

This is joint work with Senan Doyle and Darren Wraith, post-doctoral fellows in MISTIS *and Eric Frichot who was intern in the team in 2010.*

Considering specifically multidimensional data, I addressed the issue of weighting the various dimensions differently. I developed a clustering model in which weights are introduced and estimated in a Bayesian framework. This technique was successfully applied to the detection of brain lesions from multiple MR sequences (see Section 4.1.4).

Combining sequences is a data fusion issue which, in a probabilistic setting, naturally becomes an issue of combining probabilistic distributions. This relates to the so-called *pooling* of distributions in the statistical literature[20]. Examples include *linear*

[20] C. Genest, K. J. McConway, and M. J. Schervish. Characterization of externally Bayesian pooling

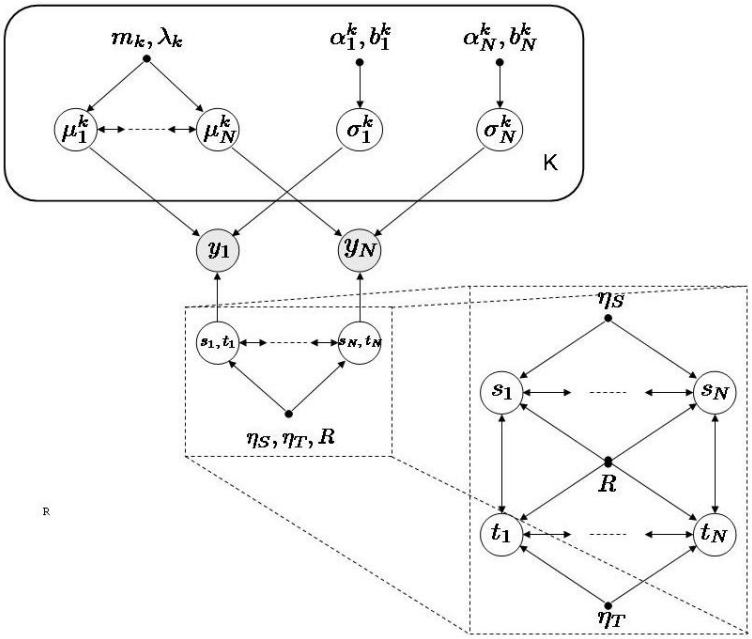


FIGURE 5 – Graphical model representation for the cooperative clustering model developed in [28] : two label sets are considered, resp. $\{s_1, \dots s_N\}$ and $\{t_1, \dots t_N\}$. The dashed lines show a zoom that specifies the interactions at the $\{s_i, t_i\}$ level (within label set, MRF, and between label sets interactions). MRF regularization also occurs at the hyperparameters level via smoothing of the means $\{\mu_1^k \dots \mu_N^k\}$.

and *logarithmic* pooling. The former corresponds to a mixture of distributions, while the latter consists of combining the distributions into a product where each component is raised to a power. This power can be viewed as a weight. In this work we considered logarithmic pooling for it appears that it was more appropriate for our segmentation framework. Note however that the link to pooling although interesting is only mentioned for information and that our approach could be presented without referring to such aspects.

We consider a finite set S of N data items. We denote by $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ the observed values. Each $\mathbf{y}_i = \{y_{i1}, \dots, y_{iM}\}$ is itself a vector of M values (*e.g.* corresponding to M different MR sequences). Our goal is to assign each item i to one of K classes considering the observed features data $\mathbf{y}$. For instance in our brain tissue segmentation task (Section 4.1.4), we consider in general 3 tissues plus some possible additional classes to account for lesions in pathological data. We denote the hidden classes by $\mathbf{z} = \{z_1, \dots, z_N\}$, and the z_i's take their values in $\{1 \dots K\}$. We consider nonnegative weights $\omega = \{\omega_i, i \in S\}$ with $\omega_i = \{\omega_{i1}, \dots, \omega_{iM}\}$. In our general setting the weights are dimension and item-specific. The rationale is that relevant information

operators. *Ann. Statist.*, 14 :487–501, 1986.

is not usually uniformly localized so that the weights cannot be fixed equally for all the items in a given dimension but may depend on the index i. Spatial dependencies between items are then introduced through Markov random field modeling.

The clustering or segmentation task is recast into a missing data framework in which $\mathbf{y}$ are the observations and $\mathbf{z}$ are the missing variables. Their joint distribution $p(\mathbf{y}, \mathbf{z}|\omega; \theta)$ is governed by the weights ω and parameters θ, which are both unknown and need to be estimated within the segmentation procedure. A prior distribution $p(\omega)$ is defined on the weights. Taking advantage of the fact that Bayesian inference does not differentiate between missing data and random parameters, we propose a framework in which the weights ω are viewed as additional missing variables. Denoting the parameters by $\theta = \{\psi, \beta\}$, we assume that the joint distribution $p(\mathbf{y}, \mathbf{z}, \omega; \theta)$ is a MRF with the following energy function :

$$H(\mathbf{y}, \mathbf{z}, \omega; \theta) = H_{\mathbf{Z}}(\mathbf{z}; \beta) + H_W(\omega) + \sum_{i \in S} \log f(\mathbf{y}_i | z_i, \omega_i; \psi) \tag{3}$$

where the energy term $H_W(\omega)$ involving only ω does not depend on θ and the $f(\mathbf{y}_i|z_i, \omega_i; \psi)$s are probability density functions of $\mathbf{y}_i$. The three terms in this energy are further specified below.

Data term. The data term $\sum_{i \in S} \log f(\mathbf{y}_i|z_i, \omega_i; \psi)$ in (3) corresponds to the modeling of class-dependent distributions. We considered two cases : first the particular simpler case where the distributions are M-dimensional Gaussian distributions with diagonal covariance matrices. For each class k, ${}^t(\mu_{k1}, \ldots, \mu_{kM})$ is the mean vector and $\{v_{k1}, \ldots, v_{kM}\}$ the covariance matrix components. When $z_i = k$, then $\mathcal{N}(y_{im}; \mu_{z_i m}, v_{z_i m})$ represents the Gaussian distribution with mean μ_{km} and variance v_{km}. The whole set of Gaussian parameters is denoted by $\psi = \{\mu_{km}, v_{km}, k = 1 \ldots K, m = 1 \ldots M\}$. Our data term is then defined by setting

$$f(\mathbf{y}_i|z_i, \omega_i; \psi) \quad = \quad \prod_{m=1}^{M} \mathcal{N}(y_{im}; \mu_{z_i m}, \frac{v_{z_i m}}{\omega_{im}}) \ ,$$

which is proportional to $\prod_{m=1}^{M} \mathcal{N}(y_{im}; \mu_{z_i m}, v_{z_i m})^{\omega_{im}}$. This corresponds to a *modified* logarithmic pooling[20] of the M distributions $p(z_i|y_{im}, \omega_{im}; \theta)$ and $p(z_i; \beta)$. Intuitively, the impact of a larger ω_{im} is to give more importance to the intensity value y_{im} in the model. Typically an integer ω_{im} greater than one would correspond to increase ω_{im} times the number of voxels with intensity value y_{im}. When the weights are all one, a standard multivariate Gaussian case is recovered.

Then, we considered also the more general case where the class-dependent distributions are M-dimensional Gaussian distributions with full covariance matrices. For each class k, $\mu_k = {}^t(\mu_{k1}, \ldots, \mu_{kM})$ is the mean vector and Σ_k the $M \times M$ covariance matrix. When $z_i = k$, then $\mathcal{N}(\mathbf{y}_i; \mu_{z_i}, \Sigma_{z_i})$ represents the M-dimensional Gaussian distribution with mean μ_k and covariance matrix Σ_k. The whole set of Gaussian parameters is denoted by $\psi = \{\mu_k, \Sigma_k, k = 1 \ldots K\}$. More specifically, we propose to consider the

[20] C. Genest, K. J. McConway, and M. J. Schervish. Characterization of externally Bayesian pooling operators. *Ann. Statist.*, 14 :487–501, 1986.

parametrization of the covariance matrices based on their eigenvalue decompositions :

$$\Sigma_k = D_k A_k D_k^T,$$

where D_k is the matrix of eigenvectors of Σ_k and A_k is a diagonal matrix with the corresponding eigenvalues of Σ_k. The matrix D_k determines the orientation of the k^{th} Gaussian and A_k its shape. Such a parameterization has the advantage to allow an intuitive incorporation of the weight parameters via the transformation of the covariance matrices given below. Our data term is then defined by setting :

$$f(\mathbf{y}_i|z_i, \omega_i; \psi) \quad = \quad \mathcal{N}(\mathbf{y}_i; \mu_{z_i}, D_{z_i}\Delta_i A_{z_i} D_{z_i}^T) \ ,$$

where Δ_i is the $M \times M$ diagonal matrix whose diagonal components are the inverse weights $\{\omega_{i1}^{-1}, \ldots, \omega_{iM}^{-1}\}$. This means that the weights are used to act on the shape of the Gaussians but not on their orientation. As before, when the weights are all one, a standard multivariate Gaussian case is recovered.

Missing label term. The missing data term $H_{\mathbf{Z}}(\mathbf{z}; \beta)$ in (3) is set to a standard Potts model, with external field ξ and spatial interaction parameter η, and whose energy is

$$H_{\mathbf{Z}}(\mathbf{z}; \beta) = \sum_{i \in S} (\xi_{iz_i} + \sum_{j \in \mathcal{V}(i)} \eta \ \mathbb{1}_{z_i = z_j}),$$

where $\mathcal{V}(i)$ denotes the items neighboring i and $\mathbb{1}_{z_i=z_j}$ is 1 when $z_i = z_j$ and 0 otherwise. Parameter $\beta = \{\xi, \eta\}$ with $\xi = \{^t(\xi_{i1} \ldots \xi_{iK}), i \in S\}$ being a set of real-valued K-dimensional vectors and η a real positive value.

Missing weight term. The weights are assumed independent from parameters θ and independent across modalities. The simplest choice is to define a prior $p(\omega) = \prod_{m=1}^{M} \prod_{i \in S} p(\omega_{im})$ where each $p(\omega_{im})$ is a Gamma distribution with hyperparameters α_{im} (shape) and γ_{im} (inverse scale). Thus

$$H_W(\omega) = \sum_{m=1}^{M} \sum_{i \in S} ((\alpha_{im} - 1) \log \omega_{im} - \gamma_{im} \ \omega_{im}).$$

In practice, the set of hyperparameters is fixed so that the modes of each prior $p(\omega_{im})$ are located at some *expert weights* $\{\omega_{im}^{exp}, m = 1 \ldots M, i \in S\}$ accounting for some external knowledge, if available. Formally, we set $\alpha_{im} = \gamma_{im} \ \omega_{im}^{exp} + 1$ to achieve this. The expert weights can be chosen according to the specific task. For example, when observations around a specific value of interest are not numerous enough to attract a model component, increasing the expert weight for some of them will help in biasing the model toward the identification of this class.

Note that we also investigated the use of Dirichlet distributions for the weights adding the constraints that they should sum to the sample size N in each modality. However, there were no real theoretical reasons to do so and it required less stable numerical computation. This is due to the fact that the Dirichlet distribution is not a conjugate distribution in our setting. In addition, in some preliminary experiments, results were not improved compared to the simpler independent Gamma case.

The graphical representation of the model with weights is shown in Figure 6.

We proposed to use a variational EM framework to deal with the missing label and weight data and the MRF prior. The fact that the weights can be equivalently considered as missing variables or random parameters induces some similarity between our Variational EM variant and the Variational Bayesian EM algorithm presented by M. Beal and Z. Ghahramani[21,22]. Our framework differs slightly. In contrast to these latter papers, our observations are not *i.i.d.* and condition (2) in Section 3 of M. Beal and Z. Ghahramani's paper[21] is not satisfied. However, these differences are not significant. More importantly, our missing data presentation offers the possibility to deal with extra parameters (the Gaussian means and variances in our setting) for which no prior information is available. This is done in a maximum likelihood manner and avoids the use of non-informative priors that could be problematic (difficulties with non informative priors are listed in a book by A. Gelman et al.[23] p. 64-65). As a consequence, the variational Bayesian M-step of Beal and Ghahramani[21] is transferred into our E-step while our M-step has no equivalent in the formulation of Beal and Ghahramani[21].

In practice, the diagonal case leads to a series of formulas for the updating while the general case requires the use of a numerical procedure (the Flury and Gautschi algorithm[5]) for the covariance matrices.

3.1.7 Triplet Markov Fields for non standard supervised clustering

Related publications : 1 journal paper [8], 1 research report [60], 1 international conference [40].

This was the main topic of Juliette Blanchet PhD thesis that I advised from 2004 to 2007.

In more and more high-level image analysis, such as feature-based object recognition or object tracking, images are not described by intensity levels on a regular grid but rather by local affine-invariant descriptors (*e.g.* Figure 1) and by spatial relationships between these descriptors. A graph is associated with an image with the nodes representing feature vectors describing image regions and the edges joining spatially related regions. For tractability, most approach to recognition assume independence between the features which is an obvious oversimplification. Incorporating information about the spatial organization of the descriptors leads to better recognition results. Current approaches consist in augmenting the data with information coming from the spatial

[21] M. Beal and Z. Ghahramani. *The variational Bayesian EM Algorithm for incomplete data : with application to scoring graphical model structures.* Bayesian Statistics. Oxford University Press, 2003.

[22] Z. Ghahramani and M. Beal. Propagation algorithms for variational Bayesian learning. In *Advances in Neural Information Processing Systems. MIT Press*, 2001.

[23] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis.* Chapman & Hall, 2nd edition, 2004.

[5] G. Govaert G. Celeux. Gaussian parsimonious clustering models. *Pattern Recognition*, 28(5) :pages 781–793, 1995.

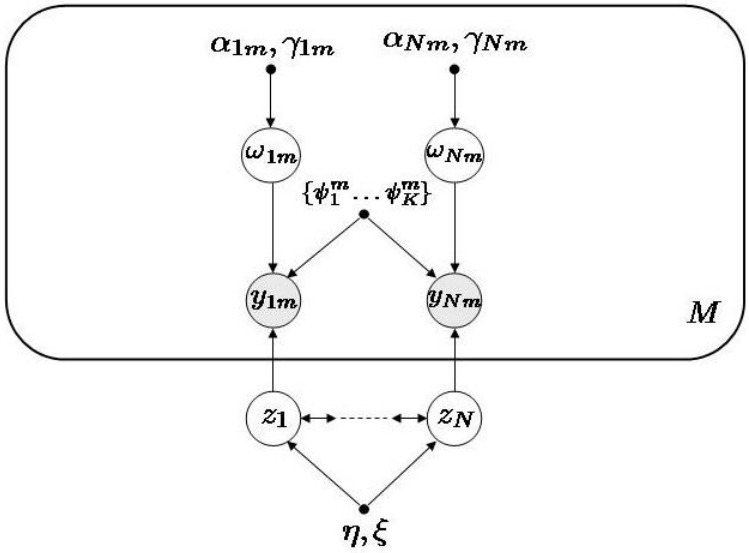


FIGURE 6 – Graphical model representation for the Bayesian weighted clustering model. Here the emphasize is on the addition of a set of weights $w_{1m}, \dots w_{Nm}$ to the standard hidden Markov field setting.

relationships, for instance by using co-occurrence statistics, but without modeling explicitly the dependencies between neighboring descriptors. In such approaches the underlying model is one where the descriptors are statistically independent variables. Our claim is that recognition results can be further improved by considering that descriptors are statistically dependent. We proposed to introduce the use of statistical parametric models of the dependence between descriptors. We used hidden Markov models which provide parametric models where the parameters have a natural interpretation and can be adjusted to incorporate a priori knowledge with respect to strength of interaction for instance. Another major source of complexity is related to the structure of the *noise* model or the distribution (say $p(\mathbf{y}|\mathbf{z};\theta)$) linking the unknown labels $\mathbf{z}$ to the observations $\mathbf{y}$. First, in real-world applications, data cannot usually be reduced to classes modelled by unimodal distributions and consequently by single Gaussian distributions. As regards the measurement process, we proposed class and site dependent mixtures of distributions and more specifically mixtures of Gaussian distributions which provide a richer class of density models than the single Gaussian distributions. To allow such more general distribution, a natural idea is to to decompose each class, given by the z_i's, into subclasses. Let us for instance assume that each of the K classes is decomposed into L sub-classes so that we can introduce additional variables $\{X_1, \dots, X_N\}$ indicating the sub-classes and then consider class and sub-class dependent distributions $P(.|x_i, z_i; \theta)$ that depend on some parameters θ, actually usually on part of θ denoted by $\theta_{x_i z_i}$. Then, a strong assumption of conditional independence of the observed data is generally used in the hidden Markov random field framework for tractability and this

standard model is sometimes referred to as HMF-IN (for Hidden Markov Field with Independent Noise). This assumption combined with the Markovianity of the hidden field has the advantage to lead to a distribution of the labels given the observations (the *posterior distribution*) which is Markovian. This last property is essential in all Markov model based clustering algorithms. However, conditional independence is too restrictive for a large number of applications such as textured or non-stationary image segmentation. For this reason, various Markov models have been proposed in the literature including Gaussian Markov fields[24] and more general hidden Markov models proposed by Pieczynski and Tebbache[25] and referred to as *Pairwise Markov models*. They are based on the observation that the conditional independence assumption is sufficient but not necessary for the Markovianity of the conditional distribution to hold. A further generalization has then been proposed by Benboudjema and Pieczynski[26] through the so-called *Triplet Markov Fields* (TMF), that is a triplet $(\mathbf{Y}, \mathbf{X}, \mathbf{Z})$ which has a Markovian joint distribution. It has larger modeling capabilities, allowing more general noise models and in particular multi-modal class distributions. We considered Triplet models different from those by Benboudjema and Pieczynski[26,27]. We proposed a class of Triplet Markov models with rich noise models that still allow standard processing as regards classification and parameter estimation. We proposed to use an Expectation Maximization framework and a mean field like approximation procedure (Section 3.2.1) originally developed for the standard hidden Markov field case. We focused in particular on texture recognition (see Section 4.1.1) but further work includes other contexts such as object recognition and tracking. Description and illustration of these models can be found in preliminary studies [60, 40] and in the journal publication [8]. The triplet models we considered in their general form are detailed in [8]. For illustration, we just show a simple example of such a triplet model. Figure 7 shows images corresponding to simulations of this example. It is given by :

$$P(\mathbf{y}, \mathbf{x}, \mathbf{z}) \propto \exp(\eta_1 \sum_{i \sim j} \mathbb{I}_{x_i = x_j} \mathbb{I}_{z_i = z_j} + \eta_2 \sum_{i \sim j} \mathbb{I}_{z_i = z_j} + \sum_{i \in S} \log f(y_i | \psi_{x_i z_i})) \qquad (4)$$

where η_1 and η_2 are real parameters, $\psi_{lk} = (\mu_{lk}, \Sigma_{lk})$ are the parameters of Gaussian distributions $f(.|\psi_{lk})$, for $l \in \{1 \dots L\}$ and $k \in \{1 \dots K\}$. The couple $(\mathbf{X}, \mathbf{Z})$ is then Markovian :

$$P(\mathbf{x}, \mathbf{z}) \propto \exp(\eta_1 \sum_{i \sim j} \mathbb{I}_{x_i = x_j} \mathbb{I}_{z_i = z_j} + \eta_2 \sum_{i \sim j} \mathbb{I}_{z_i = z_j}) \ .$$

Figure 8 then illustrates the performance of our triplet models when segmenting synthetic images with non standard noise models (see also [8] for details).

[24] G. R. Cross and A. K. Jain. Markov Random Fields texture models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5(1) :25–39, 1983.

[25] W. Pieczynski and A. Tebbache. Pairwise Markov Random Fields and segmentation of textured images. *Machine Graph. Vision*, 9 :705–718, 2000.

[26] D. Benboudjema and W. Pieczynski. Unsupervised image segmentation using triplet Markov fields. *Computer Vision and Image Understanding*, 99(3) :476–498, 2005.

[27] D. Benboudjema and W. Pieczynski. Unsupervised statistical segmentation of non stationary images using triplet Markov Fields. *IEEE Trans. PAMI*, 29(8) :367–1378, 2007.

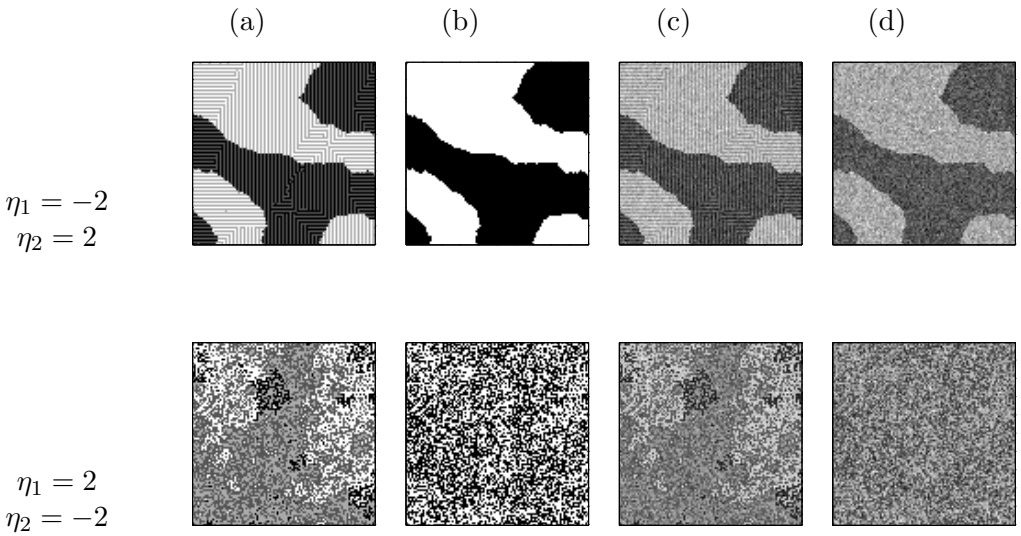


FIGURE 7 – Simulations of 2 parameter (η_1 and η_2) Triplet Markov fields defined in eq. (4) [8] when L = 2 and K = 2 with respectively $\eta_1 = -2, \eta_2 = 2$ (first row) and $\eta_1 = 2, \eta_2 = -2$ (second row) : (a) Realizations of $(\mathbf{X}, \mathbf{Z})$, (b) Realizations of $\mathbf{Z}$, (c) Realizations of $\mathbf{Y}$, (d) Realizations of a HMF-IN built by adding to images in (b) some Gaussian noise with 0 mean and standard deviation equal to 0.3. Note that in the images, each of the 4 possible values of (x_i, z_i) has been associated with a grey-level for visualization.

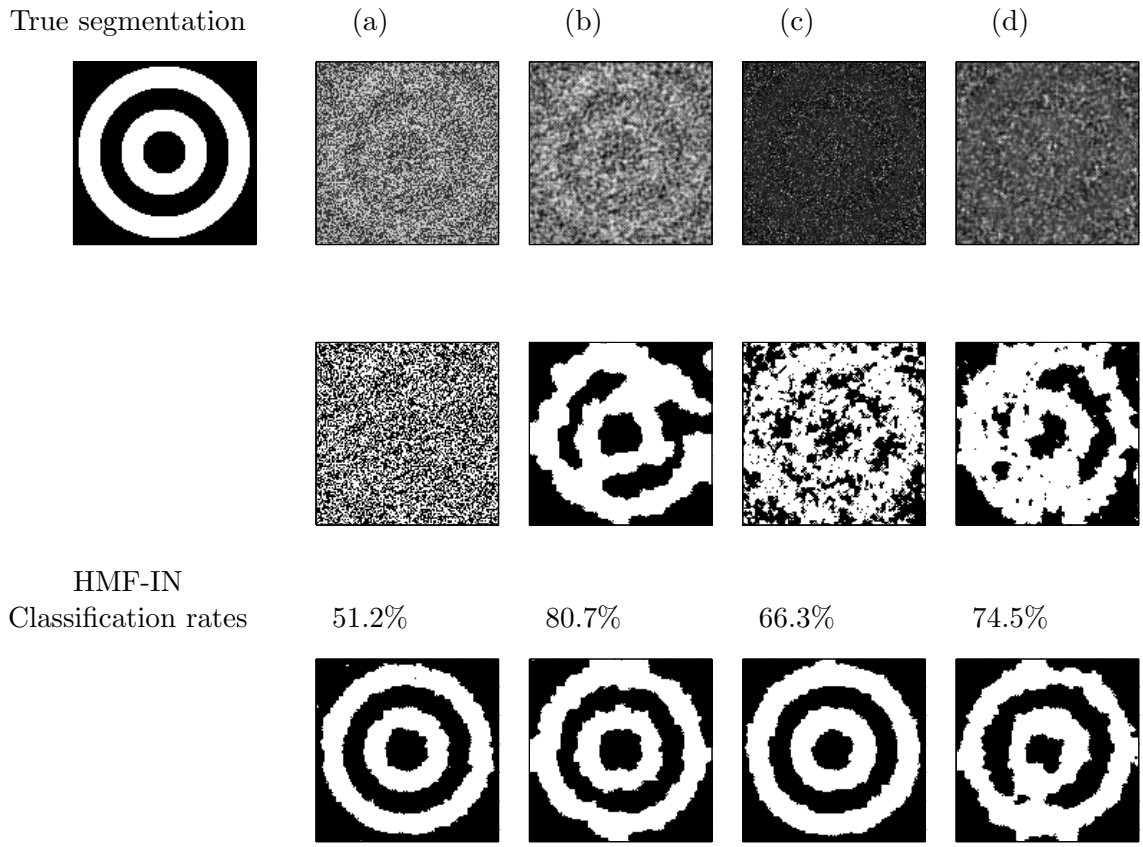


FIGURE 8 – Synthetic image segmentations using an HMF-IN model (second row) and our TMF model (third row) : the true 2-class segmentation is the image in the upper left corner and four different noise models are considered. In (a) class distributions are mixtures of two Gaussians, In (c) observations from class 1 are generated from a Gamma(1,2) distribution and observations from class 2 are obtained by adding 1 to realizations of an Exponential distribution with parameter 1. In (b) and (d) the noisy images are obtained by replacing each pixel value respectively in (a) and (c) by its average with its four nearest neighbors. Classification rates are given below each segmentation results. In the TMF model case, the selected L values using our BIC_{MF} criterion (see Section 3.3.1) are given in the last row.

Also it is important to specify the relationship between the Triplet Models and the Conditional Random Fields (CRF)[18] which have been widely and successfully used in applications including text processing, bioinformatics and computer vision. CRF's are *discriminative models* in the sense that they model directly the posterior or conditional distribution of the labels given the observations. Explicit models of the joint distribution of the labels and observations or of the noise distribution are not required. In classification issues, the posterior distribution is the one needed and it can appear as a waste of time and computational resources to deal with the joint distribution or with complex noise models. All the more so as the class conditional distributions describing the noise model may contain a lot of structure but with little effect on the posterior distribution (see for instance Figure 1.27 in C. Bishop's book[13]). However, even in classification contexts, approaches that model the joint distribution of the labels and observations are considered. They are known as *generative models*. Triplet Markov models belong to this class. Such generative models are certainly more demanding in terms of modeling but they have the advantage to provide a model of the observed data (the likelihood) allowing this way better access to theoretical properties of the estimators. This can also be useful for detecting outliers or data points that have low probability under the model and for which the predictions may be of low accuracy. Even when the main interest is classification, such outliers, if not detected, can severely bias the parameter estimates and affect the classification results. In addition, for more and more modern applications, learning data is not enough and the use of external information including a priori or expert knowledge is necessary. Such expertise is usually embedded in the data structure which can be taken into account by using generative models. Our aim was to investigate the use of Triplet Markov models seen as an alternative to Conditional Random Fields with good modeling capabilities. As generative models they better model the structure of the data. They can be used with standard Bayesian techniques and probabilistic clustering tools requiring no more algorithmic effort than CRF's. They allow theoretically well-based studies and in particular model selection to guide the user to specific modeling choices consistent with the observed data (see [60, 8]).

3.1.8 Comparing two types of models for clustering spatial data

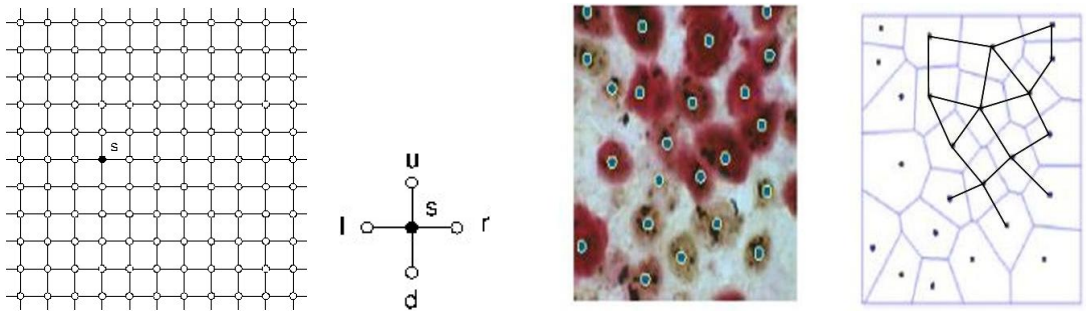
Related publication : 2 national conferences [66, 54].

This is joint work with Denis Allard from INRA, Avignon and Nathalie Peyrard from INRA Toulouse.

My expertise in Markov models includes the investigation of other related modeling solutions that offer similar processing advantages and possibly superior modeling capabilities. The pairwise and Triplet Markov random fields go in this direction and can

[18] S. Kumar and M. Hebert. Discriminative random fields. *Int. J. Comput. Vision*, 68(2) :179–201, 2006.

[13] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.



(a) (b) (c)

FIGURE 9 – Examples of different structures : (a) regular lattice and (b) a standard four nearest neighbor system, (c) biological image of cells and proposed structure based on Voronoi tesselation.

be related to the extensions we consider. With this in mind we focused on the problem of clustering or grouping observations measured at different points of the plane, taking into account the geographical distances between the points. We compared two models to solve this problem : hidden Markov random fields (HMRF) and a geostatistical model which uses Gaussian random fields. Geostatistical-type problems (Figure 9 (c)) are distinguished most clearly from lattice-type (Figure 9 (a) and (b)) problems by the ability of the spatial index to vary continuously. The goal was to study how methods from one class of problems could be borrowed from methods usually associated with another class. As regards parameter estimation we investigated an EM-like algorithm. If approximations of this algorithm exist for HMRFs (see Section 3.2.1), this is not the case for geostatistical models. We proposed [66, 54] an approach mimicing the solution in the independent case. Models performance was compared on simulated and real data : they showed the superiority of the geostatistical model in terms of classification error but the classifications obtained with the HMRF model were smoother and sometimes more satisfactory. Further work is required to better understand these possible extensions of standard HMRF.

3.2 The design and study of EM-like algorithms

As regards, estimation algorithms, non trivial hidden structures usually lead to non trivial optimization problems. I focused on a certain type of methods, appropriate for Markov structures, based on the mean field principle (Section 3.2.1) and proposed effective algorithms which show good performance in practice and for which I also studied theoretical properties (Section 3.2.2). To improve these algorithms in terms of computational efficiency and robustness to artefacts in the data, we also investigated distributed versions of EM using a multi-agent system paradigm (Section 3.2.3). We then also proposed an adaptation of the EM framework to cluster non standard multi-modal data (Section 3.2.4) and addressed the problem of slow convergence in the standard mixture model case (Section 3.2.5).

3.2.1 EM procedures using mean field-like approximations

Related publications : 1 journal paper [14], 1 national conference [56].

I initiated this direction of research in 1998 with the start of Nathalie Peyrard PhD thesis that I advised from 1998 to 2001. The tools developed during these years are still on focus today as regards their theoretical properties and practical efficiency. They are also today in our team the basis of more recent works.

The EM algorithm has an elegant formulation and when it is applied to appropriate model structures it yields parameter update procedures that are easy to derive and straightforward to implement. However, outside simple or standard cases, the EM algorithm yields update procedures that do not have closed form expressions and it is seldom tractable analytically. In particular, when focusing on image segmentation and Markov random fields estimation, difficulties arise due to the dependence structure in the models and approximations are required. A heuristic solution using mean field approximation principle has been proposed by J. Zhang[28]. The mean field approach consists of calculating quantities related to a complex probability distribution, by using a simple tractable model such as the family of independent distributions. Using ideas from this principle, we proposed [14] in the context of Markovian image segmentation a class of EM-like algorithms generalizing the work of J. Zhang which show good performance in practice.

The idea underlying these algorithms is to replace the intractable Markov distribution by a simpler distribution obtained by fixing the neighbors of each pixel to constant values. Then, an iteration of a mean field-like algorithm consists of two steps : in the first step the values for the neighbors are updated according to the observations and to the current value of the parameter. It follows an approximation of the intractable Markov distribution. The second step consists of carrying out the EM algorithm for the corresponding approximated observed likelihood to obtain an updated value of the parameter. Mean field-like algorithms can thus be related to the EM algorithm for

[28] J. Zhang. The Mean Field Theory in EM Procedures for Markov Random Fields. *IEEE Transactions on Signal Processing*, 40(10) :2570–2583, 1992.

independent mixture models, with the significant difference that the mixture model adaptively changes at each iteration depending on the current choice of the neighbors values. In [14], we compared three different ways of updating the neighbors in the first step : the mean field approximation of the conditional mean (*mean field algorithm*), an approximation of the conditional mode (*mode field algorithm*) and a simulated realization of the conditional Gibbs distribution obtained with the Gibbs sampler of Geman and Geman[29] (*simulated field algorithm*). The last choice led to a new stochastic algorithm which appeared to be the most promising for its good performance on synthetic and real image experiments. As mentioned in many sections of the present document, we are working on extending the modeling capabilities of these kind of algorithms and on providing a better understanding of their properties.

Most of my work is currently based on the mean field approximation principle which is among the variational methods [30], the simplest principle although not yet fully understood in some aspects (see for instance Section 3.2.2). We observe that in many practical applications, the mean field approximation already handles a good part of the complexity of the data. There is a trade-off then between finer approximations which may not always lead to much better (classification or estimation) results and additional computational burden. Also, very few results exist on the quality of the variational approximations when they act as a surrogate in a larger inference problem. In particular, as regards parameter estimation, results on bounds on the likelihood may not be of great help. Mean field methods can be related to *message passing* algorithms which correspond to general schemes for fitting variational approximations. Message passing algorithms include Mean Field, Loopy Belief Propagation, Expectation Propagation, Tree-reweighted message passing, Fractional Belief Propagation, Power Expectation Propagation. A unifying view and references for these different methods can be found in a report by T. Minka[31]. It appears that the difference between mean field methods and belief propagation methods is not the amount of structure they model but the measure of loss they minimize. Neither method is inherently superior. It depends on the task and goal in mind. To our knowledge, there exists no experimental comparison of the various algorithms performance on real-world networks and data.

Apart from variational methods, recent efficient techniques such as *Graph cuts* [32] are built to provide maximum a posteriori (MAP) solutions. However, in our framework they do not seem to provide the same flexibility as probabilistic techniques when parameter estimation is required. They are based on *hard membership* which allows the use of the graph cut optimization methods that have proven highly effective for solving pixel labeling problems in some cases. We chose to base our work mainly on *soft membership* adopting a more statistical point of view with some emphasize on parameter

[29] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 6 :721–741, 1984.

[30] M.I. Jordan, editor. *An introduction to variational methods for graphical models*. MIT Press, 1999.

[31] T. Minka. Divergence measures and message passing. *Microsoft Research Technical Report*, MSR-TR-2005-173, Decembre, 7 2005.

[32] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE trans. PAMI*, 23(11) :1222–1239, 2001.

estimation and in particular on the biais inherent to hard membership methods. To our knowledge comparison of MRF techniques and graph cut techniques are not always fair in the sense that they ignore the latest developments in Markov modeling.

3.2.2 MCVEM : combining Monte-Carlo simulations and variational EM

Related publications : 1 journal paper [10], 1 research report [61].

This is joint work with Gersende Fort, CNRS research scientist at LTCI TELECOM-ParisTech, Paris.

For the standard EM algorithm, parameter estimates yield increasing likelihood over the observed data and the convergence behavior of this process is well understood. However, since it is often the case that there are no other feasible choices than to resort to the mean field approximation in practical situations, it appears frequently that the mean field approximation is being used for practical problems with little consideration of important issues such as accuracy of the approximation, convergence of the algorithms and so on. As a matter of fact, in the context of Markovian segmentation, theoretical results as regards convergence properties are still missing. Convergence properties of related EM variants (GAM for Generalized Alternating Minimization) have been studied by W. Byrne, A. Gunawardana[33] and R. Neal and G. Hinton[34] but these variants cannot be applied in the MRF segmentation framework and further approximations are required. We proposed [10, 61] a new algorithm that we called MCVEM for Monte Carlo Variational EM, which is tractable in practice and for which we prove convergence results. Our algorithm has the advantage over the GAM procedures that it can be applied to perform image segmentation tasks and this on the basis of theoretical convergence results. The basis of our work is the paper by G. Fort and E. Moulines[35] which focuses on the convergence properties of the MCEM algorithm. Using similar tools, our key idea is to view the MCVEM algorithm as a stochastic perturbation of a deterministic algorithm, so called Variational EM (VEM) that is easier to study[33]. Experiments on synthetic and real images show that the algorithm performance is very closed and sometimes better than that of [14]. Additional good properties due to its stochastic nature need to be further investigated. This first effective step opens the way to a better understanding of the behavior of a lot of Markov based algorithms (see Figure 10 for an illustration of practical implementation issues). In particular, analyzing how simulation steps should be incorporated so as to interact advantageously with

[33] W. Byrne and A. Gunawardana. Convergence theorems of Generalized Alternating Minimization Procedures. *Journal of Machine Learning Research*, 1 :1–48, 2004.

[34] R.M. Neal and G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variants. In M.I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. MIT Press, 1998.

[35] G. Fort and E. Moulines. Convergence of the Monte-Carlo EM for curved exponential families. *Annals of Statistics*, 31(4) :1220–1259, 2003.

[33] W. Byrne and A. Gunawardana. Convergence theorems of Generalized Alternating Minimization Procedures. *Journal of Machine Learning Research*, 1 :1–48, 2004.

deterministic approximations seems promising [10].

3.2.3 Distributed Cooperative EM for Markov model-based segmentation

Related publications : 2 journal papers [7, 6], 1 book chapter [4].

This work was carried out during the PhD thesis of Benoit Scherrer in collaboration with Catherine Garbay from Laboratoire d'Informatique de Grenoble (LIG) and Michel Dojat from Grenoble Institut of Neuroscience.

In a different context, I also considered variants of the so-called *Incremental EM*[34] and *Variational EM*[33] which were implemented using multi-agent systems. This distributed version of EM is detailed in [7] and used in our MRI applications [6, 4] (Section 4.1.3). Classically, when considering Markov model-based image segmentation, model parameter estimation, in addition to leading to a large number of computations, is performed over the whole image, and does not reflect local image properties. We proposed to "agentify" the global MRF segmentation model by distributing a set of local MRF models within a multi-agent framework. The image is partitioned in territories where situated agents estimate the parameters of a local MRF model (see Figure 11). The global consistency of local models is ensured by modifying the EM scheme to introduce cooperation between neighboring agents. Local models estimated with our distributed local EM (DILEM) then accurately fit local image features. A priori knowledge can also be integrated into this model via the external field of the MRF models, and can be used to introduce other levels of agent cooperation. The modification of the standard EM procedure (Figure 12 (a)) leads to the introduction of additional steps and agent interactions (Figure 12 (b)).

We considered two kinds of agents : one global agent which initializes the MRF agentified segmentation and a set of local cooperating MRF segmentation agents. Local segmentation agents estimate local MRF models via local EM procedures and cooperate to ensure a global consistency of local models. We demonstrated different types of cooperations between agents that lead to additional levels of regularization compared to the standard label regularization provided by MRF. Embedding Markovian EM procedures into a multi-agent paradigm showed interesting properties that were illustrated on magnetic resonance (MR) brain scan segmentation (Section 4.1.3).

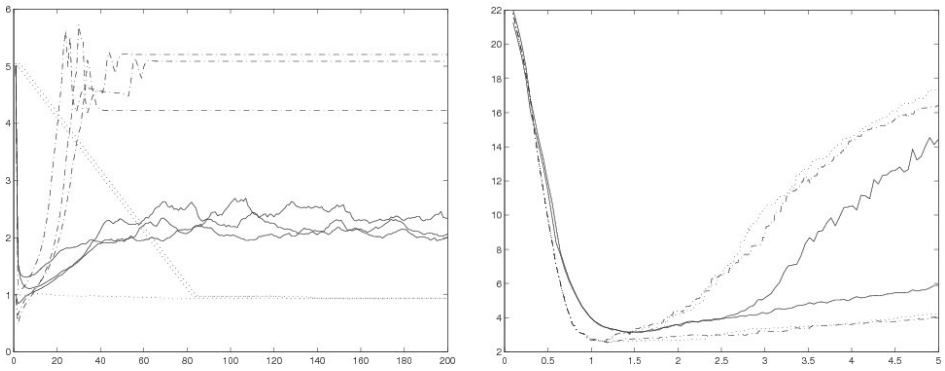
3.2.4 Conjugate EM for multimodal data clustering

Related publication : 1 journal paper [1].

This part has been developed during the PhD thesis of Vasil Khalidov in the context of the European project POP. This is also joint work with Radu Horaud from team

[34] R.M. Neal and G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variants. In M.I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. MIT Press, 1998.

[33] W. Byrne and A. Gunawardana. Convergence theorems of Generalized Alternating Minimization Procedures. *Journal of Machine Learning Research*, 1 :1–48, 2004.



(a) (b)

FIGURE 10 – Estimation and effect of the interaction parameter η in standard hidden Potts model (HMRF) : Figure (a) shows the η estimates trajectory versus the number of iterations for different parameter starting values, with *Mean Field* (dot line), *Simulated Field* (solid line) and *MCVEM* (dash-dot line). It appears that the starting value is crucial for the limiting behavior of *Mean Field*. The trajectories of *Simulated Field* do not converge to some fixed limiting value but the behavior of the different trajectories is similar. We believe that convergence of the *Simulated Field* algorithm has to be understood in a different way. An approach similar to what is done for the so-called stochastic EM algorithm is more appropriate. Averages of the parameters should converge and this suggests to replace the current implementation of *Simulated Field* algorithm by an averaging procedure. Note that for this data set, despite the variations in the estimation of the interaction parameter, the corresponding segmentations were quite stable : the mean error rate is in the range $(2.86\%, 2.92\%)$ for *MCVEM*, $(2.82\%, 3.10\%)$ for *Mean Field* and $(3.42\%, 3.65\%)$ for *Simulated Field*. Figure (b) shows the classification error rate versus η obtained by *Mean Field* (dot line), *Simulated Field* (solid line) and *MVCEM* (dash-dot line), when the segmentation algorithm is started from two different initial classifications. For large values of η, the segmentation is greatly dependent of the initial segmentation. In addition, the curves give an idea of the value that corresponds to the minimum error rate. For *MCVEM* and *Simulated Field*, this naive computation is not far from the estimates obtained by running the full algorithms when all the parameters are unknown. This is not the case for *Mean Field*, thus showing that the *Mean Field* segmentation may depend on the implementation of the algorithm.

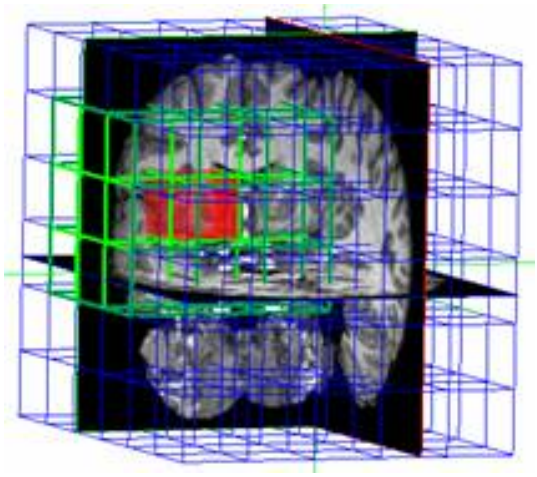


FIGURE 11 – Distributed segmentation : the volume of the 3D brain image is partionned into a number of sub-volumes on which agents are going to act locally. The global consistency of the resulting local treatments is then ensured through cooperations. A Markovian regularization takes place via cooperation of each agent (in red) with its neighbors (in green).

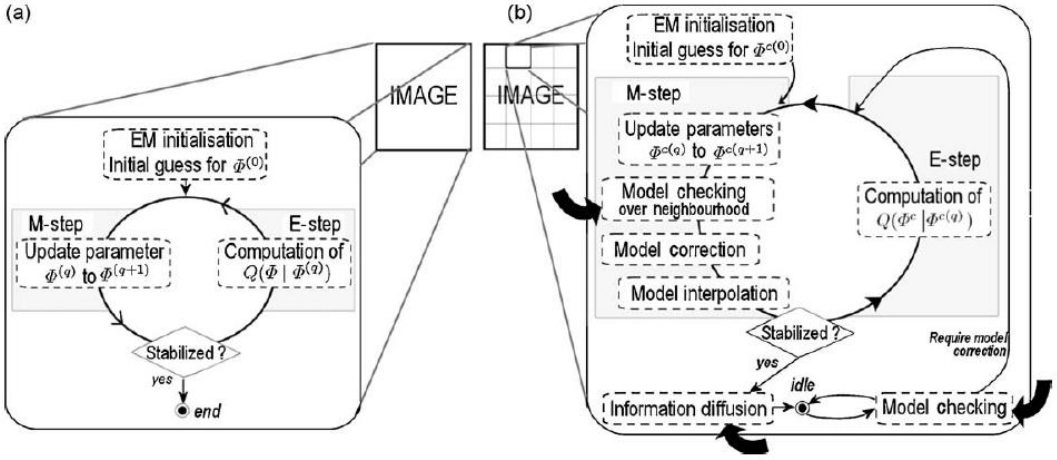


FIGURE 12 – The left Figure (a) shows a synthetic view of a classical global EM estimation scheme. The right Figure (b) shows the agentified distributed local EM (DILEM) scheme with cooperations and coordinations (bold arrows) to ensure a global consistency of local models.

PERCEPTION.

In Section 3.1.2, I mentioned the problem of clustering multi-modality data and presented a conjugate mixture models framework that provides consistent clustering results in multiple feature spaces. In [1], we formulated this problem as a likelihood maximization task. We proposed a variant of the EM algorithm specifically designed to estimate object-space parameters that are indirectly observed in several sensor spaces. We explicitly derived the expectation and maximization steps of the associated EM algorithm. While the E-step of the proposed algorithm is standard, the M-step implies non-linear optimization of the expected complete-data log-likelihood with respect to the object parameters. We investigated efficient local and global optimization methods. More specifically, we prove that, provided that the object-to-sensor functions as well as their first derivatives are Lipschitz continuous, the gradient of the expected complete-data log-likelihood is Lipschitz continuous as well. The immediate consequence is that a number of recently proposed optimization algorithms specifically designed to solve Lipschitzian global optimization problems[36] can be used within the M-step of the proposed algorithm. Several of these algorithms combine a local maximum search procedure with an initializing scheme to determine, at each iteration, good initial values from which the local search should be performed. This implies that the proposed EM algorithm has guaranteed convergence properties. We then discussed several possible local search initialization schemes, leading to different convergence speeds. We proposed and compared two possible strategies to initialize the EM algorithm. The number of objects is determined using the BIC criterion (see also Section 3.3.2).

3.2.5 Accelerated EM for finite mixtures

Related publication : 1 journal paper [16].

This work has been carried out with G. Celeux, S. Chrétien and A. Mkhadri when A. Mkhadri was visiting the former IS2 *team at INRIA and S. Chrétien was a post-doc in the former* NUMOPT *team.*

One of the most documented problems occurring with EM is its possible slow convergence. In the finite independent mixture case, we proposed and studied the so-called CEMM : component-wise EM for Mixtures algorithm [16] which aims at overcoming the problem of slow convergence.

In its general formulation, the EM algorithm involves the notion of a complete data space, in which the observed measurements and incomplete data are embedded. An advantage is that many difficult estimation problems are facilitated when viewed in this way.

One drawback is that the simultaneous update used by standard EM requires overly informative complete data spaces, which leads to slow convergence in some situations. In the incomplete data context, it has been shown that the use of less informative

[36] A.A. Zhigljavsky and A. Žilinskas. *Stochastic Global Optimization.* Springer, 2008.

complete data spaces, or equivalently smaller missing data spaces, can lead to faster convergence without sacrifying simplicity.

In [16], we proposed a component-wise EM for mixtures. It uses, at each iteration, the smallest admissible missing data space by intrinsically decoupling the parameter updates. Monotonicity is maintained, although the estimated proportions may not sum to 1 during the course of the iteration. However, we proved that the mixing proportions satisfied this constraint upon convergence. Chrétien and Hero[37] showed that the EM procedure could be recast into a proximal point framework and our proof of convergence relies on the interpretation of our procedure as a proximal point algorithm. .

For performance comparison, we considered standard EM as well as two other algorithms based on missing data space reduction, namely the SAGE[38] and AECME[39] algorithms. We provided adaptations of these general procedures to the mixture case. We also considered the ECME algorithm[40], which is not a data augmentation scheme but still aims at accelerating EM. The numerical experiments reported in [16] illustrate the advantages of the component-wise EM algorithm relative to these other methods.

[37] S. Chretien and A. O. Hero. Acceleration of the EM algorithm via proximal point iterations. In *IEEE International Symposium on Information Theory*, MIT Boston, 1998.

[38] J. A. Fessler and A. O. Hero. Space-Alternating generalized expectation-maximisation algorithm. *IEEE Trans. Signal Processing*, 42 :2664–2677, 1994.

[39] X.-L. Meng and D. A. van Dyk. The EM algorithm - an old folk song sung to a fast new tune (with discussion). *J. Roy. Stat. Soc. Ser. B*, 59 :511–567, 1997.

[40] C. Liu and D.B. Rubin. The ECME algorithm : A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81 :633–648, 1994.

3.3 Hidden structure model selection

Choosing the probabilistic model that best accounts for the observations is an important first step for the quality of the subsequent statistical inference and analysis. In most cases the choice is done subjectively using expert knowledge or *ad hoc* procedures and there is a striking lack of systematic data-based approaches. When recasting this choice as a problem of probabilistic model comparison, most selection criteria involve calculating integrated likelihoods for a number of models, *i.e.* the likelihoods of the observations integrated over the respective model parameters. For a lot of models of interest, these integrated likelihoods are high dimensional and intractable integrals so that most available software is generally inefficient for their evaluation. Various approximations have been proposed. In particular the Bayesian Information Criterion (BIC) approximation of Schwarz[41] is based on the Laplace method for integrals. Many other approaches can be found in the literature on model selection (see for instance the list of references in Kass and Raftery's paper[42]). BIC has become quite popular due to its simplicity and its good results in cases where other standard model selection procedures were not satisfactory.

In the HMRF context, I proposed [15] variational approximation tools to deal with the HMRF selection issue in practice using the Bayesian Information Criterion (BIC) (Section 3.3.1). Then, we also developed a so-called *Conjugate BIC* (Section 3.3.2) appropriate in the context of our conjugate EM algorithm mentioned in Section 3.1.2. At a different level, I also investigated the deviance information criterion (DIC) which is not so naturally defined for missing data models (Section 3.3.3).

3.3.1 Approximations for selecting hidden Markov models

Related publication : 1 journal paper [15].

This section reports on research done with Nathalie Peyrard during her PhD 1998-2001.

Our aim was to apply variational approximation tools[30] to built and select a model from a collection of hidden Markov random fields in the context of spatial data clustering and in particular image segmentation. In these situations, exact calculation of selection criteria is not possible and simulation methods such as Monte Carlo Markov Chains (MCMC) methods reach their limits. In [15], we focused on the use of BIC[41] and proposed two ways to use variational approximations to get tractable criteria. We illustrated the performance of these approximations on the issue of selecting the number of classes and compared with another criterion called PLIC[43] for *Pseudo Likelihood*

[41] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6 :461–464, 1978.

[42] R. Kass and A. Raftery. Bayes factor. *Journal of the American Statisticial Association*, 90 :733–795, 1995.

[30] M.I. Jordan, editor. *An introduction to variational methods for graphical models*. MIT Press, 1999.

[43] D.C. Stanford and A.E. Raftery. Approximate bayes factors for image segmentation : the pseudolikelihood information criterion (plic). *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(11) :1517 – 1520, nov. 2002.

Information Criterion. We obtained generally better results with our so called BIC^{MF} but its theoretical properties remain to be investigated. In this case, we have no specific results on the quality of BIC as an approximation of the integrated likelihood and this choice as a selection criterion is arguable. However, the question of the criterion ability to asymptotically choose the correct model can be addressed independently of the integrated likelihood approximation issue. As an illustration, E. Gassiat[44] proved that for the more specialized but related case of hidden Markov chains, under reasonable conditions, the *maximum penalized marginal likelihood* estimator of the number of hidden states in the chain is consistent. This estimator is defined for a class of penalization terms that includes the BIC correction term and involves an approximation of the maximized log-likelihood which is not necessarily good, namely the maximized log-marginal likelihood. In particular, this criterion is consistent even if there is no guarantee that it provides a good approximation of the integrated likelihood. This suggests that a good approximation of the maximized log-likelihood is not a strong requirement to obtain consistent criteria. A key point in E. Gassiat's paper[44] seems to be the decomposition of the criterion as a sum of identically distributed terms. The criteria proposed in our paper [15] can also be written as a sum because of the factorization property of the distributions involved. The generalization is not straightforward but a next step could therefore be to investigate if consistency results can be deduced in a similar way.

3.3.2 Information criteria for clustering multimodal data

This work has been carried out with Vasil Khalidov and is reported in Chapter 6 of his PhD thesis[45]*. No other published document is available yet.*

As already presented in Section 3.1.2, a multimodal data setting is a combination of multiple data sets each of them being generated from a different sensor. The data sets live in different physical spaces with different dimensionalities and cannot be embedded in a single common space. Clustering such multimodal data raises the question of how to perform pairwise comparisons between observations living in different spaces. We have proposed a solution within the framework of Gaussian mixture models and the EM algorithm in [1]. Each modality is associated to a modality specific Gaussian mixture which shares with the others a number of common parameters $\{o_1, \ldots, o_K\}$ in equation (2) of Section 3.1.2 and a common number of components (K). Each component corresponds to a common multimodal event that is responsible for a number of observations in each modality.

As this number of components is usually unknown, we have proposed information criteria, based on a penalized maximum likelihood principle, for selecting this number from the data. A consistency result for the estimator of the common number of components is given under some assumptions. To our knowledge, there has been no procedure

[44] E. Gassiat. Likelihood ratio inequalities with application to various mixtures. *Annales de l'institut Poincaré*, 2002.

[45] V. Khalidov. *Conjugate mixture models for the modelling of visual and auditory perception*. PhD thesis, University of Grenoble, October 2010.

so far that properly selects the model dimensionality for multimodal case in a consistent manner. Standard results on information criteria are shown for identically distributed data which is typically not the case in the multimodal setting. In this contribution, we have been able to derive an appropriate information criterion with a BIC like penalty and illustrate the performance of this *conjugate BIC* score on a challenging two modality task of detecting and localizing audio-visual objects.

3.3.3 Deviance Information Criteria for missing data models

Related publication : 1 journal paper [13].

This work was initiated in the former IS2 *team with G. Celeux now from team* SELECT*, INRIA futur, when Mike Titterington from University of Glasgow in Scotland was visiting* IS2 *at INRIA Grenoble. It resulted in a common publication [13] with G. Celeux, M. Titterington and Christian Robert, from* CEREMADE*, Paris Dauphine.*

The deviance information criterion (DIC) has been introduced by Spiegelhater et al.[46]. The advantage of DIC over other criteria in the case of Bayesian model selection is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. AIC and BIC require calculating the likelihood at its maximum over the possible parameter value θ, which is not readily available from the MCMC simulation. But DIC follows directly from simple approximations that consists of computing the average of the deviance (minus 2 log-likelihood) over the samples of θ, and the value of this deviance evaluated at the average of the samples of θ.

The DIC is directly inspired by linear and generalized linear models, but it is not so naturally defined for missing data models. In [13], we have considered different possible variations depending in particular whether or not the missing variables are treated as parameters. We have reassessed the criterion for such models, and compared different DIC constructions, testing the behavior of various extensions in the cases of independent mixture and random effect models. We have shown that the deviance information criterion of Spiegelhater et al. and the corresponding effective dimension allow for a wide range of interpretations and extensions outside exponential families, as was already apparent from the published discussion of Spiegelhater et al.'s paper. What we have found in addition through theoretical and experimental studies is that some of these extensions, while as "natural" as the others, are simply not adequate for evaluating the complexity and fit of a model, either because they give *negative* effective dimensions or because they exhibit too much variability from one model to the next. While Spiegelhater et al. argue that negative p_D's are indicative of a possibly poor fit between the model and the data, there is no explanation of that kind in our cases : for the same data and the same model, some DICs are associated with positive p_Ds and others are not.

As illustrated by the associated discussion in [13], a lot of questions and issues are

[46] D. J. Spiegelhalter and al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, series B*, 64 :1–34, 2002.

still open regarding the use of DIC as a measure of complexity and a well-grounded criterion.

4 Application domains

In this section more details are given for the various applications mentioned in the text. Most Markov models applications arise naturally from image analysis (**Section 4.1**. Other non image data applications are related to biology and computational neuroscience (**Section 4.2**).

4.1 Image Analysis

Several areas of image analysis can be covered using the tools I developed. During the ACI Movistar project 2003-06 (see Section 6 in the detailed CV part), we addressed in collaboration with team LEAR at INRIA Grenoble Rhône-Alpes, issues about object and class recognition and about the extraction of visual information from large image data bases (Section 4.1.1). I also worked in the context of the European project POP (see Section 6 of my detailed CV) on the consistent estimation of disparities and boundaries or surface normals from stereoscopic images (Section 4.1.2).

Other applications in medical imaging are natural. I worked on brain MRI data in collaboration with the Grenoble Institute of Neuroscience (GIN), team VISAGE at INRIA Rennes and team MAGMA of the Computer Science Lab. in Grenoble (LIG) (Sections 4.1.3 and 4.1.4). My investigations extended to functional MRI with the GIN and Laboratoire de Neuroimagerie Assistée par Ordinateur (LNAO) from NeuroSpin CEA in Saclay (Section 4.1.5). I also dealt with breast MRI data in collaboration with the Statistics Department of University of Washington, Seattle (Section 4.1.6).

I then considered other statistical 2D fields coming from other domains such as remote sensing, in collaboration with Laboratoire de Planétologie de Grenoble in the context of the ANR Vahine project 2008-11 (see Section 6 of my detailed CV).

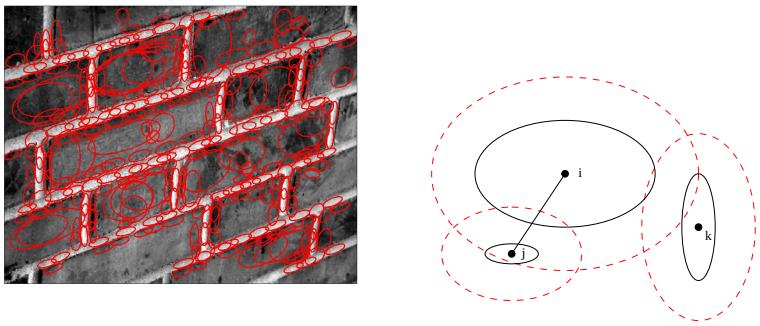
4.1.1 Triplet Markov Random Fields for recognizing textures

Related publications : 1 journal paper [8], 3 international conferences [44, 45, 40], 2 national conferences [52, 53], 1 research report [60].

In the context of Juliette Blanchet's PhD thesis, this is joint work with Cordelia Schmid, LEAR*, INRIA Grenoble Rhône-Alpes.*

We proposed a new probabilistic framework for recognizing textures in images. Images are here described by local affine-invariant descriptors such as SIFT descriptors[1] (Figure 1) and by spatial relationships between these descriptors. See an illustration in Figure 13. We then proposed to introduce in texture recognition the use of statistical parametric models of the dependence between descriptors. We chose hidden Markov models and followed the method described in Section 3.1.7 involving Triplet MRF. Using sample images, which have been gathered over a wide range of viewpoints and

[1] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2) :91–110, 2004.



(a) (b)

FIGURE 13 – Spatial relationships between descriptors : (a) detected interest points (ellipse centers) and associated scales (ellipse sizes and orientations) using Laplace detector and SIFT descriptors[1], (b) construction of a neighborhood system between the detected points : points within the range of an enlarged ellipse are connected to the ellipse center.

scale changes, textures are then learned as hidden MRF and a set of estimated parameters is associated to each texture. In our data set there were $K = 7$ texture classes illustrated in Figure 14. Using Triplet models, we introduced Gaussian subclasses for each texture class so as to allow a better modeling of the textures which cannot be capture by single Gaussian distributions. For simplicity, the number of subclasses to describe each class distribution was set to $L = 10$ for each texture. Selecting L using BIC is also possible, but in this application we did not observe significantly better recognition results. At recognition time then, another Markov model is used to compute, for each feature vector, the membership probabilities to the different texture classes. Experiments showed very promising results (*e.g.* [8]) as illustrated in Figure 15.

4.1.2 Depth recovery from stereo matching using coupled Markov fields

Related publications : 2 international conferences, [35, 25], 1 research report [59].

This is part of Ramya Narasimha's PhD thesis and is joint work with Elise Arnaud, Miles Hansard and Radu Horaud from team PERCEPTION*, INRIA. This was part of the European STREP POP (Perception On Purpose).*

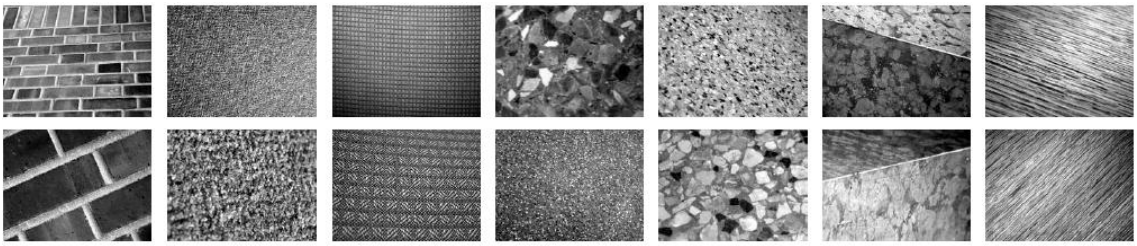


FIGURE 14 – Texture sample : two different images (rows) of each of the 7 texture classes (columns).

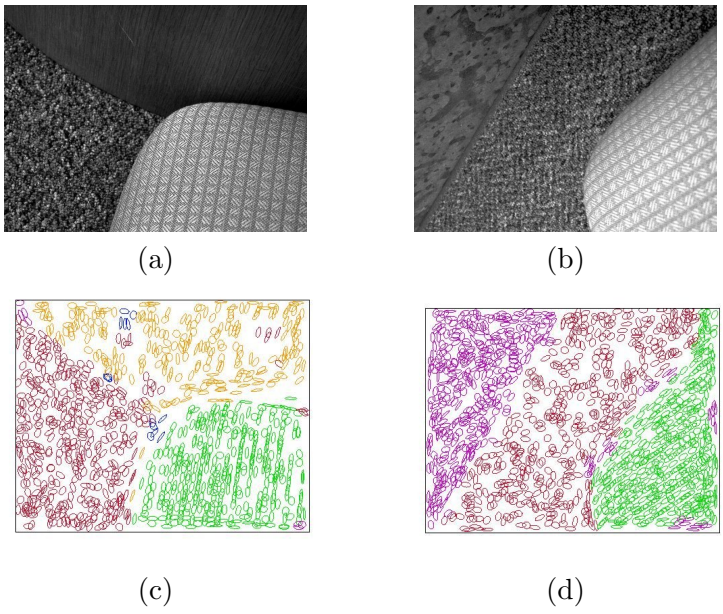


FIGURE 15 – Texture recognition using Triplet Markov Models : (a) and (b) original multi-texture images (three textures are visible in each), (c) and (d) classification results (ellipses represent interest points and associated regions). The different colors correspond to different texture assignments.

A typical stereo camera system captures a scene from two different view points (left and right). The goal is then to find the locations in both images that correspond to the same physical points in space. With this information, along with the geometry of the stereo set up, it is possible to determine the three-dimensional locations of all the points in the image. In this work, we focused on the so-called stereo correspondence issue. It consists of determining the locations in each camera image that are the projection of the same physical point in 3D space. After camera rectification, the search for the corresponding points can be reduced to finding disparities, *i.e.* differences in the x-coordinates of the corresponding left and right pixel locations. This disparity is related to the depth in 3D space.

Methods for stereo correspondence include then techniques that use Markov random fields to specify the local interaction between the disparities. They are used as priors which encode the smoothness of the disparities across locations in space. The cost from the stereo image intensities is introduced as the likelihood. In such approaches the objective then is to maximize the posterior probability over all possible disparity maps. However, even though such modeling techniques capture the local interactions between the neighbouring disparities and incorporate the stereo image information, some crucial problems still remain :

1) Some areas in the scene that are visible in one image may be occluded in the other and this can lead to incorrect matches.

2) Regularization term in the model could smooth over all disparities and lead to poor solutions at the object boundaries.

3) Incorporating just the stereo image intensities and smoothness term would model disparities which may not be consistent with the geometric properties of the surface.

In order to tackle these issues some extra information or constraints are required in modeling the correspondence problem. Monocular cues such as gradient, edges or color information pertaining a single image could be used within the model to provide better solutions for disparity. In addition some extra geometric constraints have to be incorporated to obtain surface-consistent solutions for the disparity. In this context, we worked on two approaches :

- **Cooperative disparity estimation and object boundary extraction.**
  The first method proposes to cooperatively estimate disparities and object boundaries in a joint probabilistic framework. The idea here is to tackle the problem of localizing discontinuities in disparity which correspond to the object boundaries in the real world, along with that of disparity estimation. This scheme involves incorporation of gradient information from a single image as a monocular cue. While the disparities are detected using the stereo cue (the left and the right images), the monocular cues help in correcting the disparity at the discontinuities and finding the object boundaries. We modelled both the stereo and monocular cues within a joint MRF-framework. Further details can be found in [35]. An illustration is given in Figure 16.
- **Estimating surface consistent disparities .**
  The second method incorporates surface differential geometric constraints into

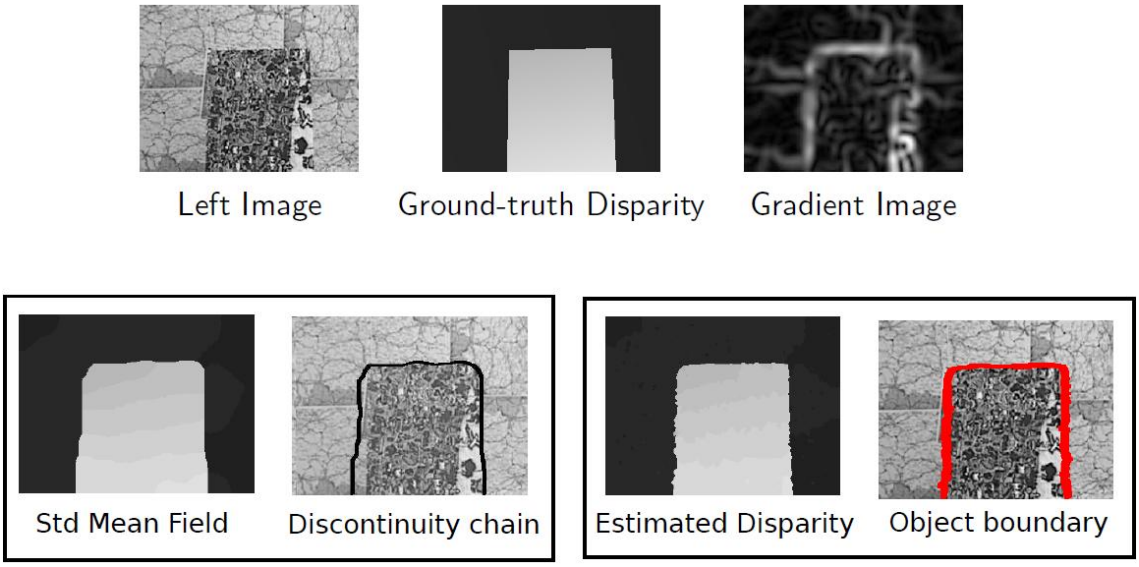


FIGURE 16 – Cooperative disparity estimation and object boundary extraction : our method (bottom right) improves over standard mean field (bottom left) in which boundaries are improperly localized.

the disparity model. These constraints are derived from the surface normals in disparity space. The idea is to model the disparities in such a way that they lie on the plane defined by the surface normals. This constraint leads to solutions that are consistent with the surface geometric properties of the scene. The idea is to simultaneously estimate the disparity and surface normals, considering explicitly the influence of one on the other. This is done by modeling both the disparities and the normal in a joint framework. This work corresponds to [25, 59].An illustration is given in Figure 17.

The major challenge in both of the above mentioned methods is to incorporate these cues and constraints within a single joint probabilistic setting, in which the relationship between the disparities and the variables under consideration (object boundary or surface normals) can be explicitly established. In this regard we used the idea of coupled Markov random fields, which permit the influence of one variable on the other to be made explicit within the model. Such a probabilistic set up also allows for separate optimization techniques to be used for maximizing the posterior distributions pertaining to each of the variables, providing further flexibility in modeling and optimization. An Alternating Maximization procedure was then used to achieve the overall optimization.

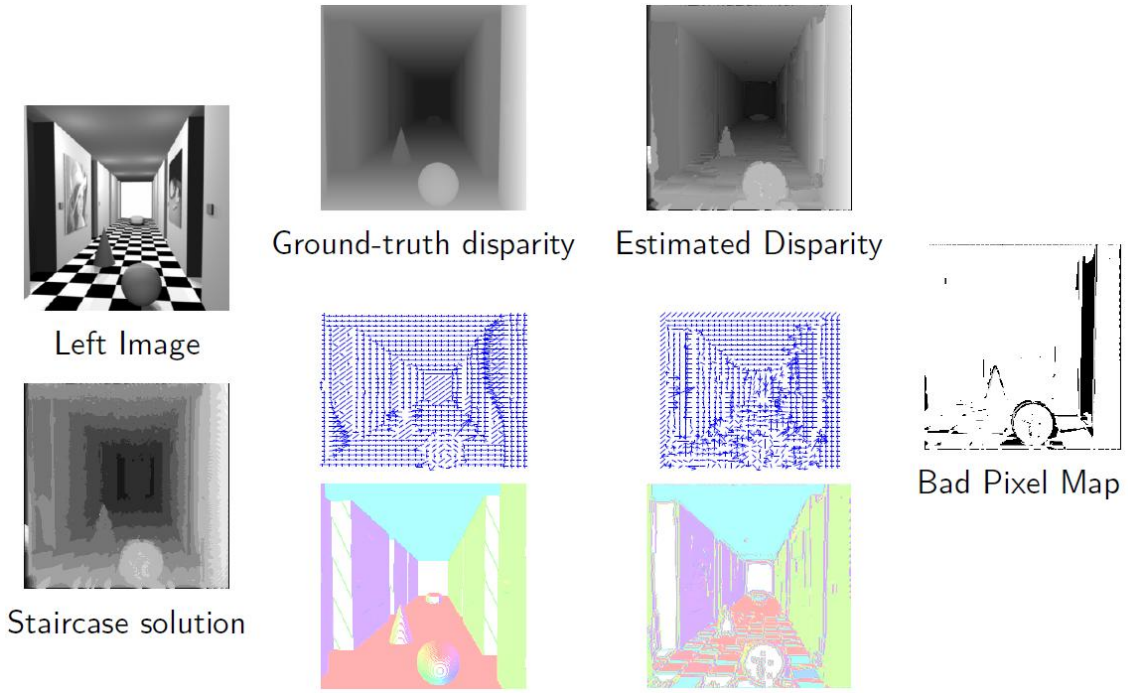


FIGURE 17 – Estimating disparity for slanted and curved surfaces : the improvement over the staircase solution that does not use normals is visible.

4.1.3 Distributed and cooperative Markov modeling for local atlas registration and robust brain tissue and structure segmentation

Related publications : 2 journal papers [7, 6], 1 book chapter [4], 5 international conferences [28, 31, 36, 37, 41], 2 national conferences [48, 49].

This is joint work with Benoit Scherrer, Michel Dojat and Catherine Garbay from Grenoble Institut des Neurosciences and team MAGMA *of Laboratoire d'Informatique de Grenoble (LIG).*

Difficulties in automatic MR brain scan segmentation arise from various sources. The nonuniformity of image intensity results in spatial intensity variations within each tissue, which is a major obstacle to an accurate automatic tissue segmentation. The automatic segmentation of subcortical structures is a challenging task as well. It cannot be performed based only on intensity distributions and requires the introduction of a priori knowledge usually encoded via a pre-registered atlas. Most of the proposed approaches share two main characteristics. First, tissue and subcortical structure segmentations are considered as two successive tasks and treated relatively independently although they are clearly linked : a structure is composed of a specific tissue, and knowledge about structures locations provides valuable information about local intensity distribution for a given tissue. Second, tissue models are estimated globally through the entire volume and then suffer from imperfections at a local level. Alternative local procedures exist but are either used as a preprocessing step or use redundant information to ensure consistency of local models.

In a first approach we proposed [6, 7, 36, 37, 41], we obtained good results using an innovative local and cooperative approach. It performs tissue and subcortical structure segmentation by distributing through the volume a set of local Markov random field (MRF) models which better reflect local intensity distributions. Local MRF models are used alternatively for tissue and structure segmentations. For parameter estimation, we used the algorithms we proposed in [14] (see Section 3.2.1) based on EM and variational approximations. They showed interesting results for tissue segmentation but were not sufficient for structure segmentation without introducing a priori anatomical knowledge (atlas). In this first attempt, Brain anatomy was then described by fuzzy spatial relations between structures that express general relative distances, orientations or symmetries. This knowledge was incorporated into a 2-class Markov model via an external field. This model was used for structure segmentation. The resulting structure information was then incorporated in turn into a 3 to 5-class Markov model for tissue segmentation via another specific external field. Tissue and structure segmentations thus appear as dynamical and cooperative MRF procedures whose performance increases gradually. This approach was implemented into a multi-agent framework, where autonomous entities, distributed into the image, estimate local Markov fields and cooperate to ensure consistency (see Section 3.2.4).

Although satisfying in practice, these tissue and structure MRFs did not correspond to a valid joint probabilistic model and were not compatible in that sense. As a consequence, important issues such as convergence or other theoretical properties of

the resulting local procedure could not be addressed. In addition, in [36, 37], cooperation mechanisms between local models were somewhat arbitrary and independent of the MRF models themselves. Our second contribution [4, 31] was then to propose a fully Bayesian framework in which we define a joint model that links local tissue and structure segmentations but also the model parameters so that both types of cooperations, between tissues and structures and between local models, are deduced from the joint model and optimal in that sense. This model has the following main features : 1) cooperative segmentation of both tissues and structures is encoded via a joint probabilistic model specified through conditional MRF models which capture the relations between tissues and structures. This model specifications also integrate external a priori knowledge in a natural way ; 2) intensity nonuniformity is handled by using a specific parametrization of tissue intensity distributions which induces local estimations on subvolumes of the entire volume ; 3) global consistency between local estimations is automatically ensured by using a MRF spatial prior for the intensity distributions parameters. Estimation within our framework was defined as a maximum a posteriori (MAP) estimation problem and carried out by adopting again an instance of the EM algorithm. We showed that such a setting could adapt well to our conditional models formulation and simplified into alternating and cooperative estimation procedures for standard hidden MRF models.

Then, we proposed to go further towards coupling methods by constructing a model that performed local affine registration of an atlas in addition to tissue segmentation using local Markov random fields and MRF segmentation of structures. The idea was to capture in a single model all the relationships that could be formalized between these tasks. Our basis toward this third solution was similar to that in [31] with the major difference that therein a joint model was not explicitly given but defined through the specification of a number of compatible conditional MRF models. In [28], we specified directly a joint model from which the conditional models were derived. As a result, cooperation between tissues and structures was treated in a more symmetric way which resulted in new even more consistent conditional models. In addition, interaction between the segmentation and registration steps was easily introduced. An explicit joint formulation has the advantage to provide a strategy to construct more consistent or complete models that are open to incorporation of new tasks. For estimation, we provided again an appropriate variational EM framework allowing a Bayesian treatment of the parameters. The evaluation performed on both phantoms and real 3T brain scans showed good results and demonstrated the clear improvement provided by coupling the registration step to tissue and structure segmentation (see Figure 18).

4.1.4 Bayesian Weighting of Multiple MR Sequences for Brain Lesion Segmentation

Related publications : 3 international conferences [38, 27, 26].

This is joint work with Senan Doyle, post-doctoral fellow in MISTIS*, Michel Dojat (Grenoble Institute of Neuroscience), Daniel Garcia-Lorenzo and Christian Barillot (INRIA*

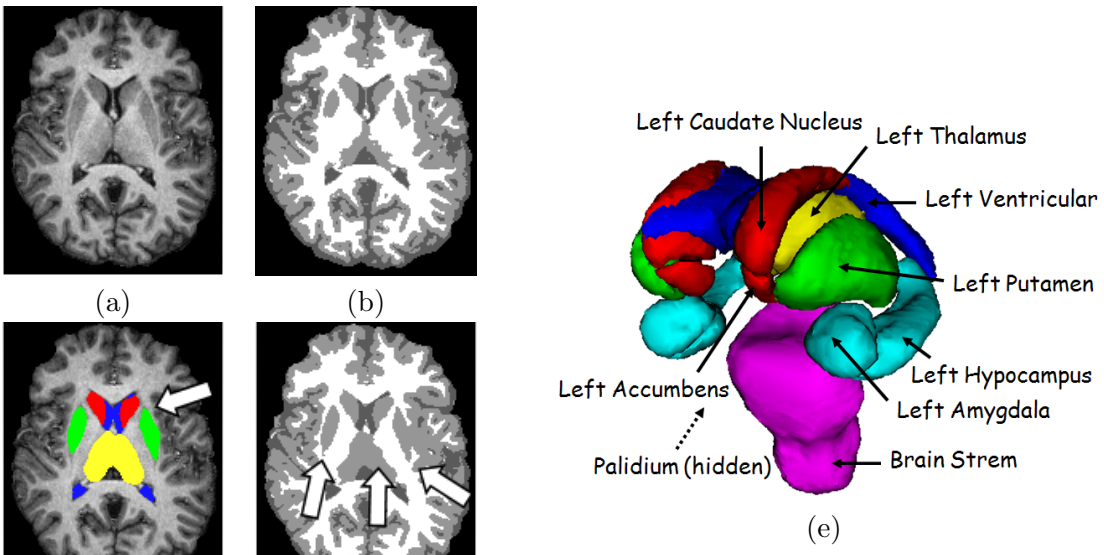


FIGURE 18 – Evaluation of the cooperative approach on a real 3T brain scan (a). For comparison the segmentation of tissues only is given in (b). The results obtained with our approach are shown in the second line. Major differences between tissue segmentations (images (b) and (d)) are pointed out using arrows. Image (e) shows the corresponding 3D reconstruction of 17 structures segmented using our approache. The names of the left structures (use symmetry for the right structures) are indicated in the image.

Team VISAGES*).*

The investigation described in the previous section has been carried out for healthy brain for one type (T1) of MR images with no temporal information. In an ARC project SeLMIC 2007-08 (see Section 6 in my detailed CV), we planned to extend our tools to include multidimensional MR sequences corresponding to other types of MR modalities and longitudinal data. This was investigated in collaboration with team VISAGES from INRIA Rennes. Such extensions are particularly relevant in the study of some central nervous system disorders such as Multiple Sclerosis (MS) and brain attacks [38]. After the end of the SeLMIC project, we went on investigating the delineation of brain lesions from multiple MR images. The delineation and quantification of brain lesions is critical to establishing patient prognosis, and for charting the development of pathology over time. Typically, this is performed manually by a medical expert, although automatic methods have been proposed (see the review od Seghier et al.[47]) to alleviate the tedious, time consuming and subjective nature of manual delineation. A healthy brain is generally segmented into three tissues : cephalo spinal fluid, grey matter and white matter, and one or two extra classes are generally added to account for lesions. Statistical based approaches usually aim to model probability distributions of voxel intensities with the idea that such distributions are tissue-dependent. Automated or semi-automated brain lesion detection methods can then be classified according to their use of multiple sequences, *a priori* knowledge about the structure of normal brain, tissue segmentation models, and whether or not specific lesion types are targeted. A common feature is that most methods are based on the initial identification of *candidate regions* for lesions. In most approaches, normal brain tissue *a priori* maps are used to help identify regions where the damaged brain differs, and the lesion is identified as an outlier. Existing methods frequently avail of complementary information from multiple sequences. For example, lesion voxels may appear atypical in one modality and normal in another. This is well known and implicitly used by neuroradiologists when examining data. Within a mathematical framework, multiple sequences enable the superior estimation of tissue classes in a higher dimensional space.

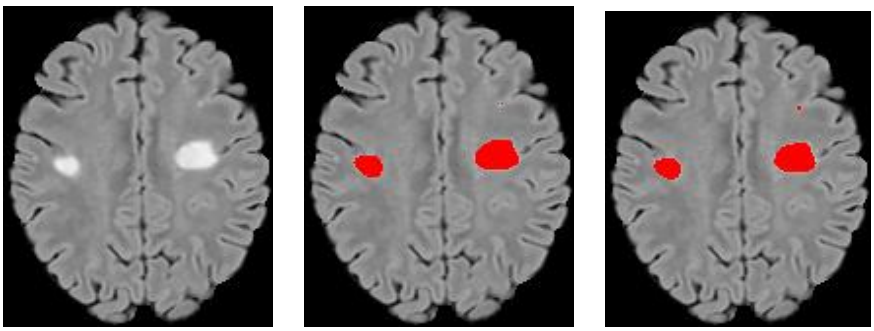
For multiple MRI volumes, intensity distributions are commonly modelled as multi-dimensional, often Gaussian, distributions. This provides a way to combine the multiple sequences in a single segmentation task but with all the sequences having equal importance. However, given that the information content and discriminative power to detect lesions varies between different MR sequences, the question remains as to how to best combine the multiple channels. Depending on the task at hand, it might be beneficial to weight the various sequences differently.

In our work [27, 26], rather than trying to detect lesion voxels as outliers from a normal tissue model, we adopt an incorporation strategy whose goal is to identify lesion voxels as a fourth component additional to the usual three tissues. Such an explicit modeling of the lesions is usually avoided. It is difficult for at least two reasons : 1)

[47] M.L. Seghier, A. Ramlackhansingh, J. Crinion, A.P. Leff, and C. J. Price. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage*, 41 :1253–1266, 2008.

most lesions have a widely varying and inhomogeneous appearance (*e.g.* tumors or stroke lesions) and 2) lesion sizes can be small (*e.g.* multiple sclerosis lesions). In a standard tissue segmentation approach, both reasons usually prevent accurate model parameter estimation resulting in bad lesion delineation. Our approach aims to make this estimation possible by modifying the segmentation model with an additional weight field. We propose to modify the tissue segmentation model so that lesion voxels become inliers for the modified model and can be identified as a genuine model component. Compared to *robust estimation* approaches (*e.g.* the work by Van Leemput et al.[48]) that consist of down-weighting the effect of outliers on the main model estimation, we aim to increase the weight of candidate lesion voxels to overcome the problem of under-representation of the lesion class.

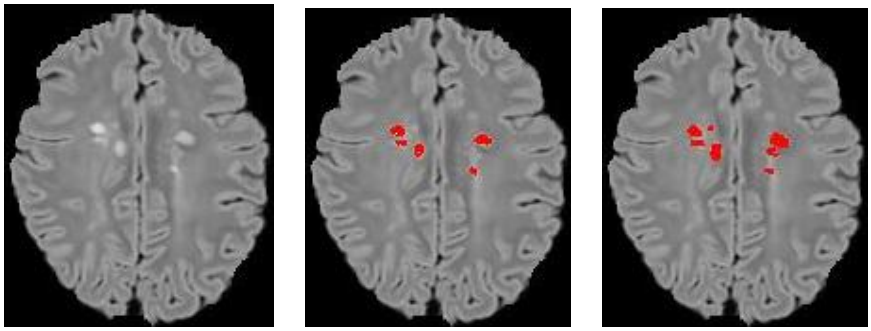
Using the model presented in Section 3.1.6, we introduce weight parameters in the segmentation model and then solve the issue of prescribing values for these weights. This has the advantage to avoid the specification of *ad-hoc* weight values and to allow the incorporation of expert knowledge through a weight prior distribution. Furthermore, in the absence of explicit expert knowledge, we show how the weight prior can be specified to guide the model toward lesion identification. Experiments on artificial (Table 1) and real lesions (Table 2, Figures 19, 20, 21) of various sizes are reported to demonstrate the good performance of our approach. We performed a quantitative evaluation using the Dice similarity coefficient (DSC). This coefficient measures the overlap between a segmentation result and the gold standard. Denoting by TP_k the number of true positives for class k, FP_k the number of false positives and FN_k the number of false negatives the DSC is given by : $d_k = \frac{2\mathrm{TP}_k}{2\mathrm{TP}_k+\mathrm{FN}_k+\mathrm{FP}_k}$ and d_k takes its value in $[0, 1]$ where 1 represents the perfect agreement.



(a) (b) (c)

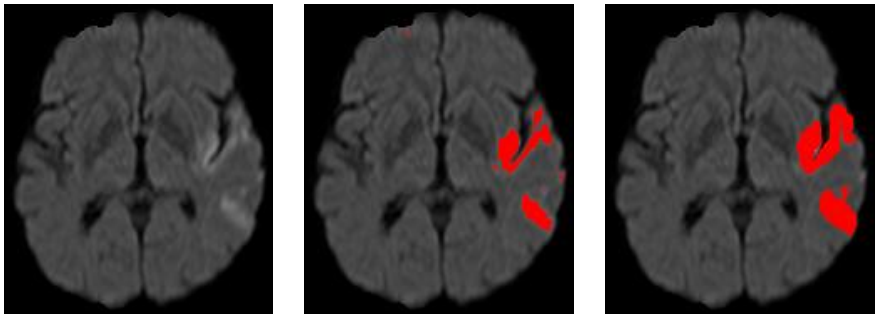
FIGURE 19 – Real MS data, patient 1. (a) : Flair image. (b) : identified lesions with our approach (DSC 82%). (c) : ground truth .

[48] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Medical Imaging*, 20(8) :677–688, 2001.



(a) (b) (c)

FIGURE 20 – Real MS data, patient 3. (a) : Flair image. (b) : identified lesions with our approach (DSC 45%). (c) : ground truth .



(a) (b) (c)

FIGURE 21 – Real stroke data. (a) : DW image. (b) : identified lesions with our approach (DSC 63%). (c) : ground truth.

Method	3%	5%	7%	9%
Mild lesions (0.02% of the voxels)				
AWEM	**72 (+5)**	55 (-15)	**39 (+5)**	**22 (+18)**
[G]	67	**70**	34	0
EMS	56	33	13	4
[R]	52	NA	NA	NA
Moderate lesions (0.18% of the voxels)				
AWEM	**86 (+7)**	80 (-1)	**77 (+18)**	**73 (+36)**
[G]	72	**81**	59	29
EMS	79	69	52	37
[R]	63	NA	NA	NA
Severe lesions (0.52% of the voxels)				
AWEM	**93 (+8)**	88 (0)	**78 (+6)**	**74 (+33)**
[G]	79	**88**	72	41
EMS	85	72	56	41
[R]	82	NA	NA	NA

0% non-uniformities

-

Method	3%	5%	7%	9%
Mild lesions (0.02% of the voxels)				
AWEM	50 (-25)	0 (-65)	0 (-20)	0 (-30)
[G]	**75**	**65**	**20**	**30**
EMS	58	27	13	6
Moderate lesions (0.18% of the voxels)				
AWEM	64 (-12)	66 (-10)	66 (-1)	0 (-48)
[G]	75	**76**	**67**	**48**
EMS	**76**	64	47	31
Severe lesions (0.52% of the voxels)				
AWEM	**88 (+2)**	**84 (+1)**	**80 (+6)**	**67 (+9)**
[G]	75	83	74	58
EMS	86	74	62	45

40% non-uniformities

TABLE 1 – DSC results (%) on MS Brain Web simulated data, for various lesion sizes, noise and non-uniformity levels. Comparison of different methods : AWEM for our Adaptive Weighted EM, Garcia-Lorenzo & al's method [G], Van Leemput & al.'s method [EMS] and Rousseau & al.'s method [R]. The corresponding gain/loss over the best comparable results is given in parentheses. NA means not available.

4.1.5 Variational approach for the joint estimation-detection of Brain activity from functional MRI data

Related publications : 1 journal paper [2], 1 international conference [24].

This is joint work with Michel Dojat from Grenoble Institute of Neuroscience and Philippe Ciuciu from Laboratoire de Neuroimagerie Assistée par Ordinateur(LNAO), NeurosSpin, CEA Saclay.

	LL	EMS	AWEM
Patient1	0.42	62	**82 (+20)**
Patient2	1.71	54	**56 (+2)**
Patient3	0.29	**47**	45 (-2)
Patient4	1.59	65	**72 (+7)**
Patient5	0.31	**47**	45 (-2)
Average		55 +/-8	60 +/-16

TABLE 2 – Lesion load or percentage of lesion voxels (LL), DSC results (%) for Van Leemput & al.'s method [48] (EMS) and for our Adaptive Weighted EM (AWEM), for 5 patients with MS.

More recently, we have started investigating the extension of our methodology for anatomical MRI to functional MRI (fMRI) data. The latter are from rather different nature as they incorporate an additional time course. Since the first report of the Blood Oxygen Level Dependent (BOLD) effect in humans, fMRI represents a powerful tool to non-invasively study the relation between sensori-motor or cognitive tasks and the hemodynamic BOLD response reflecting evoked neuronal activity indirectly. fMRI allows us to determine which cortical regions elicit activations when a human subject is submitted to specific stimuli. One of the major goals of fMRI is to show the existence of common cognitive functions based on the exploration of a population of subjects (typically about 15) submitted to the same experimental paradigm. This has to be based on a careful within-subject statistical analysis. This analysis is traditionally conducted using a Generalized Linear Model (GLM) built upon an experimental design matrix that specifies the onsets of each stimulus occurrence. Such a GLM also postulates a preset form of the so-called Hemodynamic Response Function (HRF). However, it appears that fixing the latter function to some canonical shape is too restrictive and misleading. In fact, as demonstrated by various experiments, the HRF may vary between subjects and even between cortical regions in a given subject's brain.

The context of our work is then the conjoint brain activity detection and HRF estimation on the whole brain. To get reliable estimates, HRF estimation has to be performed at a coarser spatial scale than the voxel level given the poor signal to noise ratio of fMRI data. To this end, Makni et al. in 2005[49] and in 2008[50] have introduced a *joint estimation detection* (JED) framework that implements using a MCMC framework such a joint procedure on a prior parcellation of the function brain mask.

We have started investigating the possibility of using variational approximation

[49] S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline. Joint detection-estimation of brain activity in functional MRI : a multichannel deconvolution solution. *IEEE Trans. Signal Processing*, 53(9) :3488–3502, Sep. 2005.

[50] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3) :941–969, July 2008.

techniques as an alternative to the MCMC based method[50]. The 5-month internship of Alexandre Janon enabled us to initiate this activity [2, 24] which is now pursued with a new collaboration with Philippe Ciuciu from LNAO, NeuroSpin, CEA in Saclay.

4.1.6 Model-based Region-of-Interest Selection in dynamic breast MRI

Related publications : 2 journal papers [12, 17], 3 research reports [65, 64, 62].

This is joint work with Chris Fraley and Adrian Raftery from University of Washington in Seattle, with Dave Goldhaber and Dianne Georgian-Smith, M.D. from Harvard Medical School, Massachussets General Hospital.

Magnetic resonance imaging has also emerged as a powerful tool for the diagnosis of breast abnormalities. Dynamic analysis of the temporal pattern of contrast uptake has been applied in differential diagnosis of benign and malignant lesions to improve specificity. Signal intensity time course data are useful for differentiating benign from malignant enhancing lesions. The overall shape of the time-signal intensity curve is an important criterion, while a single attribute of the curve, such as the enhancement rate, may not be enough.

Selecting a region of interest (ROI) is an almost universal step in the process of examining the contrast uptake characteristics of a breast lesion. I proposed [64, 62, 12] an ROI selection method that combines model-based clustering of the pixels with Bayesian morphology [17]. I then investigated tools for subsequent analysis of signal intensity time course data in the selected region.

Results on a data base of 19 patients were promising. The method provided informative segmentations and good detection rates were obtained. The investigation indicated that our proposed statistical methods, which enable us to take into account more than a single enhancement measure, were quite promising for tumor identification. There is a clear gain in combining segmentation with kinetics analysis. Associating the location and shape of a lesion with its pattern of uptake proved to be useful in resolving questionable cases. The trade-off between smoothness and resolution needs to be assessed by further empirical research on other images. Our study was limited to the determination of feasibility for the proposed computational methods. Clinical value would have to be assessed in more extensive and controlled studies, which in the light of our initial experience may be warranted.

4.2 Biology and Computational Neuroscience

A second domain of applications concerns biomedical statistics and molecular biology. I have considered the use of missing data models in epidemiology and investigated statistical tools for the analysis of bacterial genomes beyond gene detection (Section

[50] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3) :941–969, July 2008.

4.2.1). I also considered applications in population genetics (Section 4.2.2) and computational neuroscience (POP and HUMAVIPS European projects, Section 4.2.3).

4.2.1 Integrated Markov models on irregular graphs for clustering gene expression data

Related publications : 1 journal paper [5], 2 international conference [42, 43], 4 national conferences [47, 50, 51, 46].

This work was carried out with Matthieu Vignes during his PhD thesis 2003-07. For the additional treatment of missing data, the work also involved Juliette Blanchet.

Spatial gene clustering. Because of the increasing amount of genetic data generated by researchers, there is a great need to develop methodologies to analyse and to use the information contained in this data. A major challenge in bioinformatics is to reveal interactions between components of living organisms and discover the corresponding networks responsible for their biological complexity. In this framework, clustering of genes into groups sharing common characteristics is a useful exploratory technique. It is frequently used as the basis for further computational analysis. As a simple example, the function of a gene can be predicted according to known functions of other genes from the same cluster.

A wide range of clustering algorithms have been proposed to analyze gene expression data but most of them consider the genes as independent entities or include relevant information on gene interactions a posteriori. We proposed [5] a probabilistic model that has the advantage to take into account individual features (*e.g.* expression) and pairwise data (*e.g.* interaction information coming from biological networks) simultaneously. As mentioned in Section 3.1.4, our model is based on hidden Markov random fields in which parametric probability distributions account for the distribution of individual data for each gene. Data on pairs are included through a graph where the nodes represent the genes and the edges are weighted according to pair data, for instance in order to reflect distances or similarity measures between genes. This model has many interesting features. It leads to various possible statistical criteria to select automatically the number of clusters. It is also able to incorporate many types of data. It is flexible in the sense that its generalization to include missing data, that often occur when dealing with expression data, is straightforward. Its extension to overlapping clustering methods, to deal with more realistic situations where genes can belong to many groups at the same time, can also be considered. Preliminary investigations are reported in [42, 43, 50, 51]. We have illustrated and validated the approach on simulated data as well as on yeast expression data combined with pathway neighborhoods [5].

Expression data with missing values. DNA microarray technologies provide means for monitoring in the order of tens of thousands of gene expression levels quantitatively and simultaneously. However data generated in these experiments can be noisy and have missing values. When it is not ignored, the last issue has been solved by imputing the expression matrix in order to keep going with traditional analysis methods.

Although it was a first useful step, it is not recommended to use value imputation to deal with missing data. Moreover, as previously mentioned, appropriate tools are needed to cope with noisy background in expression levels and to take into account a dependency structure among the genes under study. Various approaches have been proposed but to our knowledge none of them has the ability to fulfil all these features. In [5], we mentioned the ability of a straightforward extension of the model therein to deal with missing values. I therefore initiated [47] the study of clustering algorithms that explicitly account for dependencies within a biological network and for missing value mechanism to analyze microarray data. I proposed to tackle these issues in a unique statistical framework. Theoretical developments have been specified in Juliette Blanchet's PhD thesis. The approach has been implemented and proved to be successful at dealing with different absence patterns either on simulated or real biological data sets. It has demonstrated enhanced results over existing approaches. Juliette Blanchet and Matthieu Vignes emphasized[51,52] that our model could be useful in a great range of applications for clustering entities of interest (such as genes, proteins, metabolites in post-genomics studies). The method is neither organism- nor data-specific. More generally, the method is of interest in a wide variety of fields where missing data is a common feature : social sciences, computer vision, remote sensing, speech recognition and of course biological systems. The analysis workflow is illustrated in Figure 22.

4.2.2 Modeling and inference of population structure from genetic and spatial data

Related publications : 2 journal papers [11, 9].

This is joint work with Chibiao Chen (INRIA post-doctoral fellow in 2006) and Prof. Olivier François from team TimB in TIMC laboratory in Grenoble.

In applications of population genetics, it is often useful to classify individuals in a sample into populations which become then the units of interest. However, the definition of populations is typically subjective, based, for example, on linguistic, cultural, or physical characters as well as the geographic location of sampled individuals. Pritchard et al[53] proposed a Bayesian approach to classify individuals into groups using genotype data. Such data, also called multilocus genotype data, consists of several genetic markers whose variations are measured at a series of loci for each sampled individual. Their method is based on a parametric model in which there are K groups (where K may be unknown), each of which is characterized by a set of allele frequencies at each

[51] J. Blanchet and M. Vignes. Combined expression data with missing values and gene interaction network analysis : a Markovian integrated approach. In *7th IEEE BIBE Conference*, pages 366–373, Boston, USA, 2007.

[52] J. Blanchet and M. Vignes. A model-based approach to gene clustering with missing observations reconstruction in a Markov random field framework. *Journal of Computational Biology*, 16(3) :475–486, 2009.

[53] J.K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155 :945–959, 2000.

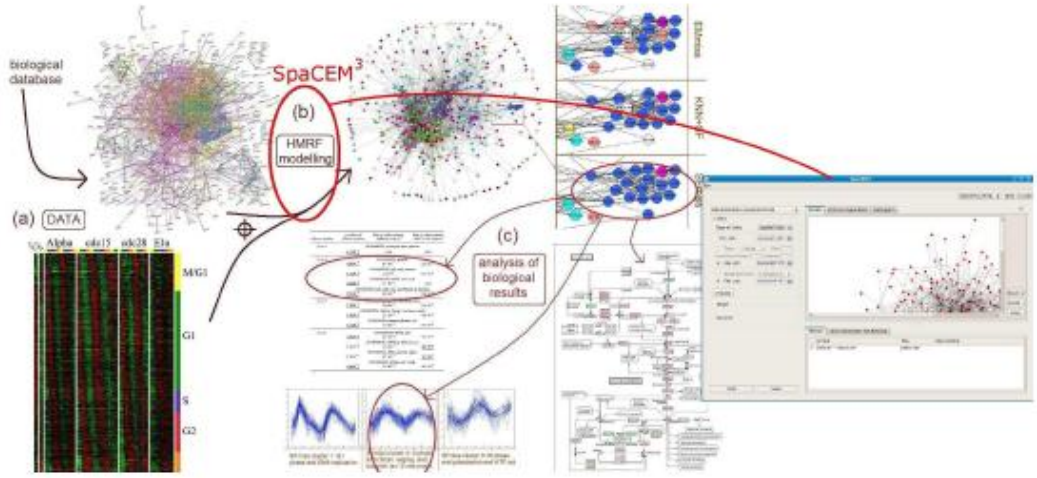


FIGURE 22 – Graphical summary of an analysis workflow : (a) data extraction from relevant databases, (b) specification of the HMRF setting using our Spacem3 software, model inference and result visualization, (c) downstream biological analysis for biological cluster relevance : modularity, over-represented GO-terms, expression level profiles and link to pathways.

locus. Group allele frequencies are unknown and modelled by a Dirichlet distribution at each locus within each group. A MCMC algorithm is then used to estimate simultaneously assignment probabilities and allele frequencies for all groups. In such a model, individuals are assumed to be independent, which does not take into account their possible spatial proximity.

The main goal of our work [9] was to introduce spatial prior models and to assess their role in accounting for the relationships between individuals. In this perspective, we proposed to investigate particular Markov models on graphs and to evaluate the quality of mean field approximations for the estimation of their parameters. Maximum likelihood estimation of such models in a spatial context is typically intractable but mean field like approximations within an EM algorithm framework, in the spirit of [14] can be considered and result in an alternative procedure to MCMC approaches. With this in mind, we first considered the EM approach in a non spatial case, as an alternative to the traditional Bayesian approaches. This resulted in a new computer program (see Section *Software* in the next part) and promising results were reported in [11].

This first approach was based on traditional hidden Markov models for which a standard conditional independence assumption holds. Dependencies between individuals are described through spatial correlations between groups meaning that spatially close individuals are likely to belong to the same group. In the joint work with INRA Avignon [66, 54] (Section 3.1.8), we tried to weaken the standard but somewhat unrea-

listic conditional independence assumption to describe dependencies at the observations level through a spatial correlation model inspired from geostatistical models . Another goal of this work would then be to investigate such extension in the context of population genetics.

4.2.3 Multi-speaker Localization with Binaural Audition and Stereo Vision

Related publications : 1 journal paper [1], 3 international conferences [32, 33, 34].

This is part of Vasil Khalidov's PhD thesis and is joint work with Elise Arnaud, Miles Hansard, Radu Horaud and Ramya Narasimha from team PERCEPTION. *This work took place in the context of the POP European project (see Section 6 of my detailed CV part) and included further collaborations with researchers from University of Sheffield, UK.*

The context is that of multi-modal sensory signal integration as already mentioned in Section 3.1.2. We have focused on audio-visual integration. Fusing information from audio and video sources can result in improved performance in applications such as tracking. However, crossmodal integration is not trivial and requires some cognitive modeling because at a lower level, there is no obvious way to associate depth and sound sources. Combining expertise from team PERCEPTION and University of Sheffield, we addressed the difficult problem of integrating spatial and temporal audio-visual stimuli using a geometrical and probabilistic framework. We developed the conjugate model of Section 3.1.2 for this purpose.

Audio-visual objects can be characterized both by their locations in space and by their auditory status, *i.e.*, whether they are emitting sounds or not. These object characteristics are not directly observable and hence they need to be inferred from sensor data, *e.g.*, cameras and microphones. These sensors are based on different physical principles, they operate with different bandwidths and sampling rates, and they provide different types of information. On one side, light waves convey useful visual information only indirectly, on the premise that they reflect onto the object surfaces. A natural scene is composed of many objects/surfaces and hence the task of associating visual data with objects is a difficult one. On the other side, acoustic waves convey auditory information directly from the emitter to the receiver but the observed data is perturbed by the presence of reverberations, of other sound sources, and of background noise. Moreover, very different methods are used to extract information from these two sensor types. A wide variety of computer vision principles exist for extracting 3D points from a single image or from a pair of stereoscopic cameras but practical methods are strongly dependent on the lighting conditions and on the properties of the object surfaces (presence or absence of texture, color, shape, reflectance, etc.). Similarly, various algorithms have been developed to locate sound sources using a microphone pair based on interaural time differences (ITD) and on interaural level differences (ILD)[54,55], but these cues

[54] D. Wang and G. J. Brown, editors. *Computational Auditory Scene Analysis : Principles, Algorithms, and Applications.* Wiley-IEEE Press, September 2006.

[55] H. Christensen, N. Ma, S.N. Wrigley, and J. Barker. Integrating pitch and localisation cues at a

are difficult to interpret in natural settings due to the presence of background noise and of other reverberant objects. A notable improvement consists in the use of a larger number of microphones. Nevertheless, the extraction of 3D sound source positions from several microphone observations results in inaccurate estimates. We showed then in [1] that our conjugate mixture models framework may help in the task of locating several speaking persons. The multimodal data consists of M visual observations $\mathbf{f}$ and of N auditory observations $\mathbf{g}$. We considered data that were recorded over a short time interval such that one can reasonably assume that the audio-visual (AV) objects have a stationary spatial location. Nevertheless, it is not assumed here that the AV objects, *e.g.*, speakers, are static : lip movements, head and hand gestures are tolerated. We addressed the problem of estimating the spatial locations of all the objects that are both seen and heard. Let K be the number of objects and in this case each object is described by a three dimensional parameter vector $o_k = (x_k; y_k; z_k)$. The AV data are gathered using a pair of stereoscopic cameras and a pair of omnidirectional microphones, *i.e.*, binocular vision and binaural hearing. For $j = 1$ to M, a visual observation vector $f_j = (u_j; v_j; d_j)$ corresponds to a 2D image location $(u_j; v_j)$ and to an associated binocular disparity d_j. Considering a projective camera model it is straightforward to define an invertible function $\mathcal{F} : \mathbb{R}^3 \to \mathbb{R}^3$ that maps $o = (x; y; z)$ onto $f = (u; v; d)$:

$$\mathcal{F}(o) = {}^t(\frac{x}{z}; \frac{y}{z}; \frac{1}{z}) \quad \text{and} \quad \mathcal{F}^{-1}(f) = {}^t(\frac{u}{d}; \frac{v}{d}; \frac{1}{d}) \; .$$

This model corresponds to a rectified camera pair and it can be easily generalized to more complex binocular geometries[56,57].

Similarly one can use the auditory equivalent of disparity, namely the interaural time difference (ITD) widely used by auditory scene analysis methods. The function $\mathcal{G} : \mathbb{R}^3 \to \mathbb{R}^3$ maps $o = (x; y; z)$ onto a 1D audio observation :

$$g = \mathcal{G}(o) = \frac{1}{c}(||o - o_{M1}|| - ||o - o_{M2}||).$$

Here c is the sound speed and o_{M1} and o_{M2} are the 3D locations of the two microphones in the sensor-centered coordinate system. Each isosurface defined by the above equation is represented by one sheet of a two-sheet hyperboloid in 3D. Hence, each audio observation g constrains the location of the auditory source to lie onto a 2D manifold (see Figure 23).

With this setting, we then performed audiovisual clustering based on the conjugate EM algorithm mentioned in Section 3.2.4 (See Figures 24 and 25). We refer to [1] for more illustrations. Simulated data experiments allowed us to assess the average method behaviour in various configurations. They showed that the obtained clustering results

speech fragment level. In *Proc. of Interspeech*, pages 2769–2772, 2007.

[56] M. Hansard and R. Horaud. Patterns of binocular disparity for a fixating observer. In *Proc. of Second International Symposium of Advances in Brain, Vision, and Artificial Intelligence*, pages 308–317. Springer, 2007.

[57] M. Hansard and R. Horaud. Cyclopean geometry of binocular vision. *Journal of the Optical Society of America A*, 25(9) :2357 2369, September 2008.

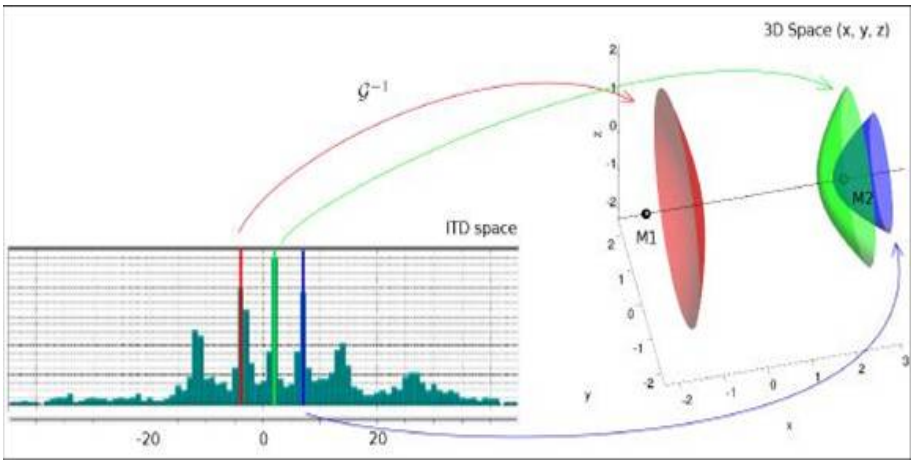


FIGURE 23 – Audio mapping from the 3D object space to the 1D audio space (ITD).

were precise as regards the observation spaces under consideration. They also illustrated the theoretical dependency between the precisions in observation and parameter spaces. Real data experiments then showed that the observed data precision was high enough to guarantee high precision in the parameter space.

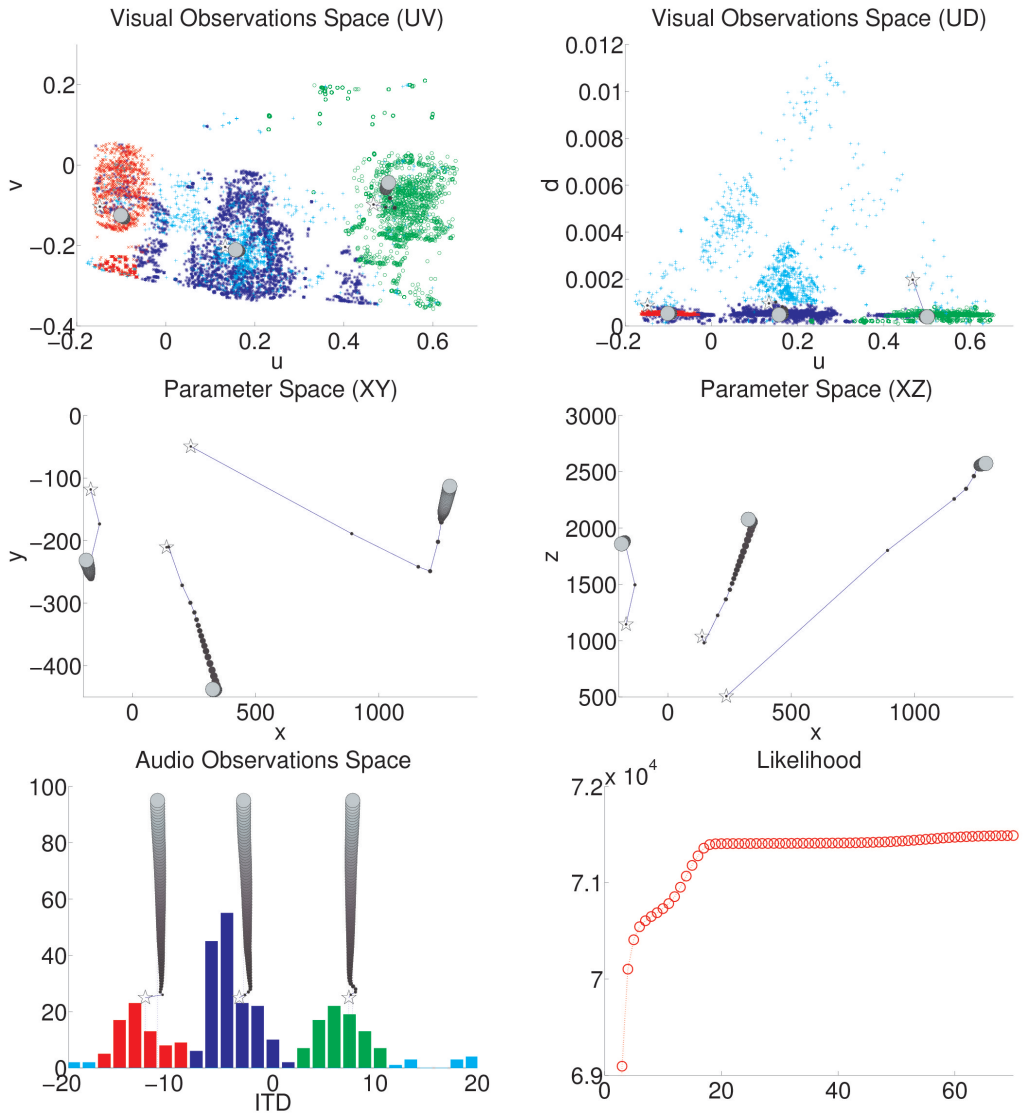
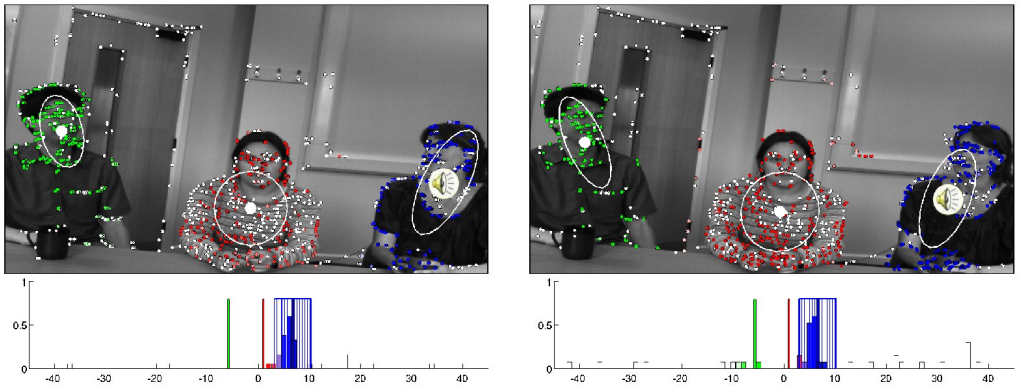
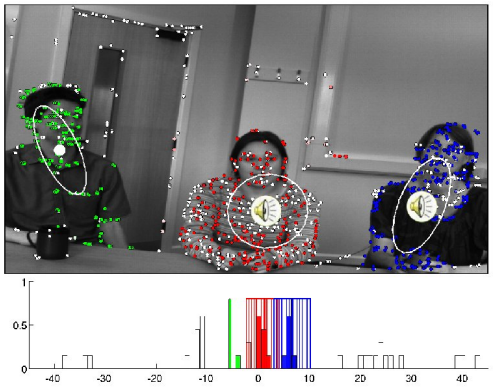


FIGURE 24 – An example of applying the proposed EM algorithm to a time interval of 20 seconds of the meeting scenario. The results are shown in the visual and auditory observation spaces as well as in the parameter space. The initial parameter values are shown with three stars while the parameter evolution trajectories are shown with circles of increasing size. The final observation-to-cluster assignments are shown in color : red, blue, and green for the three Gaussian components and light-blue for the outlier component. The log-likelihood curve (bottom-right) shows that the algorithm converged after 20 iterations.

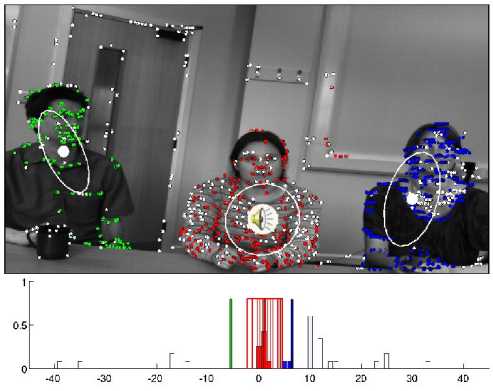


(a) frames 1001-1010

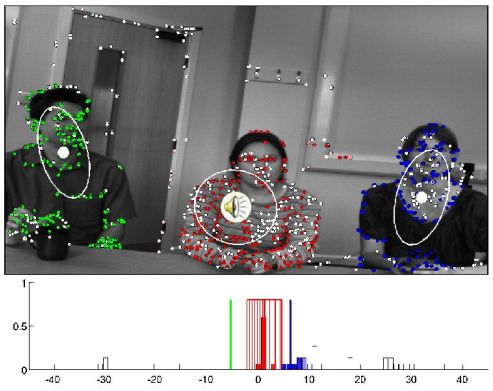
(b) frames 1011-1020



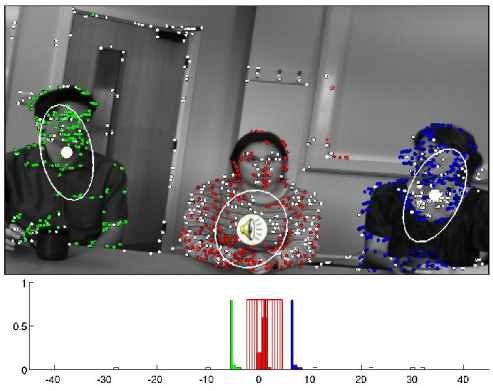
(c) frames 1021-1030



(d) frames 1031-1040



(e) frames 1041-1050



(f) frames 1051-1060

FIGURE 25 – Results obtained in the case of the meeting scenario shown overlapped onto the left image. Sixty frames (1001 to 1060) were split into six segments. Parameter initialization and model selection were performed on the first segment (frames 1-10) and are not shown. The "visual" covariance matrices associated with the 3 Gaussian components are projected onto the image plane. The white dots correspond to the projected 3D locations estimated by the algorithm. The blue, green, and red colors encode the observation-to-cluster assignments and the active speaker is marked with a corresponding symbol. The algorithm correctly estimates speech sources, even in the case when two speakers are active.

Technology development

1 Software

More than a contribution to the development of a specific software, what I would like to emphasize in this section are my efforts to develop a unifying approach of software development in MISTIS. I initiated this approach with the advising of Lemine Abdalah in 2005 who was charged at that time with transforming a first prototype into a flexible piece of software with a clear structure that would become our SpaCEM^3 software. I then successfully applied in 2007 for an ODL position and hired then Sophie Chopart to help us. The goal was not only to reach a higher level in software development but also to generate new habits among researchers in MISTIS. This goal has been reasonably reached today with most of our recent software and prototypes on the INRIA forge.

I have been myself involved in the following software :

The SpaCEM^3 software. The SpaCEM^3 (Spatial Clustering with EM and Markov Models) software replaces the former, still available, SEMMS (Spatial EM for Markovian Segmentation). SpaCEM^3 proposes a variety of algorithms for image segmentation, supervised and unsupervised classification of multidimensional and spatially located data. The software is written in C++ (35000 lines) and is available with a new user interface at `http://spacem3.gforge.inria.fr`. A user guide and example data sets are also available there. The software has been registered at APP (Software Protection Agency) in 2007 under the CeCILLB license. A description of its main functionalities are also available in French [23]. See Figure 26 for an illustration of the interface.

The FASTRUCT software. Bayesian model-based clustering programs have gained increased popularity in studies of population structure since the publication of the software STRUCTURE. These programs are generally acknowledged as performing well, but their running-time may be prohibitive. FASTRUCT is a non-Bayesian implementation of the classical model with no-admixture uncorrelated allele frequencies. This program relies on the Expectation-Maximization principle, and produces assignment rivaling other model-based clustering programs. In addition, it can be several-fold faster than Bayesian implementations [11]. The software consists of a command-line engine, which is suitable for batch-analysis of data, and a MS Windows graphical interface, which is convenient for exploring data. It has been registered at APP in 2006, under the CeCILLB license.

The TESS software. TESS is a computer program in C++ that implements a Bayesian clustering algorithm for spatial population genetics [9]. Is it particularly useful for seeking genetic barriers or genetic discontinuities in continuous populations. The method is based on a hierarchical mixture model where the prior distribution on cluster labels is defined as a Hidden Markov Random Field. Given individual geographical locations, the program seeks population structure from multilocus genotypes without assuming predefined populations. TESS takes input data files in a format compatible to existing non-spatial Bayesian algorithms (*e.g.*, STRUCTURE). It returns graphical displays of cluster membership probabilities and geographical cluster assignments from its Graphical User Interface. It has

been registered at APP in 2006 under the CeCILLB license. New models have been implemented until recently by O. Francois and collaborators from TimB, `http://membres-timc.imag.fr/Olivier.Francois/tess.html`.

The ECMPR software. The ECMPR (Expectation Conditional Maximization for Point Registration) package implements [39, 3]. It registers two (2D or 3D) point clouds using an algorithm based on maximum likelihood with hidden variables. The method can register both rigid and articulated shapes. It estimates both the rigid or the kinematic transformation between the two shapes as well as the parameters (covariances) associated with the underlying Gaussian mixture model. It has been registered at APP in 2010 under the GPL license.

The LOCUS software. The LOCUS software analyses in a few minutes a 3D MR brain scan and identifies brain tissues and a large number of brain structures. An image is divided into cubes on each of which a statistical model is applied. This provides a number of local treatments that are then integrated to ensure consistency at a global level. It results a low sensitivity to artefacts. The statistical model is based on a Markovian approach which enables to capture the relations between tissues and structures, to integrate a priori anatomical knowledge and to handle local estimations and spatial correlations. A description and a video of the software are available at the web site `http://locus.gforge.inria.fr`, which is still under construction. The software is written in C++ (50000 lines). It has been registered at APP in 2010 under an owner license. See Figure 27 for an illustration of the interface.

The POPEYE software. POPEYE contains software modules and libraries jointly developed by three partners within the POP STREP project : INRIA, University of Sheffield, and University of Coimbra. It includes kinematic and dynamic control of the robot head, stereo calibration, camera-microphone calibration, auditory and image processing, stereo matching, binaural localization, audio-visual speaker localization. Currently this software package is not distributed outside POP.

2 Industrial transfer of research result

I have started (October 2009) coordinating, on the MISTIS side, a *Pôle de compétitivité MINALOGIC* project including VI-Technology and Pige electronique as industrial partners (see Section 6 of my detailed CV). This collaboration is an excellent opportunity to transfer my expertise in statistical image analysis and Bayesian techniques to the AOI (Automatic Optical Inspection) market. I am developing new models, for expert knowledge integration and defect detection from multiple image analysis, that have been already implemented and will be transferred to VI-Technology during the project.

We have also started for the development of the Locus software in September 2010 a project supported by GRAVIT. GRAVIT is a structure devoted to stimulating innovation that brings together 7 academic institutions, attached to 200 research labs. It is a shared tool, for the benefit of researchers and industry, to adapt the technology transfer

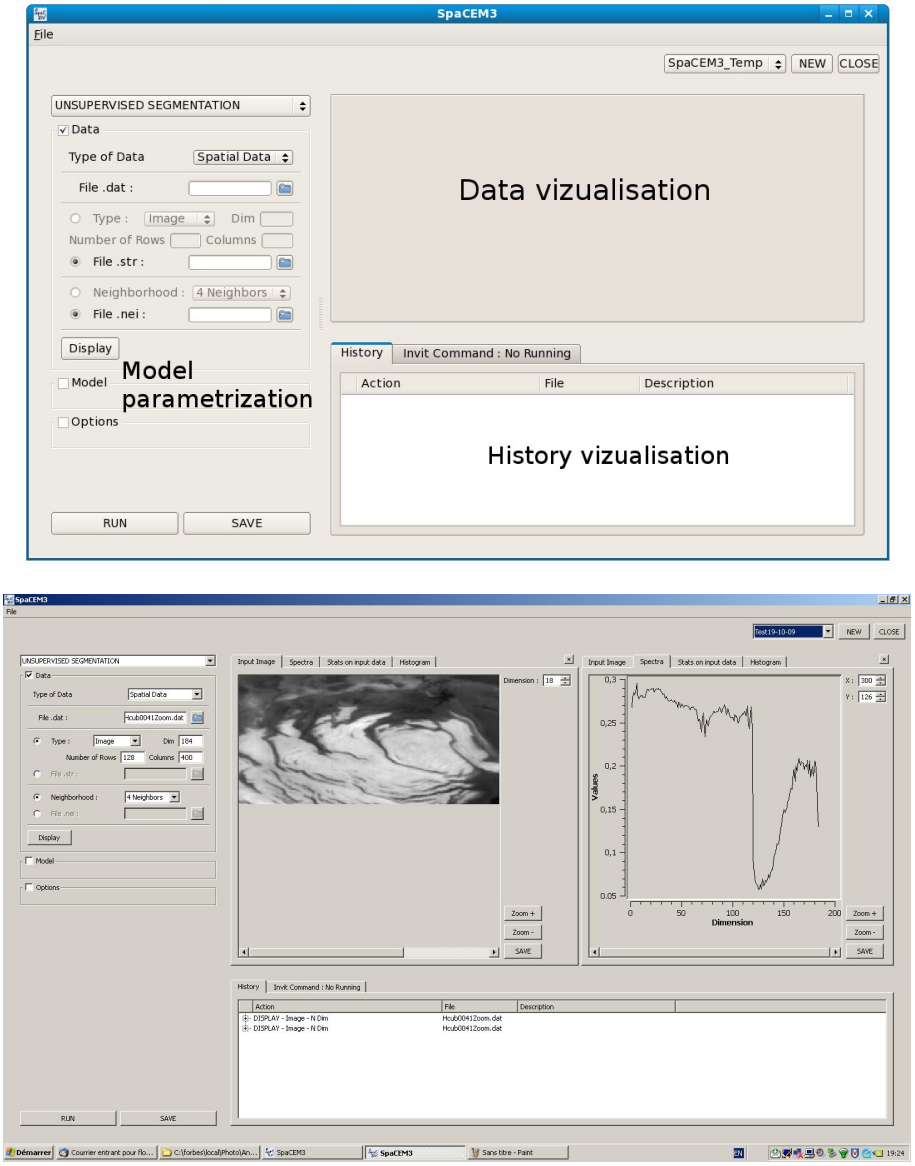


FIGURE 26 – Top : Composition of the SpaCEM^3 GUI window. Bottom : Hyperspectral image of Mars visualized in the GUI. The left image shows dimension 18 of the data set, the right image shows the spectrum at coordinates (300, 126).

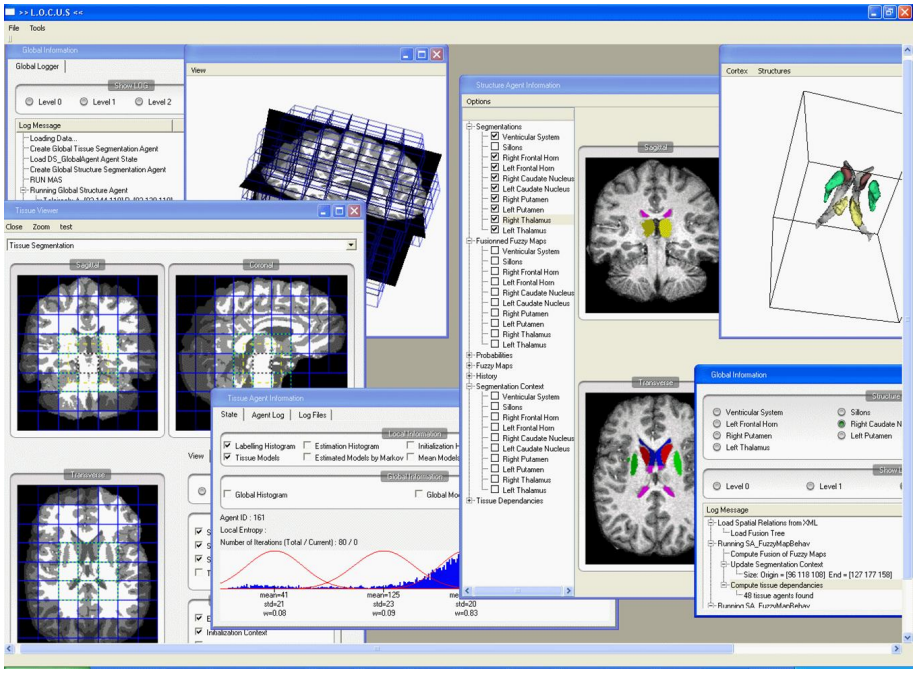


FIGURE 27 – Illustration of the Locus GUI window : data sets and segmentation results can be visualized with functionalities similar to standard MRI visualization tools.

to companies's innovation needs. In this context we are participating in the Grenoble Innovation Fair in October 2010 to explore new opportunities for collaboration among industrial groups, start-ups and research labs.

Summary of main contributions

Main contributions

To summarize, my main contributions can be divided into three axes :

1. Variational approximations for inference in hidden Markov random fields
2. Models for complex and structured data
3. Model-based clustering for the analysis of MR brain scans

1 Variational approximations for inference in hidden Markov random fields

Issues involving hidden (or missing) data are typical settings where exact inference is not tractable. Approximations are required and are often based on simulation techniques such as Monte-Carlo Markov Chains (MCMC) techniques. As a computationally more efficient alternative, I developed deterministic approximations based on the so-called mean field principle and proposed in the context of hidden Markov random fields (HMRF) a class of EM-like algorithms which shows good performance [14]. I focused on the mean field approximation principle which is among the variational methods, the simplest principle although not yet fully understood in some aspects. Convergence results exist for the related Variational Expectation Maximization (VEM) algorithm[33] but its application is restricted to very simple missing data structure. The variants I proposed in [14] extend the application domain, in particular to image segmentation and more generally to spatial data clustering. I also investigated how the variational approximation principle could be used for the selection of such models [15]. In these papers however, I did not address the proof of convergence results. As a matter of fact, in most settings of practical interest, theoretical results regarding accuracy and convergence properties are still missing. This is not the case for simulations methods as the convergence of MCMC based algorithms has been widely studied. I worked then with G. Fort to show that combining both types of methods to design new algorithms could greatly improve accuracy and modeling flexibility in missing data settings. This resulted in [10] in the so-called MCVEM algorithm which is tractable for HMRFs and for which we prove convergence results.

Variational approximations are one of my main research interests. I initiated this direction of research in 1998 with N. Peyrard's PhD thesis that I advised from 1998 to 2001. This was a natural follow-up after my post-doc in Seattle as I had started working there on statistical image analysis using HMRFs [17]. This latter work was itself using my more theoretical background in graphical models and Markov random fields which was the central subject of my thesis (1993-96) [22, 21, 20, 19, 18].

Originality and difficulty. Before the introduction of variational techniques for estimating HMRFs, most approximation techniques were based on simulation methods

[33] W. Byrne and A. Gunawardana. Convergence theorems of Generalized Alternating Minimization Procedures. *Journal of Machine Learning Research*, 1 :1–48, 2004.

at a much higher computational cost. I contributed to the development of efficient alternative methods based on the mean field principle [14] while providing also the tools for efficient model selection [15] and this was not standard at that time. The main difficulty then was related to proving theoretical properties of the resulting algorithms. Such theoretical issues are still under consideration and largely open questions today. A subsequent original direction of my research was then to consider *hybrid* approaches with the hope to combine fast and reasonable approximate estimates with simulation methods offering more tools for theoretical studies. The MCVEM algorithm [10] is an example of such a combination that allowed me to prove the first, to my knowledge, convergence result for this kind of algorithm. I believe this first effective step opened the way to a better understanding of the behavior and theoretical properties of a lot of Markov model based algorithms. In particular, analyzing how simulation steps should be incorporated so as to interact advantageously with deterministic approximations seems promising.

Validation and impact. The approximation methods I proposed provide fast and reasonable approximate parameter estimates in many scenarios and are the basis of various other works in team MISTIS. They favored our designing and developing of complex structure (Markovian) models outside our main domain of expertise. The PhD work of J. Blanchet included new *Triplet Markov models* (see contribution 2) and used these tools for inference. They were then applied successfully to a texture recognition task in the context of the ACI Movistar project. In the PhD work of M. Vignes our methodology was applied to gene expression data clustering using information from gene networks. More recently the PhD thesis of B. Scherrer on brain MRI segmentation and our work on brain lesion detection are also based on this inference tools (see contribution 3). Similarly the PhD thesis of R. Narasimha in the context of the European project POP, uses MRF modeling and variational approximation methods. Eventually, I obtained in collaboration with researchers from INRA Clermont-Ferrand-Theix a 5 year financial support from INRA for the PhD thesis and post-doc of L. Azizi. The goal is to investigate the use of HMRF for disease mapping from spatio-temporal data based on our ability to design and estimate such models.

My expertise in HMRF and variational methods is also generating interaction with O. Francois and collaborators from TIMC lab. (see Section 7 of my detailed CV). O. Francois is developing MCMC based methods for application to population genetics. Currently, we are interacting about the work of L. Azizi as it appears that more recent needs in population genetics would be potentially challenging for MCMC methods and then open to variational methods similar to what we are planning to use in L. Azizi's thesis.

I should also mentioned that in a number of our past and current projects (see Section 6 of my detailed CV), the participation of MISTIS is explicitly based on the presence of this expertise in the team. This was or still is the case for ACI Movistar, ANR Vahine, Minalogic project I-VP and ARC SeLMIC project. In addition, my former students N. Peyrard and M. Vignes are now researchers in spatial statistics at INRA Toulouse.

Dissemination. The journal publications related to the methodology are [14, 15, 10]. An extension to new types of HMRFs is given in [8].

Publications using these tools with a specific application focus are also mentioned in contributions 2 and 3. As regards texture recognition using image descriptors in the context of the Movistar ACI project, they correspond to conference papers [40, 44, 45, 52, 53]. The journal paper [5] and conference papers [42, 43, 50, 51, 47] illustrate the application to gene expression data clustering. As regards brain MRI analysis, they correspond to journal papers [7, 6, 2], book chapter [4] and conference papers [27, 26, 24, 28, 31, 36, 37, 41, 48, 49]. I also applied these tools to the consistent depth and boundary/surface normal estimation from stereo image data [25, 35, 59].

The inference tools mentioned above are also the basis of the spaCEM^3 software that we are developing (see Section 1 of the Technology development part).

2 Models for complex and structured data

I have contributed to statistical modeling by developing theoretical concepts and computational tools to handle properly complications in data-generating processes such as issues generated by dependencies between variables and by high dimensional, heterogeneous, possibly incomplete and noisy data. More specifically, I considered the following :

1. Issues related to the **the measurement process** :

High dimensional data : I proposed an HMRF framework for high dimensional data [40, 8] by extending the work of MISTIS researchers[9].

Missing observations : I addressed the issue of clustering multi-dimensional data when not all dimensions are available for all observed items in particular for the clustering of gene expression data [47]. I also worked on an EM procedure to estimate the so-called tail index from undersampled network traffic data [29].

Heterogeneous data : I proposed a *conjugate clustering* model [1] able to deal with observations that live in spaces with different physical properties such as dimensionality, space metric, sensor sampling rate, *etc.* A typical such issue, that we addressed within the European POP project, was audio-visual speaker detection [32, 33, 34]. I also proposed an elegant and efficient way to recast the complex rigid and articulated point registration problem, *e.g.* in computer vision, into a statistically well-based clustering task[39, 3].

2. Issues related to the existence of **dependencies** between observed and hidden variables :

Constraint clustering : I proposed a method to include gene network information when clustering gene expression data [5]. I also addressed the issue of taking into account

[9] C. Bouveyron, S. Girard, and C. Schmid. High dimensional data clustering. *Computational Statistics and Data Analysis*, 52 :502–519, 2007.

spatial information in the estimation of population structure from genetics data. I first proposed a non-Bayesian implementation [11] of the classical model with no-admixture uncorrelated allele frequencies and then developed a Bayesian clustering algorithm for *spatial* population genetics [9].

Cooperative clustering : I considered the task of producing more than one partitioning using the same data, the later partitionings being mutually linked. I proposed the notion of *cooperative clustering* which was very successfully applied to the joint segmentation of brain tissues and structures (see contribution 3). I used similar coupling ideas for the consistent depth and boundary/surface normal estimation from stereo image data [25, 35, 59].

Clustering with weights : Considering specifically multidimensional data, I addressed the issue of weighting the various dimension differently. I developed a clustering model in which weights are introduced and estimated in a Bayesian framework. This technique was successfully applied to the detection of brain lesions from multiple MR sequences (see contribution 3).

3. Non-standard **"noise" model** in an HMRF context :

HMRF extensions : I proposed a class of so-called *Triplet Markov models* where the HMRF structure is enriched by introducing a set of sub-class variables. They allow to account for more complex class-dependent distributions using standard hidden Markov models tools [8].

My contribution was in general to formalize and set the main modeling aspects. This implied to design new models or to extend existing ones. I also provided the inference step by designing appropriate estimation algorithms. I provided both these steps in my work on Markov models for gene clustering, on conjugate and cooperative clustering, on multidimensional data clustering with weights, and on the definition of new triplet Markov models. My contribution was focused on the modeling step, when dealing with missing and high-dimensional observations, and was related to the inference step in the population genetics and network traffic cases.

Originality and difficulty. I coordinated and combined the expertise of different members of team MISTIS including myself, which led to an original HMRF inference framework able to deal with dependent possibly incomplete high dimensional data. There are not so many existing tools that do both. Similarly, we mixed our expertise in missing data models and extreme value theory to produce with team RESO the original work in [29] which was accepted in a selective conference (SIGMETRICS, acceptance rate 15% in 2009).

Regarding my contribution to the POP project, it seems that there has been no attempt to use a generative model, such as ours, for the task of multimodal data interpretation. Similarly, as detailed in [3], there are intrinsic difficulties when one wants to cast the point registration problem in an EM framework.

In population genetics, my work [11] relies on the EM principle, and produces assignment rivaling other clustering programs. The work in [9] is particularly useful for

seeking genetic barriers or genetic discontinuities in continuous populations. It was one of the first models able to handle spatial population genetics data.

My cooperative clustering approach is powerful in that it provides general guidelines to decompose a joint process into parts to account for various types of interactions between multiple label sets.

Similarly, the framework I proposed to introduce weights in a multidimensional case has the advantage of allowing 1) the incorporation of a priori knowledge on the information content of each measure and 2) a weighting scheme which is modified adaptively. The model, applied to the detection of brain lesions, showed promising results.

The originality of the Triplet Markov models [8] comes from their ability to allow non standard assumptions while being tractable via standard algorithmic tools.

Validation and impact. The possibility to handle very high dimensional data and missing observations in an HMRF framework was an important achievement. It allowed us to prove the relevance of our models in a number of applications including gene clustering, image analysis (ACI Movistar) and hyper-spectral data clustering (ANR Vahine).

The work on multi-sensory data achieved in POP was selected for publication in ICT-RESULT [1]. Our work was presented as the main scientific results of the project at CogSys 2010, the 4th International Conference on Cognitive Systems in Zurich, Switzerland.

We were granted a new European project (HUMAVIPS 2010-13) within the highly competitive FP7-ICT program of the European Union, which will build on the innovation in POP.

My expertise in modeling spatial dependencies also generated collaborations which resulted in two papers [11, 9] and two software packages, FASTRUCT and TESS, the latter of which is still being developed by O. Francois and collaborators at TimC. I was part of the PhD committee of two students advised by O. Francois in 2006 and 2009, and O. Francois was in the committee of M. Vignes in 2007.

The impact on the medical imaging community is specified in the next section.

Dissemination. The related publications consist of 1 selective conference for the traffic data application [29], 5 conferences [32, 33, 34, 35, 25], 1 journal paper [1], 1 submitted paper [59] and the POPEYE software (see Technology development part) for our contribution in the POP project, 1 journal paper [5] and 5 conferences [42, 43, 50, 51, 47] for the genomics application. The population genetics application generated 2 journal papers [11, 9] and the FASTRUCT and TESS software registered at APP in 2006, under the CeCILLB license. High dimensional data and Triplet Markov models are developed in 1 journal paper [8] and 1 conference [40]. Our approach to point registration generated 1 journal paper [3] and 1 conference [39] with a corresponding software ECMPR (see Technology development part). Most of the other developments, except the MRI

1. `http://cordis.europa.eu/ictresults/index.cfm?section=news&tpl=article&BrowsingType=Features&ID=90953`

one, which has required a specific treatment, have been implemented in the SpaCEM^3 software available at `http://spacem3.gforge.inria.fr`. A user guide and data sets are also available there. The software has been registered at APP in 2007 under the CeCILLB license. A description of its main functionalities is also available in French [23].

3 Model-based clustering for the analysis of MR brain scans

In 2006, I initiated a collaboration with M. Dojat from the Grenoble Institute of Neuroscience (GIN) and C. Garbay (team MAGMA , Laboratoire d'informatique de Grenoble) for the analysis of Magnetic Resonance (MR) brain scans. We first focused on healthy brain MRI segmentation as it is a central step for a number of applications in neuroscience. From MR images, neuroradiologists are able to delineate tissues such as grey matter and structures such as Thalamus but this delineation cannot be performed easily automatically due to a number of artefacts. The first model we developed is based on the *cooperative clustering* framework I proposed (see contribution 2). It allows to capture the relationships between tissues and structures, to integrate a priori anatomical knowledge (via an atlas) and to handle local estimations and spatial correlations. It resulted in a procedure that is able to jointly register the atlas properly and to segment both tissues and structures with the three processes gradually and mutually improving each other. In parallel, I extended the collaboration to team VISAGES at INRIA in Rennes via the ARC project SeLMIC. The goal was to add two important features : 1) the possibility to deal with multiple complementary MR sequences and 2) the possibility to deal with abnormalities (lesions). I proposed then a probabilistic model for brain lesion segmentation, able to take into account multiple MR volumes and to weight these volumes differently depending on their informative power. In 2009, we started considering another type of MRI data, namely functional MRI data. The modeling is quite different and required that we develop new collaborations with other experts in the field. To this end, I started working with P. Ciuciu from CEA-Neurospin in Saclay [2, 24].

During the 2006-09 period, I co-advised one PhD-student (B. Scherrer), one post-doctoral fellow (S. Doyle), and one Master student (A. Janon). I also coordinated the ARC project SeLMIC. My participation in the collaboration was an important step toward the use of probabilistic and statistical tools. My contribution was first to provide the statistics and probability background and then to formalize the targeted brain imaging task. I designed the models and the estimation techniques.

Originality and difficulty. The originality comes from the successful combination of the teams' respective strengths *i.e.* expertise in distributed computing, in neuroimaging data processing and in statistical methods. It was a challenge to reach rapidly a high level in a domain outside my main domain of expertise. The relevance of our approach has been acknowledged by the community with the MICCAI best paper award for the

work in [31].

Validation and impact. MISTIS is the first statistics team to reach such a level of collaboration with the GIN. This is unique at INRIA Grenoble and among other statistics groups in Grenoble. No other team is involved in brain medical imaging or in a collaboration with the GIN. This work allowed me to initiate new collaborations with other prestigious experts in the field, such as with INRIA team VISAGES and CEA-Neurospin. As an additional achievement, B. Scherrer got a post-doctoral position at the prestigious Harvard Medical School in Boston.

Dissemination. Publications represent 3 journal papers [7, 6, 2], 1 book chapter [4], 9 international conferences [28, 31, 36, 37, 38, 41, 27, 26, 24], 2 national conferences [48, 49]. This includes 3 papers at the major medical imaging conference, MICCAI, among which a best paper award in 2008 [31]. It resulted in a software package called Locus, `http://locus.gforge.inria.fr` (website still under construction) that was selected for a regional technology fair in Grenoble, Forum 4i, May 2009 and for the Grenoble Innovation Fair (GIF) in October 2010.

Research Program

A statistical methodology for coupling models and methods

Faced with complex processes, with multiple sources of observations from various sensing devices, and with various degrees of scientific knowledge, the modeling task needs to take into account and to combine several sources of information. As a step toward the combination of methods that would deal alternatively but separately with different sources or sub-tasks, I consider *coupling* as a paradigm to allow for mutual interactions between components of a model. Such a coupling can be naturally expressed in a statistical framework via the definition of joint distributions. However, although it may be intuitive to consider processes from a joint perspective, it can present serious challenges in terms of statistical modeling. My goal is to develop a number of directions of research that involve coupling issues to eventually reach the definition of general guidelines to deal with complex joint processes from a statistical perspective. More specifically, I am planning to consider the following axis.

1 Bayesian conditional modeling for multimodal data analysis

When studying complex phenomena, it is crucial to be able to deal with various information sources. Not only several data sources can improve model identification by limiting the noise impact but also they can provide complementary information. However, the potential heterogeneity of data usually prevent the use of a joint model for inference. Even when it is possible to specify a joint multivariate distribution, inference may be possible only through equivalent conditional models. A coherently linked system of conditional models is often more appropriate. The approach I propose is to deal with joint modeling by replacing the joint distributions by a series of conditional models which fully define the joint model but not necessarily in an explicit way as in hierarchical modeling. Such models can be considered from either a classical or Bayesian perspective. I propose to focus on the latter one for its additional modeling capabilities in a number of applications. The *conjugate and cooperative clustering* approaches that I have recently developed, respectively for the analysis of multi-sensory and MR brain scans data, will be the basis of this research.

Audio-visual integration. In particular, I will go on studying the integration of audio-visual information through my participation in the new European STREP HUMAVIPS (Humanoids with auditory and visual abilities in populated spaces) which has started in February 2010 with partners including INRIA teams PERCEPTION and MISTIS, Czech Technical University, Aldebaran Robotics, IDIAP, and Bielefeld University. The challenge is to integrate statistical machine learning techniques with interactive robotic applications. Indeed, one of the most challenging and difficult tasks is to integrate unconstrained auditory, visual, and motor information such that a humanoid robot can learn how to localize, recognize, and interact in real-time with people. Hence, modeling the coupling of data

from physically different sensors plays a crucial role in this case.

Anatomical and functional MRI. I will also further address similar coupling issues in MRI and functional MRI analysis with my collaborators from the Grenoble Institute of Neuroscience and INRIA team VISAGES.

As regards studies on human cognitive processes, most current approaches are based on functional MRI data only. They do not usually consider the anatomical information related to the data. By contrast, it is important for interpretation purposes to relate functional information to anatomical structures. Also the injection of anatomical and/or functional a priori can greatly facilitate the detection of activation and limite false positive rate. A large number of successful approaches have involved statistical modeling. I believe this is one of the right approaches for the task although to my knowledge no real unifying framework has been proposed to date that fully exploits the modeling capabilities of probabilistic settings. My aim is to build on my previous achievements to go towards that goal. In particular, we will address the question of combining anatomical features with functional information in various forms.

Joint estimation-detection framework of brain activity. As regards functional MRI, a new collaboration with CEA Neurospin is on its way.

Traditional functional MRI analysis postulates a preset form of the so-called Hemodynamic Response Function (HRF) although, as demonstrated by various experiments, the HRF may vary between subjects and even between cortical regions in a given subject's brain. It appears then that fixing the HRF to some canonical shape is too restrictive and misleading. I will therefore address the issue of jointly detecting brain activity and estimating the HRF. Existing approaches along this line are based on simulation intensive Markov Chain Monte Carlo (MCMC) methods which prevent from easily integrating other prior information such as the subject brain anatomy or structure mentioned above. I will investigate the use of deterministic variational approximation techniques as a more efficient and more open alternative.

Spatio-temporal disease mapping. Then I will develop our collaboration with INRA and in particular the Clermont-Ferrand-Theix unit through the PhD thesis of L. Azizi. The goal is to couple spatial and temporal information to study epidemiology data. I will exploit the similarities that exist with the analysis of spatio-temporal audio-visual signals as encountered in the HUMAVIPS project.

2 Extreme value theory and structured data

This second axis represents another level of coupling related to the combination of ideas from different methodologies, namely extreme value theory and spatial statistics or Markov random field modeling.

My goal is to combine the expertise already available in MISTIS to develop a new domain, namely, *dependent extreme value theory*. Standard extreme value theory is a relatively recent domain in statistics which is part of MISTIS scientific bases. Taking

into account dependencies when studying extreme values is then a more recent area of research in which the MISTIS team has some advantages and some part to play. The two main directions I will consider are conditional extremes, to account for possible co-factors, and coupling traditional extreme value theory with spatial statistics analysis. In particular, the few existing approaches are based on simulation intensive inference techniques. The Markov Random Field modeling and variational approximations approaches that I have developed have not been used yet to this end, and they are promising. Applications include environmental issues and in particular rainfalls studies as initiated in MISTIS by S. Girard and L. Gardes in their collaboration with LTHE (Hydrology lab. in Grenoble).

3 Hybrid inference algorithms

This third axis refers to the design of algorithms for inference of complex structure models that would use both deterministic variational approximations and simulation techniques. While variational methods provide fast and reasonable approximate estimates, simulation methods offer more consideration of important theoretical issues such as accuracy of the approximation and convergence of the algorithms but at a much higher computational cost. The idea of hybrid approaches is to combine the main features and advantages of both simulation and deterministic methods. This kind of approach has not been fully exploited yet and the development of my *coupling* objectives above makes this exploitation more promising. Also I wish to keep a balanced activity between the main three parts of a statistical study, namely those related to modeling, estimation algorithms and theoretical properties.

4 Statistical image processing and cognition

My goal is then to investigate also the possibility to create a new team that would unify activities in medical imaging. The medical imaging activities in Grenoble are distributed between different labs with common interest but without any common structured action. There is a striking lack of structure around this topic although clear motivations and goals are expressed by various teams from some important laboratories. The GIN is the central entity as regards MRI due to its proximity to the Grenoble hospital and the MR device itself. Moreover, GIN researchers and engineers have strong expertise in MRI data acquisition and analysis. An overall theme that I foresee as a good structuring topic for such a team would be *Statistical image processing and cognition*. The collaboration between MISTIS and the GIN gives us some visibility and some advantage to be leader in such a project. Also, in terms of human resources, this should help attracting new researchers (permanent or non permanent) either from the GIN side which is affiliated to INSERM, Univ. Joseph Fourier and CEA, or from the MISTIS side which is an INRIA team also in the Laboratoire Jean Kuntzman affiliated to CNRS, Univ. Joseph Fourier, Univ. Pierre Mendès France and INPG.

5 Application domains and technology transfer

In addition to the applications already mentioned in the previous axis, I will go on developing a number of techniques in particular for solving issues in new applications related to microelectronics : process variability and defect detection in PC boards. Recent collaborations to be further developed include ones with CEA-Leti and VI-Technology, a world leader in Automatic Optical Inspection. I have started (October 2009) coordinating, on the MISTIS side, a *Pôle de compétitivité MINALOGIC* project including VI-Technology and Pige electronique as industrial partners (see Section *Industrial transfer* in the previous Technology development part). Also, the Locus software for brain MRI segmentation is now the object of a GRAVIT project whose goal is to seek for potential industrial users and partners.

6 International collaborations

In addition, the context of my research also includes collaborations with a number of international partners such as the Statistics Department of University of Washington in Seattle, the Speach and Hearing group of University of Sheffield, UK and new partners involved in the HUMAVIPS project. I will work at strengthening these collaborations to effectively combine the teams' respective strengths.

Detailed Curriculum Vitae

Last Name : FORBES First Name : Florence
Date and place of birth : 31/12/1970, Monaco (Monaco)
Citizenship : French Sex : F
Mailing address : INRIA Grenoble Rhône-Alpes
EPI MISTIS
655 avenue de l'Europe
Montbonnot 38334 Saint Ismier Cedex
Telephone : (+33) 4 76 61 52 50
E-mail : Florence.Forbes@inrialpes.fr
Web page : http://mistis.inrialpes.fr/people/forbes/

1 Diplomas

Ph.D. :
Title : Modèles markoviens de ressources partagées (*Markov models for resource sharing*)
Date of the defense of the Ph.D. : September 27, 1996
Granting institution : Joseph Fourier Univsersity (UJF), Grenoble
Host institution : SMS team, Laboratoire LMC, IMAG, Grenoble

Other diplomas :

- Ingénieur ENSIMAG (National institute for advanced study in computer science and applied mathematics), Grenoble,MSc in Computer Science and Mathematics. With highest honors.
  Graduation rank : 1/120.
- DEA (degree for doctoral candidacy), Joseph Fourier University, Grenoble, in applied probability with highest honors (mention TB).
  Graduation rank : 1/30.

2 Professional history

Current professional status

Position and statute : Senior Research Scientist (CR1)
Institution : INRIA Grenoble Rhône-Alpes
Start : January 2000

Previous professional experiences

Start	End	Institutions	
January 1998	Dec. 1999	INRIA Grenoble Rhône-Alpes	Junior Research Scientist (CR2)
October 1996	Dec. 1997	U. of Washington, Seattle, USA	postdoctoral fellow

3 Prizes and awards

My paper [31] with B. Scherrer, C. Garbay and M. Dojat, entitled "Fully Bayesian Joint Model for MR Brain Scan Tissue and Structure Segmentation" received the young investigator award at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) in 2008 in New-York, USA.

4 Supervision of research activities

PhD. students

Nathalie Peyrard [1998-2001, MENRT, UJF, Grenoble]. PhD thesis title : Approximation de type champ moyen des modèles de champ de Markov pour la segmentation de données spatiales. In this work we investigated the use of variational approximations (Mean-Field) in the context of Markov Random Fields. I initiated this activity in team IS2 and co-advised **[80%]** the work with G. Celeux (former team IS2). This work was at the core of my research program at that time. It became the basis of my activity regarding spatial data analysis (*e.g.* images) and implied a number of other theoretical and practical developments. The basis of our SpaCEM^3 software come from this work. Main related publications : [14, 15].

Nathalie is a senior researcher (CR1) at INRA in Toulouse.

Juliette Blanchet [2004-2007, MENRT, UJF, Grenoble]. PhD thesis title : Modèles markoviens et extension pour la classification de données complexes. We investigated and proposed extensions of standard Markov models, in particular the so-called *Triplet Markov models*. Texture segmentation was a key application. I advised **[90%]** this work with C. Schmid from team LEAR, INRIA Rhône-Alpes, in the context of a common ACI project called Movistar. The SpaCEM^3 software was initiated at that time. Main related publications : [8, 44].

Juliette is a post-doctoral fellow at the WSL Institute for Snow and Avalanche Research SLF, in Davos, Switzerland.

Matthieu Vignes [2003-2007, AC, Ecole Normale Supérieure de Lyon]. PhD thesis title : Modèles markoviens graphiques pour la fusion de données individuelles et d'interaction- application à la classification de gènes. We investigated the use of graphical Markov models in genomics, and in particular for

the clustering of gene expression data. I advised [**100%**] this work. It gave us the opportunity to initiate collaborations with statisticians and bio-statisticians from INRA Jouy-en-Josas and Toulouse. We were also involved in two ARC projects coordinated by M-F. Sagot on related subjects. Main related publication [5].

Matthieu is a junior researcher (CR2) at INRA Toulouse.

Benoit Scherrer [2005-2008, MENRT, INPG, Grenoble]. PhD thesis title : Segmentation des IRM cérébrales : agents markoviens et formulation bayesienne. I co-advised [**70%**] this work with M. Dojat (Grenoble Institut des Neurosciences) and C. Garbay (Laboratoire d'Informatique de Grenoble). I provided all the probabilistic and statistical background and most of the probabilistic modeling ideas. We developed a statistical framework for the analysis of MR brain scans (tissue and structure segmentations) that is able to combine several sources of information. This work is the key-stone in the collaboration I developed with the Grenoble Institute of Neuroscience. Main related publications : [4, 6, 7, 28, 31, 36].

Benoit is a post-doctoral fellow at the Harvard Medical school, Boston, USA.

Vasil Khalidov [2006-2010, POP European STREP, UJF Grenoble]. PhD thesis title : Conjugate Mixture Models for the Modeling of Visual and Auditory Perception. I co-advised [**70%**] this work with R. Horaud (PERCEPTION team, INRIA Grenoble Rhône-Alpes). The thesis was part of the European STREP named POP for Perception On Purpose. This PhD work contributed to the development of theoretical and algorithmic models based on probabilistic and statistical modeling of both the input and the processed data. Bayesian theory and hidden Markov models in particular have been combined with efficient optimization techniques in order to confront physical inputs and prior knowledge. Research along this line is going to be further addressed in the new STREP HUMAVIPS started in February 2010. Related publications : [1, 32, 33, 34].

Vasil is starting a post-doctoral position at IDIAP in Switzerland in November 2010.

Ramya Narahsimha [2006-2010, POP STREP, INPG, Grenoble]. PhD thesis title : Depth Recovery from Stereo Matching Using Coupled Random Fields. I co-advised [**50%**] this work with E. Arnaud and R. Horaud from team PERCEPTION. We investigated joint probabilistic frameworks for both disparity, boundary and surface normal estimation in the context of stereo matching problems. This work was initiated in the European STREP POP and is another part of my collaboration with computer vision experts. Related publications : [25, 35, 59].

Lamiae Azizi [2008-2011, CJC INRA , UJF, Grenoble]. I am co-advising [**70%**] this work with M. Garrido and D. Abrial from INRA Clermont-Ferrand-Theix.

We investigate the use of hidden Markov random fields for spatio-temporal disease risk mapping as an alternative to more standard hierarchical Gaussian models. No related publications yet.

Temporary engineer

Sophie Chopart [Sept. 2007-May 2009]. I advised [**50%**] the software development activity of Sophie related to the SpaCEM^3 software. Related publications : [23] and a SpaCEM^3 user guide.

Sophie is now a software engineer for the EB solutions company in Saint Cyr au Mont d'or, France.

Interns

Jean-Baptiste Durand [March-June 1999, ENSIMAG, MASTER statistics UJF, Grenoble]. I was partly involved [**20%**] in this Master work regarding gesture recognition using hidden Markov chains.

Romain Neugebauer [March-June 2000, Ecole Nationale Supérieure Agronomique, Montpellier, MASTER Biostatistics]. I advised [**100%**] this work on the use of statistical clustering techniques to identify breast carcinoma using angio-MRI data. This internship and the following were part of a collaboration with Adrian Raftery and Chris Fraley from University of Washington, Seattle and Toshiba inc. San Francisco. The overall goal was the analysis of dynamic breast MRI.

Armelle Regeasse [March-June. 2001, MASTER Biostatistics Univ. Montpellier]. I advised [**100%**] this work to investigate the use of data analysis techniques to help with the diagnosis of breast cancer.

Franz Chouly [June-Sept. 2001, ENSIMAG, 2nd year research project]. I co-advised [**50%**] this internship with Nathalie Peyrard. We generalized the use of Potts model for image segmentation by showing the impact of using an external field appropriately. The results of this study were an important step toward our contribution to the use of Markov models for image segmentation.

Cyril Martin [June-Sept. 2002, ENSIMAG, 2nd year research project]. The goal of this work that I advised [**100%**] was to investigated an alternative to hidden Markov Random Fields for image segmentation and more generally spatial data clustering. We studied conditional autoregressive Gaussian processes.

Matthieu Vignes [July-Sept. 2002, Ecole Normal Supérieure de Lyon, MASTER mathematics, UCB Lyon]. I was involved [**20%**] in this Master project

on the use of parsimonious higher order Markov chains for the analysis of DNA sequences.

Juliette Blanchet [March-June 2004, ENSIMAG, MASTER statistics, Univ. Paul Sabatier, Toulouse]. I advised [**100%**] this work on the use of hidden Markov field to cluster image descriptors. This study initiated our collaboration with team LEAR, which was then centralized in our common ACI project Movistar.

Jean-Baptiste Masson [March-June 2005, ENSIMAG, MASTER applied Maths. UJF, Grenoble]. I co-advised [**50%**] this Master work with G. Fort. The goal was to study the convergence properties of our mean-field like algorithms. This initiated my collaboration with G. Fort.

Lemine Abdalah [August-October 2005]. I supervised [**50%**] with J. Blanchet (MISTIS) the software development work of Lemine related to the SpaCEM^3 software.

Rajendran Narayanan [June-July 2007, INRIA Interships program]. I advised [**100%**] this work whose context was that of our ANR project Vahine. The goal was to study the concept of overlapping clustering for segmenting complex remote sensing data. Interesting preliminary results were obtained but the use on real hyperspectral data is still under consideration.

Lamiae Azizi [March-August 2008, bioMerieux, MASTER statistics, UPMF, Grenoble]. I co-advised this work [**30%**] with Laurent Gardes (MISTIS) and René Vachon from bioMerieux, Grenoble in the context of a 6 month contract with bioMerieux. We investigated the use of the Slice Inverse Regression (SIR) method for the analysis of a quantification process.

Luce Ponsar [March-June 2008, Veolia-eau, 3rd year research project, Ecole Centrale de Lyon]. I co-advised this work [**50%**] with J-B. Durand (MISTIS) in the context of a 4 month contract with Veolia-eau in Lyon. The goal was to study and possibly detect groups of individuals in time series describing various quantities linked to water consumption and billing in the Lyon area.

Alexandre Janon [April-August 2009, Ecole Normal Supérieure de Lyon, MASTER statistics, UJF, Grenoble]. I co-advised [**50%**] this work with Michel Dojat (Grenoble Institute of Neuroscience). The goal was to investigate the use of variational approximation techniques for the joint estimation-detection of brain activity in functional MRI data. This work initiated a new activity which is pursued with a new collaboration with Philippe Ciuciu from Neurospin, CEA in Saclay.

Eric Frichot [February-June 2010, ENSIMAG, 2nd year research project]. I advised [**100%**] the work of Eric on the extension of our Bayesian weighted model for brain lesion segmentation to the case of full covariance matrices.

5 Responsibilities

Organization of conferences and workshops

I was a member of the technical and organization committees of the following workshops :

The 3rd French-Danish workshop on spatial statistics and image analysis, Saint Pierre de Chartreuse, France, October 2003.

The "New directions in Monte-Carlo Markov chain methods" workshop, Fleurance, France, June 2007, `http://www.adapmc07.enst.fr/`.

The first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing (IEEE WHISPERS-09), August 2009, Grenoble, France, `http://www.ieee-whispers.com/2009/`.

The IEEE Signal Processing Society conference : Machine LEAR**ning for Signal Processing** (MLSP 2009), Sept. 2009, Grenoble, France, `http://mlsp2009.conwiz.dk`. I organized with W. Pieczynski from Telecom Sud Paris, a special session on "New trends in Markov models and related learning to restore data" [30].

The 17th summer session of the *Model-based clustering workshop*, Grenoble, July 19-23, 2010. See Section 7 for more details.

The 3th annual session of the *Statlearn workshop* (see the precedent edition webpage at `http://samos.univ-paris1.fr/-Statlearn-10-`) on statistical learning and complex data, Grenoble, March 17-18, 2011.

Committees

2009-2012 : Head of the committee for post-doctoral candidate selection at INRIA Grenoble Rhône-Alpes **("Comité des Emplois Scientifiques")**. I have been a member of this committee since 2005.

2009-2012 : Member of the INRIA national committee, "Comité d'animation scientifique", in charge of analyzing and motivating innovative activities in applied mathematics.

2007-2012 : Member of the INRA committee (CSS MBIA) in charge of evaluating INRA (French National Institute of Agricultural Research) researchers once a year.

2005-2007 : Member of the GTAI group in charge of incentive initiatives in the Scientific and Technological Orientation Council **(COST)** of INRIA.

2000-2008 : Member of the committee for Faculty member selection (Commissions de Spécialistes) at Institut Polytechnique de Grenoble (INPG).

1998-2004 : Member of the committee for Faculty member selection (Commissions de Spécialistes) at University Pierre Mendès France (UPMF, Grenoble II).

2002 : Member of the committee for research scientist candidate (CR) selection at INRIA Grenoble Rhône-Alpes.

PhD thesis and HDR Committees

November 2005 : Christelle Melo de Lima from University Lyon 1. PhD title : Développement d'une approche markovienne pour l'analyse de l'organisation des génomes.

September 2006 : Mathieu Emily from TimC lab. INPG, Grenoble. PhD title : Modèles statistiques du dévelopement de tumeurs cancéreuses.

September 2009 : Eric Durand from TimC Lab. INPG, Grenoble. PhD title : Modèles statistiques pour la structure génétique des populations : organisation spatiale et liens de parentés.

March 2010 : Tomas Crivelli from team VISTA INRIA Rennes, Univ. Rennes I. PhD title : Mixed state Markov models for image motion analysis.

November 2010 : Lotfi Châari from University Paris-Est. PhD subject : Reconstruction d'images médicales d'IRM à l'aide de représentations en ondelettes.

December 2010 : Nicolas Wicker is assistant professor at Strasbourg University. I am a reviewer for the HDR (Qualification to coach researchers) thesis of Nicolas Wicker on computational statistics.

6 Management

Research team MISTIS

Scientific leadership : I am the scientific leader of the MISTIS group `http://mistis.inrialpes.fr` since its creation in 2003. I have defined its scientific bases and I have initiated several interdisciplinary collaborations and industrial contracts. I wrote several European, ANR, ARC project proposals that were eventually granted.

Projects : The team is currently involved in 1 European STREP project (HUMAVIPS), 2 ANR projects (VAHINE and MEDUP), 2 Minalogic project (I-VP and OPTYMIST II).

Past projects include a STREP project (POP), 1 ACI project (MOVISTAR), 3 ARC projects (SeLMIC, ChromoNet and IBN), 3 industrial contracts (CEA, BioMerieux, Veolia), 1 CIFRE contract with Xerox.

Projects leadership : I have been directly involved in HUMAVIPS, VAHINE, I-VP, OPTYMIST II, POP, MOVISTAR, SeLMIC, ChromoNet, IBN, bioMerieux and Veolia projects.

Team composition : In September 2010, the team is composed of 15 persons including 2 INRIA senior researchers (S. Girard and myself), 2 University Pierre Mendès France (UPMF) faculty members (L. Gardes and M-J. Martinez) and 1 ENSIMAG (INPG) faculty member (J-B. Durand). The current number of PhD students is 4, the current number of post-doctoral fellows is 5 and there is 1 temporary engineer.

Supervision : Among the 15 members, I am advising 2 PhD students and I am directly working with 3 post-doctoral fellows and 1 engineer.

MISTIS is also part of the Statistics Department of Laboratoire Jean Kuntzman.

The table below shows the budget allocated to the project-team via external funding, in keuros, over the 2006-2012 period.

2006-2012	k euros
INRIA Research Initiatives	
SeLMIC (ARC†)	27
National initiatives	
VAHINE (ANR)	104.7
European projects	
POP STREP FP6	140.6
HUMAVIPS STREP FP7	130
Industrial contracts	
bioMerieux	1.5
Veolia	4.2
Xerox	9
CEA	20
MINALOGIC (I-VP)	442
Scholarships	
PhD * L. Azizi (INRA)	100
PhD * A. Lekina (Cordi-INRIA)	100
Post Doc* C. Chen (INRIA)	46
Post Doc* S. Doyle (SeLMIC)	46
Post Doc* MEDUP (ANR)	69
Post Doc* D. Wraith (INRIA)	69
Post Doc* OPTYMIST II (UJF)	92
ODL# S. Chopart	90
Total (k euros)	1491

† INRIA Cooperative Research Initiatives
∗ other than those supported by one of the above projects
engineer supported by INRIA

Research Projects

European STREP HUMAVIPS (2010-13) MISTIS is involved in a new three-year European project (STREP) started in February 2010. The project is named HUMAVIPS (Humanoids ables with auditory and visual abilities in populated spaces) and was in 2009/10 the only INRIA coordinated project granted in the highly competitive FP7-ICT program of the European Union. The partners involved are team Perception and Mistis from INRIA Rhone-alpes (coord.), the Czech Technical University CTU Czech Republic, Aldebaran Robotics ALD France, Idiap Research Institute Switzerland and Bielefeld University BIU Germany. The goal is to develop humanoid robots with integrated audio-visual perception systems and social skills, capable of handling multi-party conversations and interactions with people in realtime. My contribution will consist in developing statistical machine learning techniques for interactive robotic applications.
The financial support for MISTIS is of **130 keuros**.

The MINALOGIC I-VP project (2009-12). Since October 2009, I am coordinating new activities for the three-year *Pôle de competitivité Minalogic* project I-VP (for Intuitive Vision Programming) supported by the French government. The project is led by VI Technology (`http://www.vitechnology.com`), a world leader in Automated Optical Inspection (AOI) of a broad range of electronic components. The other partners involved are the CMM (Centre de Morphologie Mathematiques) in Fontainebleau, and Pige Electronique in Bourg-Les-Valence. The NOESIS company, which is a leader in the field of image processing and analysis software, in Crolles, is also involved to provide help with software development. Our goal is to exploit more intensively statistical and image processing techniques to improve defect detection capability and programming time based on existing AOI principles so as to reach eventually a reliable defect detection with virtually zero programming skills and efforts. The financial support for MISTIS in about **442 keuros** and implies the hiring and coordination of 102 man/month.

The OPTYMIST II project (2009-12). I am involved in this three year *Pôle de competitivité Minalogic* project through a collaboration with CEA - LETI addressing variability issues when designing electronic devices. The partners involved are University Joseph Fourier (LJK), CEA-LETI, and Raise Partner, ASYGN and Dolphin Integration companies. The financial support corresponds for MISTIS to a 2-year post-doctoral position (**92 k euros**).

ANR VAHINE (2008-10). In the MDCO (Masse de Données et Connaissances) program, this three-year project is called "Visualisation et analyse d'images hyperspectrales multidimensionnelles en Astrophysique" (VAHINE). It aims at developing physical as well as mathematical models, algorithms, and software able to deal efficiently with hyperspectral multi-angle data but also with any other kind of large hyperspectral dataset (astronomical or experimental). It involves the Laboratoire de Planétologie (LPG) de Grenoble, the Gipsa-lab, Observatoire

de la Côte d'Azur (Nice), and University Strasbourg I. I initiated this collaboration and organized numerous meetings with the collaborators beforehand. I am in charge of coordinating the MISTIS activities which involves an 18 months post-doctoral fellowship. The financial support was of **104 keuros**. More details on the web site `http://mistis.inrialpes.fr/vahine/dokuwiki-2008-05-05/`.

European STREP POP (2006-08). I coordinated the MISTIS part for the POP European STREP (Perception On Purpose) coordinated by INRIA and involving teams MISTIS and PERCEPTION. The other partners involved were the University of Sheffield, UK, University of Osnabrck, University Hospital Hamburg-Eppendorf in Germany and University of Coimbra in Portugal. The objective was to put forward the modeling of perception (visual and auditory) as a complex attentional mechanism that embodies a decision taking process. The part I had in charge included the PhD works of Vasil Kalidhov and Ramya Narasimha. The financial support for MISTIS was of **140 keuros**. Our contribution was to the development of theoretical and algorithmic models based on probabilistic and statistical modeling of both the input and the processed data. The final review of the project was held on December 11 and 12, 2008 with in particular a live demo running on the POP audio-visual head regarding multispeaker localisation using the algorithm we developed [1]. Further details on the project web site `http://perception.inrialpes.fr/POP/`.

The SeLMIC project (2007-08). I coordinated (2007-2008) the 2-year ARC project SeLMIC supported by INRIA, `http://euterpe.ujf-grenoble.fr/selmic/doku.php`. SeLMIC stands for *Segmentation Longitudinale et multimodale par IRM d'anormalités cérébrales*. It aimed at developing new statistical methods for the segmentation of multidimensional MR sequences corresponding to different types of MRI modalities and longitudinal data. The applications included the detection of brain abnormalities and more specifically strokes and Multiple Sclerosis lesions. The partners involved were team VISAGES from INRIA in Rennes, the INSERM Unit U594 (Grenoble Institute of Neuroscience) and Laboratoire d'Informatique de Grenoble (LIG).

The funding included a one-year post-doctoral fellowship and about **27 keuros**.

ACI MOVISTAR (2003-06) MISTIS got a Ministry grant (Action Concertée Incitative Masses de données) for a three-year project involving other partners, team LEAR from INRIA, SMS from University Joseph Fourier and Heudiasyc from UTC, Compiègne. I initiated this project which aimed at investigating visual and statistical models for image recognition and description and learning techniques for the management of large image databases. The PhD work of Juliette Blanchet and Charles Bouveyron were related to this project.

7 Collaborations, visits

Long visits

1996-1997 (15 months) : Post-doctoral position in the Statistics Department of University of Washington, Seattle, USA.

1998-2002 : 5 long visits (3 to 4 months each) to the Statistics Department of University of Washington.

1997-2010 : Participation to the summer session of the Model-based clustering workshop, on invitation only.

I visited the **Statistics Department of University of Washington in Seattle,** as a post-doctoral fellow from October 1996 to December 1997. I initiated there common work with Adrian Raftery `http://www.stat.washington.edu/raftery/` on models for clustering and more specifically for statistical image segmentation [17]. I then visited the department on a regular basis for periods of 3 to 4 months. **I made 5 such long visits from 1998 to 2002.** In 1999, I visited the department with N. Peyrard whose thesis was mainly on Markov model based segmentation using the Expectation-Maximization (EM) algorithm and variational approximation techniques. Markov model selection was also considered in this work. The focus was on BIC criterion and A. Raftery's expertise on Bayesian model selection was very relevant to the collaboration. As a matter of fact, in 2001, A. Raftery was involved in N. Peyrard's PhD committee. In addition, I was also involved with N. Peyrard, A. Raftery and C. Fraley (another member of the Statistics Department) in a project involving MRI data with partners such as Toshiba Inc. and the Seattle University Hospital. In the following years, model-based clustering techniques (and the various theoretical and practical aspects mentioned above) was the main topic of the collaboration. Medical image analysis and more specifically breast MRI data played a central part. I was involved in 2 projects contracted by the University of Washington with NIH (2001) and ONR (1998). Three technical reports (1999 [65], 2001 [64], 2004 [62]) show the evolution of our collaboration regarding both statistical methods and clinical data. The journal paper [12] shows that results have been obtained that point out the relevance of statistical methodology to the MRI community. In addition, since 1997, I have been involved nine times in an invitation only *Model-based clustering Workshop* organized every summer by A. Raftery and collaborators (http ://www.stat.washington.edu/raftery/Onrgroup/onrgroup.html). Each year various teams are represented including students and researchers from the Univ. of Washington Statistics Department, INRIA (MISTIS, SELECT, and more recently PERCEPTION), INRA (Avignon, Toulouse), University College London, University of London, Trinity College and University College Dublin, University of Montreal, Texas University, Microsoft Research, etc. The workshop was in Seattle until 2006 but the 2007 and 2009 sessions took place resp. in Trinity College, Dublin and IHP, Paris. For 2010, I was in charge of organizing the 2010 session that took place at INRIA Grenoble Rhône-Alpes in July.

Collaborations

One major strength of my research activity is the numerous collaborations that I have initiated and developed. This contributes to the good visibility of team MISTIS. Among the major achievements, I have been able to initiate and develop the following strong collaborations :

Statistics Department, University of Washington, Seattle, USA. I already mentioned above this long-term collaboration.

Grenoble Institute of Neuroscience (GIN). The collaboration with the GIN, on the analysis of MR brain scans, has included a co-advised PhD student, a post-doctoral fellow, a Master student, a number of co-publications (2 journal papers [7, 6], 1 book chapter [4], 8 international conferences [28, 31, 36, 37, 38, 41, 27, 26], 2 national conferences [48, 49]) and a software (LOCUS). This collaboration also includes C. Garbay from **team** MAGMA **of LIG** for multi-agent systems aspects. I also coordinated the ARC project SeLMIC (Section 6) which allowed to initiate collaboration with team VISAGES from INRIA in Rennes (C. Barillot and collaborators).

CEA Saclay, Neurospin. More recently I have started working with P. Ciuciu from CEA-Neuropsin on functional MRI [2, 24]. More details are given in the Application domains section.

Lab. of Planetology (LPG) and Gipsa-Lab. Grenoble. This collaboration with the Gipsa-lab (Jocelyn Chanussot, Christian Jutten and collaborators) and Lab. of Planetology in Grenoble (Sylvain Douté and collaborators) was initiated before the Vahine ANR project (See Section 6). It includes two post-doctoral fellows and a number of common publications (1) and communications (4). In addition, I was a member of the technical and organization committees of the following workshops coordinated by J. Chanussot and C. Jutten (see Section 5 for details) : The first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing (IEEE WHISPERS-09) and the IEEE Signal Processing Society conference : Machine LEARning for Signal Processing (MLSP 2009).

SandH, University of Sheffield, UK. This collaboration with the Speach and Hearing group (SandH) was initiated with the European STREP POP. We organized several cross-visits alternatively in Sheffield and Grenoble for periods of 1 to 2 weeks on average. This was essential for us to combine our respective expertise in image analysis and auditory data analysis. In addition, it resulted in the creation of a data-base with audio-visual sequences made from binaural and binocular recordings [34]. The data-base is publicly available and quite unique.

TimB team, TIMC Lab. Grenoble. I also have common interest with Prof. Olivier Francois and collaborators from TimB in the TIMC lab. I have started working with O. Francois during my PhD thesis (1993-96) [22, 20, 19]. Later we co-advised

a post-doctoral fellow, Chibiao Chen (2006), on the application of hidden Markov models to population genetics. This resulted in two papers [11, 9] and two software, FASTRUCT and TESS, the latter of which is still developed at TimB. I was part of the PhD committee of two students advised by O. Francois in 2006 and 2009, and O. Francois was in the committee of M. Vignes in 2007. We are currently interacting for the PhD thesis of L. Azizi on the use of covariates in epidemiology study.

INRA Clermont-Ferrand-Theix. I developed a collaboration with the Clermont-Ferrand unit, with new applications to epidemiology, including a co-advised PhD student (L. Azizi) that started in October 2008 . I am also an active members of the INRA network called MSTGA on graphical models and spatial statistics, `http://carlit.toulouse.inra.fr/MSTGA/`.

INRIA team PERCEPTION**.** This collaboration includes two co-advised PhD thesis and a number of co-signed publications. My work on clustering heterogeneous multi-sensor data is part of MISTIS contribution to the European project POP and was joint with Radu Horaud, Miles Hansard and Elise Arnaud from PERCEPTION [1, 32, 33, 34]. This collaboration is going to be further developed with the new European project HUMAVIPS. In a different context, I also focused on Point Registration in computer vision [3, 39].

INRIA team LEAR. The ACI project Movistar involved C. Schmid and collaborators from team LEAR, SMS from University Joseph Fourier and Heudiasyc from UTC, Compiègne. The PhD work of Juliette Blanchet and Charles Bouveyron were related to this project. It resulted in 3 co-authored journal papers, 4 international conference papers and 3 national conference papers. I was myself involved in co-publications [44, 45, 52, 53].

INRIA team VISAGES**, Rennes.** Initiated in the SeLMIC project, this collaboration has focused on the analysis of pathological brain MR data. The work in [27, 26] shows a successful application to Multiple Sclerosis and Stroke lesion detection.

INRIA team RESO**, Lyon.** We have initiated a collaboration on the use of statistical methods for the analysis of network traffic data. This resulted in a paper accepted at the selective (15% acceptance rate) SIGMETRICS conference in 2009 [29].

8 Teaching

1993-1996. I taught at University Joseph Fourier (*DEUG*) and ENSIMAG, INPG (1st year, BSc level) in Grenoble, statistics and probability courses. Total= **96 hours** each year.

1998-2000. I taught at ENSIMAG, INPG (1st year, BSc level) a probability course (TD). Total= **30 hours** each year.

2002-2005. I taught at ENSIMAG, INPG (2nd year Telecom, MSc level) a stochastic processes course (TD). Total= **18 hours** each year.

2001-2003. I taught at Polytech Grenoble, University Joseph Fourier (2nd year, BSc level) a statistics course. Total= **16 hours** each year.

Since 2005, I am teaching at University Joseph Fourier a statistics course in the *Statistical engineering MASTER pro* (MSc level). Total= **24 hours** each year.

This corresponds to a **total of 618 hours** and an **average of 39 hours per year** from 1993 to 2010.

9 Dissemination of scientific knowledge

The LOCUS software (http ://locus.gforge.inria.fr) developed in collaboration with the Grenoble Institute of Neuroscience and LIG (MAGMA team), for brain MRI analysis, was selected to be presented at the **FORUM 4i (Technology fair)** in Grenoble in May 2009. For this event we prepared a 2 page flyer and a video for non specialists. They are available at `http://mistis.inrialpes.fr/people/forbes/dissemination`. The software is also going to be presented at the Grenoble Innovation Fair (GIF) in October 2010.

10 Other relevant information

Maternity leaves (two children) :

2003 : 6 months.

2006 : 6 months.

Publications

1 Articles in international refereed journals and book chapters

[1] V. Khalidov, F. Forbes, and R. Horaud. Conjugate mixture models for clustering multimodal data. *Neural Computation*, 6(1-2) :48–83, 2010.

[2] L. Risser, T. Vincent, F. Forbes, J. Idier, and P. Ciuciu. Min-max extrapolation scheme for fast estimation of 3D potts field partition functions. Application to the joint detection-estimation of brain activity in fMRI. 2010. Special issue of Journal of Signal Processing Systems.

[3] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang. Rigid and articulated point registration with expectation conditional maximization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010. To appear.

[4] B. Scherrer, F. Forbes, C. Garbay, and M. Dojat. A joint Bayesian framework for MR brain scan tissue and structure segmentation based on distributed Markovian agents. In I. Bichindaritz and eds. L. Jain, editors, *Computational Intelligence in Healthcare*. Springer-Verlag, Berlin, 2010. To appear.

[5] M. Vignes and F. Forbes. Gene clustering via integrated Markov models combining individual and pairwise features. *IEEE Trans. on Computational Biology and Bioinformatics*, 6(2) :260–270, 2009.

[6] B. Scherrer, F. Forbes, C. Garbay, and M. Dojat. Distributed Local MRF Models for Tissue and Structure Brain Segmentation. *IEEE Trans. Medical Imaging*, 28 :1296–1307, 2009.

[7] B. Scherrer, M. Dojat, F. Forbes, and C. Garbay. Agentification of Markov model based segmentation : Application to MRI brain scans. *Artificial Intelligence in Medicine (AIM)*, 46(1) :81–95, 2009.

[8] J. Blanchet and F. Forbes. Triplet Markov fields for the supervised classification of complex structure data. *IEEE Trans. on Pattern Analyis and Machine Intelligence*, 30(6) :1055–1067, 2008.

[9] C. Chen, E. Durand, F. Forbes, and O. Francois. Bayesian clustering algorithms ascertaining spatial population structure : a new computer program and a comparison study. *Molecular Ecology Notes*, 7(5) :747–756, 2007.

[10] F. Forbes and G. Fort. Combining Monte Carlo and Mean field like methods for inference in hidden Markov Random Fields. *IEEE Trans. on Image Processing*, 16(3) :824–837, 2007.

[11] C. Chen, F. Forbes, and O. Francois. FASTRUCT : Model-based clustering made faster. *Molecular Ecology Notes*, (6) :980–983, 2006.

[12] F. Forbes, N. Peyrard, C. Fraley, D. Georgian-Smith, D. Goldhaber, and A. Raftery. Model-based region-of-interest selection in dynamic breast MRI. *Journal of Computer Assisted Tomography*, 30(4) :675–687, July/August 2006.

[13] G. Celeux, F. Forbes, C.P. Robert, and M. Titterington. Deviance information criteria for missing data models. with discussion. *Bayesian Analysis*, 1(4) :651–706, 2006.

[14] G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognition*, 36(1) :131–144, 2003.

[15] F. Forbes and N. Peyrard. Hidden Markov model selection based on mean field like approximations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(8), 2003.

[16] G. Celeux, S. Chrétien, F. Forbes, and A. Mkhadri. A component-wise EM algorithm for mixtures. *Journal of Computational and Graphical Statistics*, 10 :699–712, 2001.

[17] F. Forbes and A. E. Raftery. Bayesian morphology : Fast unsupervised bayesian image analysis. *Journal of the American Statistical Association*, 94(446) :555–568, June 1999.

[18] F. Forbes and B. Ycart. Counting stable sets on Cartesian products of graphs. *Discrete Mathematics*, 186(1-3) :105–116, 1998.

[19] F. Forbes and O. Francois. Increasing couplings for interacting particle systems. *C.R. Acad. Sci. Paris, Tome 324, Srie I, Rubrique Probabilits*, (4) :459–464, 1997.

[20] F. Forbes and O. Francois. Stochastic comparison for Markov processes on a product of partially ordered sets. *Statististics and Probability letters*, 33(3) :309–320, 1997.

[21] F. Forbes and B. Ycart. The philosophers' process on ladder graphs. *Communication in Statististics, Stochastic Models*, 12(4) :559–583, 1996.

[22] F. Forbes, O. Francois, and B. Ycart. Stochastic comparison for resource sharing models. *Markov Processes and Related Fields*, 2(4) :581–605, 1996.

2 Articles in national refereed journals

[23] J. Blanchet, F. Forbes, S. Chopart, and L. Azizi. Le logiciel SpaCEM^3 pour la classification de donnees complexes. *La revue Modulad*, 40 :147–166, 2009.

3 Publications in international Conferences and Workshops

[24] L. Risser, T. Vincent, F. Forbes, J. Idier, and P. Ciuciu. How to deal with brain deactivations in the joint detection-estimation framework ? In *Human Brain Mapping (HBM) meeting*, Barcelone, Spain, 2010.

[25] R. Narasimha, E. Arnaud, F. Forbes, and R. Horaud. Disparity and normal estimation through alternating maximization. In *17th IEEE Int. Conf. Imag. Proc. ICIP 10, Hong-Kong*, 2010.

[26] F. Forbes, S. Doyle, D. Garcia-Lorenzo, C. Barillot, and M. Dojat. A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation. In *13th International Conference on Artificial Intelligence and Statistics (AISTATS10)*, Sardinia, Italy, 13-15 May 2010.

[27] F. Forbes, S. Doyle, D. Garcia-Lorenzo, C. Barillot, and M. Dojat. Adaptive weighted fusion of multiple MR sequences for brain lesion segmentation. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, Rotterdam, The Netherlands, 14-17 April 2010.

[28] B. Scherrer, F. Forbes, and M. Dojat. A conditional random field approach for coupling local registration with robust tissue and structure segmentation. In *MICCAI 2009, London, UK*, 2009.

[29] P. Loiseau, P. Gonçalves, S. Girard, F. Forbes, and P. Primet Vicat-Blanc. Maximum likelihood estimation of the flow size distribution tail index from sampled packet data. In *SIGMETRICS–Joint International Conference on Measurement and Modeling of Computer Systems*, Seattle, juin 2009.

[30] F. Forbes and W. Pieczynski. New trends in Markov models and related learning to restore data. In *IEEE conf. on Machine Learning and Signal Processing, MLSP 09, Grenoble, France*, 2009.

[31] B. Scherrer, F. Forbes, M. Dojat, and C. Garbay. Fully Bayesian Joint Model for MR Brain Scan Tissue and Structure Segmentation. Received the Young Investigator Award in Segmentation. In *MICCAI 2008, New-York, USA*, pages 1066–74, 2008.

[32] V. Khalidov, F. Forbes, M. Hansard, E. Arnaud, and R. Horaud. Audio-visual clustering for 3D speaker localization. In *5th joint Workshop on Machine Learning and Multimodal Interaction MLMI 2008, Utrecht, The Netherlands*, pages 86–97, 2008.

[33] V. Khalidov, F. Forbes, M. Hansard, E. Arnaud, and R. Horaud. Detection and localization of 3D audio-visual objects using unsupervised clustering. In *ACM/IEEE International Conference on Multimodal Interfaces (ICMI 08)*, pages 217–224, 2008.

[34] E. Arnaud, H. Christensen, Y-C. Lu, J. Barker, V. Khalidov, M. Hansard, B. Holveck, H. Mathieu, R. Narasimha, E. Taillant, F. Forbes, and R. Horaud. The cava corpus : synchronised stereoscopic and binaural datasets with head movements. In *ACM/IEEE International Conference on Multimodal Interfaces (ICMI 08)*, pages 109–116, 2008.

[35] R. Narasimha, E. Arnaud, F. Forbes, and R. Horaud. Cooperative disparity and object boundary estimation. In *15th IEEE Int. Conf. Imag. Proc. ICIP 08, San Diego, USA*, pages 1784–1787, 2008.

[36] B. Scherrer, M. Dojat, F. Forbes, and C. Garbay. LOCUS : LOcal Cooperative Unified Segmentation of MRI Brain Scans. In *10th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI07*, pages 219–227, Brisbane, Australia, 2007.

[37] B. Scherrer, M. Dojat, F. Forbes, and C. Garbay. MRF Agent Based Segmentation : Application to MRI Brain Scans. In *11th Conference on Artificial Intelligence In Medicine, AIME07*, pages 13–23, Amsterdam, Netherlands, 2007.

[38] Y. Kabir, M. Dojat, B. Scherrer, F. Forbes, and C. Garbay. Multimodal MRI segmentation of ischemic stroke lesions. In *29th Annual International Conference of the IEEE Engineering in Medicine and biology Society (EMBC)*, Lyon, France, 2007.

[39] G. Dewaele, F. Devernay, R. Horaud, and F. Forbes. The alignment between 3D-data and articulated shapes with bending surfaces. In *European Conf. Computer Vision, Lecture notes in Computer Science*, number 3, pages 578–591, 2006.

[40] J. Blanchet and F. Forbes. Triplet Markov fields designed for supervised classification of textured images. In *COMPSTAT, 17th symposium of the IASC*, Roma, Italy, 2006.

[41] B. Schrerrer, M. Dojat, F. Forbes, and C. Garbay. Distributed and Cooperative Markovian Segmentation of tissues and structures in MRI brain scans. In *Human Brain Mapping (HBM) meeting*, Florence Italy, 2006.

[42] M. Vignes and F. Forbes. A statistical glance at clustering models to fit biological network and expression data. In *31st conference on Stochastic Processes and their Applications*, Paris, France, 2006.

[43] M. Vignes and F. Forbes. Integrated Markov models for clustering genes combining individual features and pairwise relationships. In *4th workshop on Statistical methods for post-genomic data*, Toulouse, France, 2006.

[44] J. Blanchet, F. Forbes, and C. Schmid. Markov random fields for textures recognition with local invariant regions and their geometric relationships. In *British Machine Vision Conference*, Oxford, UK, September 2005.

[45] J. Blanchet, F. Forbes, and C. Schmid. Markov random fields for recognizing textures modeled by feature vectors. In *International Conference on Applied Stochastic Models and Data Analysis*, Brest, France, May 2005.

4 Publications in national Conferences and Workshops

[46] M. Vignes, J. Blanchet, D. Leroux, and F. and Forbes. Clustering of incomplete, high dimensional and dependent biological data with spacem3. In *Journee satellite MODGRAPH 2010 de JOBIM*, Montpellier, France, Septembre 2010.

[47] J. Blanchet, F. Forbes, and M. Vignes. Clustering with incomplete dependent data. In *39eme Journées de Statistique organisées par la Société Française de Statistique, SFDS*, Angers, France, 2007.

[48] B. Schrerrer, M. Dojat, F. Forbes, and C. Garbay. Segmentation markovienne distribuée et coopérative des tissus et des structures présents dans des IRM cérébrales. In *RFIA*, Tours, France, 2006.

[49] B. Schrerrer, M. Dojat, F. Forbes, and C. Garbay. Une Approche SMA pour la Segmentation Markovienne des Tissus et Structures Presents dans les IRM Cerebrales. In *JETIM*, Alger, 2006.

[50] M. Vignes and F. Forbes. Integrated Markov models for clustering gene expression data. In *Journées MAS de la SMAI*, Lille, France, 2006.

[51] M. Vignes and F. Forbes. Markov random fields for clustering genes. In *2eme Recontres Inter-Associations : la classification et ses applications*, Lyon, France, 2006.

[52] J. Blanchet, F. Forbes, and C. Schmid. Modèles markoviens pour l'organisation spatiale de descripteurs d'images. In *7e Conférence francophone sur l'Apprentissage Automatique, Presses Universitaires de Grenoble*, pages 113–126, Nice, France, Juin 2005.

[53] J. Blanchet, F. Forbes, and C. Schmid. Modèles markoviens pour la reconnaissance de textures à partir de descripteurs locaux et de leur organisation spatiale. In *37e Journées de Statistique de la Société Francaise de Statistique*, Pau, France, Juin 2005.

[54] N. Peyrard, F. Forbes, and D. Allard. Comparaison de deux modélisations pour la classification de données géostatistiques. In *37èmes Journées de Statistique organisées par la Société Française de Statistique*, Pau, juin 2005.

[55] F. Forbes and M. Vignes. Champs de markov cachés et fusion de données individuelles et pairées pour l'identification de groupes de gènes. In *JOBIM*, Lyon, France, Juillet 2005.

[56] G. Celeux, F. Forbes, and N. Peyrard. Modèle de Potts avec champ externe et algorithme de type EM pour la segmentation d'image. In *RFIA*, Toulouse, France, janvier 2004.

5 Internal Reports

[57] V. Khalidov, F. Forbes, and R. Horaud. Conjugate mixture models for clustering multimodal data. Rapport de recherche RR-7117, INRIA, November 2009.

[58] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang. Rigid and articulated point registration with expectation conditional maximization. Rapport de recherche RR-7114, INRIA, November 2009.

[59] R. Narasimha, E. Arnaud, F. Forbes, and R. Horaud. A joint framework for disparity and surface normal estimation. Rapport de recherche RR-7090, INRIA, October 2009.

[60] Juliette Blanchet and Florence Forbes. Triplet Markov fields for the classification of complex structure data. Research Report RR-6356, INRIA, 2007.

[61] F. Forbes and G. Fort. A convergence theorem for variational EM-like algorithms : application to image segmentation. Technical Report RR-5721, Inria Rhône-Alpes, 2005. http ://www.inria.fr/rrrt/rr-5721.html.

[62] F. Forbes, N. Peyrard, C. Fraley, D. Georgian-Smith, D. Goldhaber, and A. E. Raftery. Model-based region of interest selection in dynamic breast MRI. Technical Report 472, Stat. dept, Univ. of Washington, 2004.

[63] G. Celeux, F. Forbes, C. P. Robert, and M. Titterington. Deviance information criteria for missing data models. Technical Report RR-4859, Inria Rhône-Alpes, 2003.

[64] F. Forbes, N. Peyrard, C. Fraley, D. Georgian-Smith, D. Goldhaber, and A. E. Raftery. Region of interest selection and dynamic breast MRI data analysis using multivariate statistical methods for clustering and spatial segmentation. Technical Report RR-4249, Inria Rhône-Alpes, 2001.

[65] F. Forbes, C. Fraley, N. Peyrard, and A.E. Raftery. Spatial Statistical Analysis of Breast Magnetic Resonance Images via Model-based Clustering, Morphology and Markov Random Fields. Technical report, Report for Toshiba inc., 1999.

6 Others

[66] D. Allard, F. Forbes, and N. Peyrard. Comparing two models for clustering geostatistical data. In *Working group on model-based clustering, Tenth anniversary Summer Session*, Univ. of Washington, Seattle, USA, July 2004.

© *2010 by the author. This thesis may be reproduced, in its entirety, for non-commercial purposes.*