View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Infoscience - École polytechnique fédérale de Lausanne

# A Spin Glass Model of a Markov Random Field

B. Caputo
Computational Vision and Active Perception Laboratory
Dept. of Numerical Analysis and Computer Science
Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden
e-mail: caputo@nada.kth.se

## Abstract

*This paper presents a novel algorithm for robust object recognition. We propose to model the visual appearance of objects via probability density functions. The algorithm consists of a fully connected Markov random field with energy function derived from results of statistical physics of spin glasses. Markov random fields and spin glass energy functions are combined together via nonlinear kernel functions; we call the model Spin Glass-Markov Random Field. Full connectivity enables to take into account the global appearance of the object, and its specific local characteristics at the same time, resulting in robustness to noise, occlusions and cluttered background. We show with theoretical analysis and experiments that this new model is competitive with state-of-the-art algorithms.*

## 1 Introduction

Creating computer methods for automatic object recognition has long been one of the major goals for computer vision, and it gives rise to challenging theoretical problems: given a set of observations

- *How should we model visual objects?* Objects vary in visual appearance due to -for example- their orientation and distance from the camera. A good algorithm for object recognition should be able to generalize from a given set of observations;
- *How can we perform robust recognition?* Objects are located in different environments: they can be partially occluded by other objects in the scene. The presence of other objects can be misleading for the recognition of a specific one. Objects' appearance changes with respect to lighting conditions. For all these reasons, a good algorithm for object recognition should be robust with respect to noise, occlusion, cluttered background and illumination changes.

The present work is devoted to the above issues. We propose to model the visual appearance of objects via probability density functions. We propose a new statistical model and we study its performance for object identification. We also explore its robustness to noise, occlusion and cluttered background. We call this new model Spin Glass-Markov Random Field (SG-MRF). Its main features are:

- A fully connected Markov random field is used for estimating the probability distribution of the model objects. Full connectivity enables to take into account the global appearance of the object, and its specific local characteristics at the same time. Moreover, full connectivity makes it possible to define a neighborhood system for 3D objects in spite of pose variations. Defining a neighborhood system for Markov random field modeling of the appearance of a 3D object was an open problem which made unfeasible using Markov random fields for this task (while many successful examples can be found for 2D object recognition, see [8] and reference therein).

- The energy function, that characterizes the Markov random field to be used, is derived from results of statistical physics of spin glasses. Thus we can benefit of the theoretical knowledge developed by the physics community on those class of energies. Moreover, as they have a parametric form, we can learn the optimal parameters for each model object. This is equivalent, in a fully connected Markov random field, to achieve a globality sum of the significant localities.

- Markov random fields and spin glass energy functions can actually be combined together via nonlinear kernel functions. Kernel functions, and the wide class of algorithms that use kernel functions and thus are described as kernel methods (for instance support vector machines, kernel principal component analysis and many others [19]), have become increasingly popular within the machine learning community in the last years. Several papers have shown the potential usefulness of these algorithms for object recognition. An open challenge for kernel algorithms is the choice of the kernel type (note that, once the kernel type is fixed, it is always possible to select the kernel parameters during the training stage, for instance via cross-validation); the type of kernel chosen determines the metric space where the data are mapped, and consequently the algorithm's performance [19]. Till today, how to choose a kernel type for a given task is a lively research area; in practice, for the vast majority of kernel algorithms (and particularly for those used in object recognition) this choice is largely heuristic. The algorithm presented here has theoretical limitations regarding the kernel type which can be used, thus it eliminates the element of heuristic.

The paper is organized as follows: section 2 presents the state of the art in object recognition. Section 3 describes the derivation of the new probabilistic model and section 4 reports extensive experiments that show the effectiveness and robustness of the new model for appearance-based object recognition. The paper concludes with a summary discussion and possible future directions of research.

## 2 State of the Art

Object recognition is one of the most researched areas of computer vision. Most methods developed so far can be categorized as geometry-based or appearance-based, the main difference being object representation. Here we will focus on appearance-based methods, where objects are modeled by a set of images, and recognition is performed by matching directly the input image to the model set. Swain and Ballard [20] proposed to represent an object by its color histogram. Objects are identified by matching a color histogram from an image region with a color histogram from a sample of the object. The matching is performed using histogram intersection. The method is robust to changes in the orientation, scale, partial occlusion and changes of the viewing position. Its major drawbacks are its sensitivity to lighting conditions, and that many object classes cannot be described only by color. Schiele and Crowley [17] generalized this method by introducing multidimensional receptive field histograms to approximate the probability density function of local appearance. The recognition algorithm calculates probabilities for the presence of objects based on a small number of vectors of local neighborhood operators such as Gaussian derivatives at different scales. The method obtained good object hypotheses from a database of 100 objects using a small number of vectors. Also based on local characteristics, Schmid and Mohr [18] developed a system that can recognize objects in the case of partial visibility, image transformations and complex scenes. The approach is based on the combination of differential invariants computed at key points with a robust voting algorithm and semi local constraints. The recognition is based on the computation of the similarity (represented by the Mahalanobis distance) between two invariant vectors. Matching is performed on discriminant points of an image, and a voting algorithm is used to find the closest model to an image. Principal component analysis has been widely applied for appearance-based object recognition [21, 10, 12, 15]. The attractiveness of the approach is due to the representation of each image by a small number of coefficients, which can be stored and searched efficiently. However, methods from this category have to deal with the sensitivity of the eigenvector representation to changes of individual pixel values, due to translation, scale

changes, image plane rotation or light changes. Several extensions have been investigated in order to handle complete parameterized models of objects [10], to cope with occlusion [15] and to be robust to outliers and noise [7]. Recently, Support Vector Machine (SVM) and kernel methods have gained in interest for appearance based object recognition [13]. Pontil [14] examined the robustness of SVM to noise, bias in the registration and moderate amount of partial occlusions, obtaining good results. Roobaert et al. [16] examined the generalization capability of SVM when just a few number of views per objects are available. Barla [2] proposed to use a new class of kernels, especially designed for vision and inspired by the Hausdorff distance, for 3D object acquisition and detection. A common limitation of SVM and kernel methods proposed so far, is the heuristic in the choice of the kernel function, and in the choice of the kernel parameters; the performance of the algorithm depends heavily on these choices.

# 3 Spin Glass-Markov Random Fields

The probabilistic approach to appearance-based object recognition considers the image views $\boldsymbol{x}$ of a given object $\Omega_k, k = 1, \ldots \mathcal{K}$ as random vectors. Thus, given the set of data samples $\omega_k = \{\boldsymbol{x}_1^k, \boldsymbol{x}_2^k, \ldots, \boldsymbol{x}_{n_k}^k\}$ and assuming they are a sufficient statistic for the pattern class $\Omega_k$, the goal will be to estimate the probability distribution $P_{\Omega_k}(\boldsymbol{x})$ that has generated them. Then, given a test image $\boldsymbol{x}$, the decision step will be achieved using a Maximum A Posteriori (MAP) classifier $k^* = \text{argmax}_{k=1}^{\mathcal{K}} P_{\Omega_k}(\boldsymbol{x}) = \text{argmax}_{k=1}^{\mathcal{K}} P(\Omega_k|\boldsymbol{x})$, and, using Bayes rule,

$$k^* = \underset{k=1}{\overset{\mathcal{K}}{\text{argmax}}} P(\boldsymbol{x}|\Omega_k)P(\Omega_k). \tag{1}$$

where $P(\boldsymbol{x}|\Omega_k)$ are the Likelihood Functions (LFs) and $P(\Omega_k)$ are the prior probabilities of the classes. In the rest of the paper we will assume that the priors $P(\Omega_k)$ are constant and the same for all object classes; thus the Bayes classifier (1) simplifies to

$$k^* = \underset{k=1}{\overset{\mathcal{K}}{\text{argmax}}} P(\boldsymbol{x}|\Omega_k). \tag{2}$$

Probabilistic methods are philosophically optimal in the sense that with a posterior probability distribution over classes, selecting a maximum probability class will minimize the probability of error (see [4] and references therein). A major problem in these approaches is that the functional form of the probability distribution of an object class $\Omega_k$ is not known a priori. Assumptions have to be made regarding the parametric form of the probability distribution, and parameters have to be learned in order to tailor the chosen parametric form to the pattern class represented by the data $\omega_k$.

A possible strategy for modeling the parametric form of the probability function is to use Gibbs distributions within a Markov Random Field framework (MRF, [8]). MRF provides a probabilistic foundation for modeling spatial interactions on lattice systems or, more specifically, on interacting features. It considers each element of the random vector $\boldsymbol{x}$ (that in MRF terminology is called a *confi guration*) as the result of a labeling of all the sites representing $\boldsymbol{x}$, with respect to a given label set:

$$P(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-E(\boldsymbol{x})\right), \qquad Z = \sum_{\{\boldsymbol{x}\}} \exp\left(-E(\boldsymbol{x})\right). \tag{3}$$

The normalizing constant $Z$ is called the partition function, and $E(\boldsymbol{x}) = \sum_i f_i(x_i|x_{N_i})$ is the *energy function*. $P(\boldsymbol{x})$ measures the probability of the occurrence of a particular configuration $\boldsymbol{x}$; the more probable configurations are those with lower energies. Using MRF modeling for appearance-based object recognition, eq. (2) will become

$$k^* = \underset{k=1}{\overset{\mathcal{K}}{\text{argmax}}} P(\boldsymbol{x}|\Omega_k) = \underset{k=1}{\overset{\mathcal{K}}{\text{argmin}}} E(\boldsymbol{x}|\Omega_k). \tag{4}$$

Only a few MRF approaches have been proposed for high level vision problems such as object recognition [5], due to the modeling problem for MRF on irregular sites (for a detailed discussion on this point, we refer the reader to [5]). SG-MRFs overcome this limitation and can be effectively used for appearance-based object recognition [5]. To the best of our knowledge, SG-MRF is the first and only successful MRF-based approach to appearance-based object recognition.

The rest of this section will describe SG-MRFs (section 3.1) and how they can be derived from results of statistical physics of disordered systems (section 3.2). An experimental evaluation of the model will be presented in section 4.

## 3.1 Spin Glass-Markov Random Fields: Model Definition

SG-MRFs are a new class of MRFs that connect SG-like energy functions (mainly the Hopfield one [1]) with Gibbs distributions via a nonlinear kernel mapping. The resulting model overcomes many difficulties related to the design of fully connected MRFs, and enables us to use the power of kernels in a probabilistic framework. Consider $\mathcal{K}$ different object classes $\Omega_1, \Omega_2, \ldots, \Omega_{\mathcal{K}}$, and for each class a set of data samples $\omega_k = \{\boldsymbol{x}_1^k, \boldsymbol{x}_2^k, \ldots, \boldsymbol{x}_{n_k}^k\}, k = 1, \ldots \mathcal{K}$. The SG-MRF probability distribution is given by

$$P(\boldsymbol{x}|\Omega_k) = \frac{1}{Z} \exp\left( \frac{1}{N} \sum_{\mu=1}^{p_k} \left[ K_{d-G}(\boldsymbol{x}, \tilde{\boldsymbol{x}}^{(\mu)}) \right]^2 \right) \tag{5}$$

with $K_{d-G}$ generalized Gaussian kernel

$$K_{d-G} = \exp\{-\rho d_{a,b}(\boldsymbol{x}, \boldsymbol{y})\}, \qquad d_{a,b} = \sum_{i=1}^{m} |x_i^a - y_i^a|^b \tag{6}$$

and prototypes given by the naive ansatz:

$$\{\tilde{\boldsymbol{x}}^{\mu}\}_{\mu=1}^{p_k \equiv n_k} = \{\boldsymbol{x}_1^k, \ldots, \boldsymbol{x}_{n_k}^k\}, \qquad \rho >> \Delta_{min}. \tag{7}$$

Note that SG-MRFs can be defined on features and on raw pixel data. The sites are fully connected, which ends in learning the neighborhood system from the training data instead of choosing it heuristically. Another key characteristic of the model is that in SG-MRF the functional form of the energy is given by construction. The next section will sketches the theoretical derivation of the model. The interested reader will find a more detailed discussion in [5].

## 3.2 Spin Glass-Markov Random Fields: Model Derivation

Consider the following energy function:

$$E = -\sum_{(i,j)=1}^{N} J_{ij} s_i s_j, \tag{8}$$

where the $s_i$ are random variables taking values in $[-1, +1]^N$, $\boldsymbol{s} = (s_1, \ldots, s_N)$ is a configuration and $\boldsymbol{J} = [J_{ij}], (i,j) = 1, \ldots, N$ is the connection matrix, $J_{ij} \in [\pm 1]$. Equation (8) is the most general Spin Glass (SG) energy function [9]; the study of the properties of this energy for different $\boldsymbol{J}$s has been a lively area of research in the statistical physics community for the last 25 years.

An important branch in the research area of statistical physics of SG is represented by the application of this knowledge for modeling brain functions. The simplest and most famous SG model of an associative memory was proposed by Hopfield; it assumes $J_{ij}$ to be given by

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)} \xi_j^{(\mu)}, \tag{9}$$

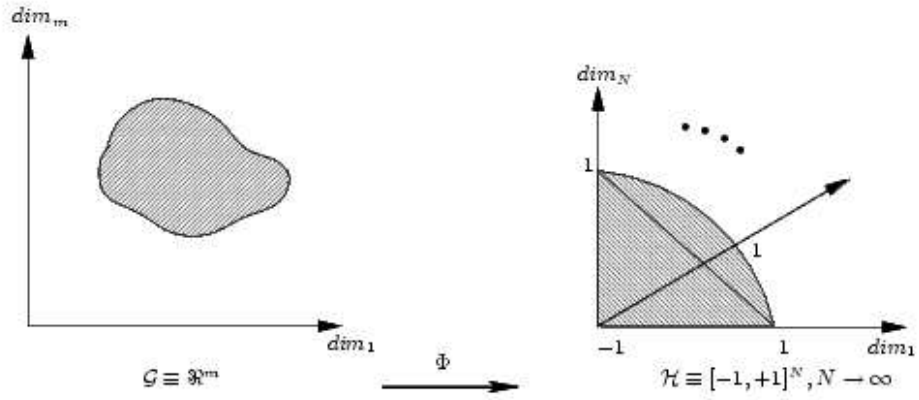


Figure 1: The kernel trick maps the data from a lower dimension space $\mathcal{G} \equiv \Re^m$ to a higher dimension space $\mathcal{H} \equiv [-1,+1]^N, N \to \infty$. This permits to use the Hopfield energy in a MRF framework.

where the $p$ sets of $\{\boldsymbol{\xi}^{(\mu)}\}_{\mu=1}^{p}, \boldsymbol{\xi}^{(\mu)} \in [-1,+1]^N$ are given configurations of the system (that we call *prototypes*) having the following properties:

$$(a) \quad \boldsymbol{\xi}^{(\mu)} \perp \boldsymbol{\xi}^{(\nu)}, \ \forall \mu \neq \nu; \quad (aa) \quad p << N, \ N \to \infty.$$

Under these assumptions it has been proved that the $\{\boldsymbol{\xi}^{(\mu)}\}_{\mu=1}^{p}$ are the absolute minima of the energy (8); for $\alpha > 0.14$ the system loses its storage capability [1]. These results can be extended from the discrete to the continuous case (i.e $\boldsymbol{s} \in [-1,+1]^N$, see [1]); note that this extension is crucial in the construction of the SG-MRF model.

It is interesting to note that the energy (8), with the prescription (9), can be written as:

$$E = -\frac{1}{N} \sum_{(i,j)=1}^{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)} \xi_j^{(\mu)} s_i s_j = -\frac{1}{N} \sum_{\mu=1}^{p} \sum_{i=1}^{N} (\xi_i^{(\mu)} s_i) \sum_j (\xi_j^{(\mu)} s_j) = -\frac{1}{N} \sum_{\mu=1}^{p} (\boldsymbol{\xi}^{(\mu)} \cdot \boldsymbol{s})^2. \tag{10}$$

Equation (10) depends on the data through scalar products, thus it can be *kernelized*, as to say it can be written as

$$E_{KAM} = -\frac{1}{N} \sum_{\mu=1}^{p} [K(\boldsymbol{\xi}^{(\mu)} \cdot \boldsymbol{s})]^2. \tag{11}$$

The idea to substitute a kernel function, representing the scalar product in a higher dimensional space, in algorithms depending on just the scalar products between data is the so called *kernel trick* [19], which was first used for Support Vector Machines (SVM); in the last few years theoretical and experimental results have increased the interest within the machine learning and computer vision community regarding the use of kernel functions in methods for classification, regression, clustering, density estimation and so on. We call the energy given by eq (11) Kernel Associative Memory (KAM). KAM energies are of interest in two different research fields: in the formulation given by equation (11) it is a non linear and higher order generalization of the Hopfield energy function [5]. The other research field is computer vision, on which we focus here. Indeed, we can look at eq (10) as follows:

$$E = -\frac{1}{N} \sum_{\mu=1}^{p} (\boldsymbol{\xi}^{(\mu)} \cdot \boldsymbol{s})^2 = -\frac{1}{N} \sum_{\mu=1}^{p} [\Phi(\boldsymbol{\xi}^{(\mu)}) \cdot \Phi(\boldsymbol{s})]^2 = -\frac{1}{N} \sum_{\mu=1}^{p} [K(\boldsymbol{\xi}^{(\mu)} \cdot \boldsymbol{s})]^2 \tag{12}$$

provided that $\Phi$ is a mapping such that (see Figure 1):

$$\Phi : \mathcal{G} \equiv \Re^m \mapsto \mathcal{H} \equiv [-1, +1]^N, N \to \infty.$$

that in terms of kernel means

$$K(\boldsymbol{x}, \boldsymbol{x}) = 1, \quad \forall \boldsymbol{x} \in \mathcal{G}, \dim(\mathcal{H}) = N, \quad N \to \infty. \tag{13}$$

If we can find such a kernel, then we can use the KAM energy, with all its properties, for MRF modeling. As the energy is fully connected and the minima of the energy are built by construction, using this energy overcomes all the modeling problems relative to irregular sites for MRF [8]. Conditions (13) are satisfied by generalized Gaussian kernels (6). Regarding the choice of prototypes, given a set of $n_k$ training examples $\{\boldsymbol{x}_1^k, \boldsymbol{x}_2^k, \ldots, \boldsymbol{x}_{n_k}^k\}$ for the object class $\Omega_k$, the condition to be satisfied by the $p_k$ prototypes of pattern class $k$ is $\boldsymbol{\xi}^{(\mu)} \perp \boldsymbol{\xi}^{(\nu)} \quad \forall \mu \neq \nu, \quad \mu = 1, \ldots, p_k, \quad p_k << N$ in the mapped space $\mathcal{H}$, that becomes $\Phi(\tilde{\boldsymbol{x}}^{(\mu)}) \perp \Phi(\tilde{\boldsymbol{x}}^{(\nu)}), \quad \forall \mu \neq \nu, \quad \mu = 1, \ldots, p_k, \quad p_k << \dim(\mathcal{H})$ in the data space $\mathcal{G}$.

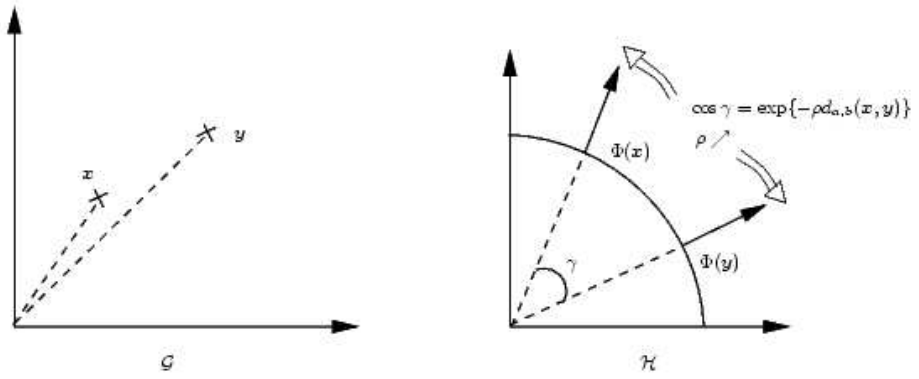


Figure 2: Generalized Gaussian kernels map the data to an infinite dimension hyper-sphere of radius unity. Thus, with a proper choice of $\rho$, it is possible to orthogonalize all the training data in that space.

The measure of the orthogonality of the mapped patterns is the kernel function (6) that, due to the particular properties of Gaussian kernels, has the effect of orthogonalizing the patterns in the space $H$ (see Figure 2). Thus, the orthogonality condition is satisfied by default: if we do not want to introduce further criteria for the choice of prototypes, the natural conclusion is to take all the training samples as prototypes. This approximation is called the *naive ansatz*.

# 4 Experiments

In this section we present experiments that show the effectiveness and robustness of SG-MRF for appearance-based object recognition. Experiments were done on the 100 objects of the Columbia [11] database. The training set consisted of 36 views per object, the testing set of 72. Features were extracted using a Multidimensional receptive Field Histogram (MFH) representation [17], that proved successful for robust object recognition. We used 2D (3D) MFH, with filters given by Gaussian derivatives along $x$ and $y$ directions (and Laplacian filter) and with $\sigma$= 1.0 (3.5);

resolution for histogram axis 16 bins. In the classification step, we compared the performance of SG-MRF with a nearest neighbor classifier with $\chi^2$ distance, which proved to be an effective comparison measurement for the MFH representation [17].

## 4.1 Recognition Results

A first set of experiments was done without any additional degradation. The obtained results testify the effectiveness of SG-MRF and can give a reference frame for evaluating its robustness in experiments reported later. Each training view was represented by a discrete 2D histogram, and we chose generalized kernels with $a = 1, 0.5$ and $b = 2, 1, 0.5$. The obtained recognition rates for $\chi^2$ and SG-MRF are reported in Table 1. For unperturbed images, SG-MRF performs comparably to $\chi^2$, for some kernels even slightly better.

| $\chi^2$ | SG-MRF | | | | | |
|---|---|---|---|---|---|---|
| 99.52 | $a = 1$ | $b = 2$ | 94.43 | $a = 0.5$ | $b = 2$ | 99.58 |
| | | $b = 1$ | 97.84 | | $b = 1$ | 99.80 |
| | | $b = 0.5$ | 99.38 | | $b = 0.5$ | **99.84** |

Table 1: Classification results for $\chi^2$ and SG-MRF

## 4.2 Robustness to Noise

Most of the experiments presented in the literature on the robustness of object recognition systems against noise use ideal views for training and introduce noise in the test set [14, 17]. This only demonstrates the capability of a system to recognize corrupted images but not its ability to generalize from degraded training images. Here we report experiments we ran to investigate this second property.

We performed several series of experiments on the COIL and Columbia databases using 36 prototypes per object in the training set. Test set consisted of the remaining views. In order to explore robustness of SG-MRF to noise, we added independent Gaussian noise with $\sigma_{noise} \in \{10, 50, 80, 120\}$. Note that since the image gray levels are bound to be between 0 and 255, adding Gaussian noise means that the noisy images were actually rescaled within the range $[0, 255]$ [14]. Some examples of noisy views are shown in Figure 3. For each level of noise, we generated four

Figure 3: COIL database, examples of noisy images: from left to right, $\sigma_{noise} \in \{0, 10, 50, 80, 120\}$

training sets containing respectively $25\%, 50\%, 75\%$ and $100\%$ of degraded views with respect to the original training set. For example, a $25\%$ training set - for a given noise level, for a given object - contains 9 noisy and 27 original views. We proceeded in the same way to generate four test sets with the same percentages of degraded images. The views to be corrupted were picked out randomly [1]. For each noise level, for each test set the training was performed

[1] A pilot experiment with a uniform distribution of noisy views led to comparable outcomes. We conclude that the chosen mixture does not affect results.

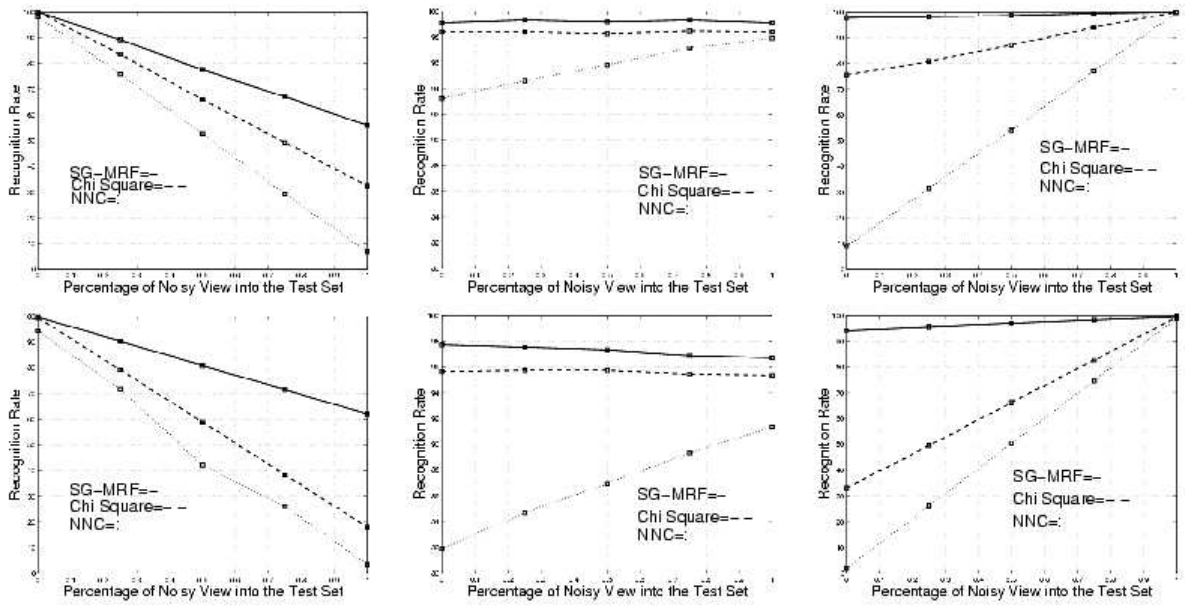


Figure 4: Recognition rates for the COIL (top) and Columbia (bottom) database, for the noise level $\sigma_{noise} = 10$ and 0% (left column), 50% (middle column) and 100% (right column) of degraded views into the training set. Results are reported as a function of the percentage of degraded views in the test set, for SG-MRF, $\chi^2$ and NNC.

respectively on the generated training sets including the ideal one. The views were represented using 2D-MFH. The kernel parameters were $a = 0.5$ and $b = 1$. Recognition time per view was 0.02 sec for the COIL database and 0.23 sec for Columbia database. Due to space constraints we present just the most significant results (the interested reader will find a comprehensive description in [5]).

**Results for** $\sigma_{noise} = 10$ Figure 4 shows the recognition rates obtained when the training set contains 0%, 50% and 100% of noisy views, as a function of increasing number of noisy views in the training set. We see that in the case of 0% noisy views in the training set (Figure 4, left column), for both databases and all classifiers, performance decreases as the percentage of noisy views in the test set increases. Still, SG-MRF performs significantly better than $\chi^2$ and NNC. Figure 4, middle column, shows the recognition rates obtained when the training set contains 50 % of noisy views, as the number of noisy views in the test set increases. Here we see a spectacular change of behavior for SG-MRF and $\chi^2$: for both classifiers, adding noisy training views leads to a remarkable robustness. Although the behavior is similar for the two classifiers, SG-MRF still performs better than $\chi^2$, particularly in the case of a high number of objects. It is interesting to note that NNC doesn't benefit from the degradation into the training set. With respect to the case of 0% noisy views in the training set, we see that recognition rate drops for 0% noisy views in the test set. Then, as the percentage of noisy views in the test set increases, the performance increases as well, in contrast with what happened in the experiment described above. This behavior is more evident for the Columbia database. For any percentage of noisy views in the test set, NNC performs worse than SG-MRF and $\chi^2$. Figure 4, right column, shows recognition performances when the training set consists of all noisy views. With respect to NNC, the behavior is similar to that observed in the previous experiments. We see that $\chi^2$ recognition performance decreases as the test set doesn't contain noisy views. This behavior is very evident for the Columbia database. On

the contrary, SG-MRF maintains its robustness properties for both databases, obtaining the best overall performance. For $\sigma_{noise} = \{50, 80, 120\}$, performances for 0% and 50% of noisy views in the training set are analogous to what reported for $\sigma_{noise} = 10$. In the case of 100 % of noisy views, all classifiers lose robustness.

#### 4.2.1 Robustness to Occlusion

We tested robustness to occlusion of SG-MRF reducing gradually the amount of visible portion of the object; in order to be sure that the windows include always portions of the objects, we took the windows centered to the whole image. The same procedure was used by Schiele in [17] for a similar experiment; it corresponds to the ideal case where the location of the object is known. For these experiments we followed a similar procedure as for the noise experiments to generate four test and training sets that contain an increasing percentage of occluded images. Thus, we reduced gradually the amount of visible portion of the object, while keeping it centered within the image. Let $o$ be the visible object portion; we computed the recognition rates for $o \in \{20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%\}$; Figure 5 shows the resulting object portions. Views were represented using 3D-histograms, with Gaussian derivatives


Figure 5: Examples of the second kind of occlusion; from left to right: visible object portions $o \in \{80\%, 70\%, 60\%, 50\%, 40\%, 30\%, 20\%\}$

$D_x D_y Lap, \sigma = 1.0$, resolution per histogram axis 16 bins. Kernel parameters were $a = 0.5$ and $b = 2$. We restrict here the representation of the results to those achieved with SG-MRF. Figure 6 show classification results of occluded images with regard to the visible portion of the object on the COIL and Columbia databases. We see that adding degraded views into the training set leads to a considerable increase in the recognition stage. We observe also that the higher the number of objects to be recognized, the sooner the performance decreases (with respect to the amount of degradation).

#### 4.2.2 Robustness to Heterogeneous Background

Many experiments presented in the literature on appearance-based object recognition considers objects as located in a homogeneous, contrasting background ([3, 6] and many others). The main motivation beyond this assumption is that, if we want to model the appearance of an object, we need information regarding that object, and that object only. On the other side, in our every day experience we meet objects in several different backgrounds; it is thus essential for every object recognition system to be robust with respect to every kind of background. Although heterogeneous background can be seen as a degradation of a given view just in a loose sense, the results shown in the previous paragraphs induce us to investigate robustness of SG-MRF with respect to heterogeneous background, using the same procedure employed for noise and occlusions. We ran experiments on the COIL and Columbia database, and tested how the robustness of SG-MRF increases when we add in the training set views containing heterogeneous background. In this set of experiments, objects from the COIL and Columbia databases were cut from the original images and put on images with heterogeneous background, having a resolution of $128 \times 128$. We chose 4 different backgrounds shown in Figure 7: one is homogeneous and the others are heterogeneous representing three different scenes. As a referring point, we performed a set of experiments training the system with views on homogeneous background. In a second

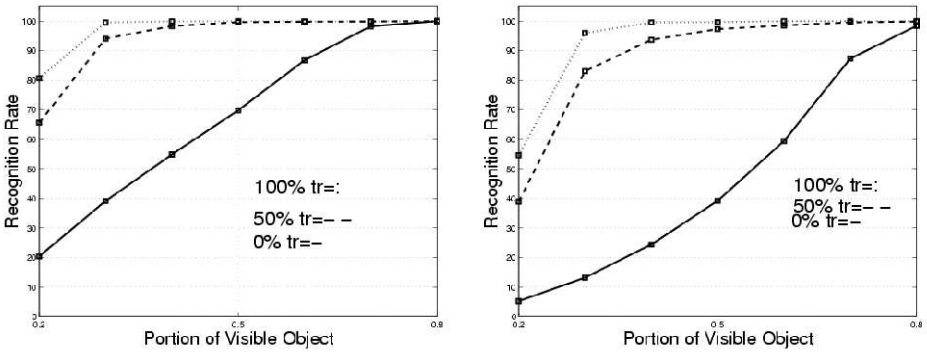


Figure 6: Recognition rates for the COIL (left) and Columbia (right) database as a function of the portion of visible object in the test set with respect to the different percentages of occlusion in the training data. The test set consists of exclusively occluded images.

step, we mixed the training data with views having heterogeneous background: 25% of each background was present in the training set; the distribution of the mixture of backgrounds was uniform. In both experiments, we performed the testing on each background separately. Features were extracted in all these experiments using 4D-MFH, with $D_x D_y$ Gaussian derivatives, $\sigma_1 = 3.5$ and $\sigma_2 = 7.0$; kernel parameters were set to $a = 1, b = 1$. The training set contained 36 prototypes per object and the test set consisted of the remaining views. Classification results achieved with the homogeneous training set are reported in Table 2. The results obtained with the mixed training set are reported in Table 3 for both databases.


Figure 7: Four different backgrounds involved in the localization experiments and their labeling: hbg0 is homogeneous, hbg1,hbg2 and hbg3 represent three different scenes.

The recognition results reported in Table 2 show that when the system is trained on ideal images, there is a dramatic decrease in the recognition rate when the objects are put on heterogeneous background. By mixing the training set with these backgrounds we achieved stability and robustness in the recognition rates (see Table 3) for both databases. This confirms once again the robustness of SG-MRF.

| | hbg0 | hbg1 | hbg2 | hbg3 |
|---|---|---|---|---|
| COIL20 | 100 | 85.00 | 87.36 | 87.29 |
| COIL100 | 100 | 59.00 | 60.75 | 66.65 |

Table 2: Recognition results on the COIL20 and the COIL100 databases, when the train set contains objects with homogeneous background

| | hbg0 | hbg1 | hbg2 | hbg3 |
|---|---|---|---|---|
| COIL20 | 99.93 | 99.93 | 99.93 | 99.93 |
| COIL100 | 99.53 | 99.37 | 99.16 | 99.15 |

Table 3: Recognition results on the COIL20 and the COIL100 databases, when the train set contains objects with a mixture of homogeneous and heterogeneous backgrounds

# 5 Summary

Recognizing objects on the basis of their visual appearance is one of the major goals in computer vision. This paper presented a new probabilistic method for object recognition which allowed us to take into account local and global visual properties of objects. We assumed that objects were represented by their visual appearance, and that, for each object represented by a collection of views, we were able to extract meaningful features. The focus has been on how to develop an effective probabilistic model on the basis of given features, rather than defining new descriptors. We derived a new probabilistic model which we have called Spin Glass-Markov Random Field. This model has been developed on the basis of knowledge and results achieved in computer vision, machine learning and statistical physics of spin glasses [9]. This results in a probability distribution estimated via a fully connected Markov random field. Full connectivity allows the model to take into account at the same time global and local characteristics of the visual pattern under consideration; hence the novelty of the approach.

A key feature of any object recognition algorithm aiming to perform well in realistic scenarios is robustness. We performed an extensive series of experiments to test robustness of Spin Glass-Markov Random Fields with respect to noise, occlusion and heterogeneous background. Experimental results have shown that Spin Glass-Markov Random Fields are highly robust to noise, occlusion and decreasing number of training views. In conclusion, we can state that Spin Glass-Markov Random Fields are a promising new probabilistic model for robust recognition of objects in real world scenes, and constitutes a valid alternative to state of the art algorithms for visual pattern recognition.

This work can be developed in many ways. Some important theoretical points should be further investigated: first of all, the choice of prototypes. Selecting prototypes from the training set is equivalent to select key views for representing the considered object. It is important to study how the naive ansatz is related to existing work on this topic, and more generally how Spin Glass-Markov Random Fields can benefit from results achieved in this research area. Second, it would be very important to develop fast algorithms for selecting the kernel parameters during the learning phase. Finally, the algorithm should be extended so to tackle object localization and scale variations. Future work will concentrate in these directions.

# References

[1] D. Amit. *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press., 1989.

[2] A. Barla, F. Odone, A. Verri. "Hausdorff Kernel for 3D Object Acquisition and Detection", *Proc. of European Conference of Computer Vision (ECCV02)*, 2002, pp 20-33.

[3] S. Belongie, J. Malik, J. Puzicha, "Matching shapes", *Proc of International Conference on Computer Vision (ICCV01)*, Vancouver, 2001, pp 454-461.

[4] C. M. Bishop, *"Neural Networks for Pattern Recognition"*, Claredon Press, Oxford, 1995.

[5] B. Caputo, S. Bouattour, and H. Niemann, "Robust appearance-based object recognition using a fully connected markov random fields", R. Kasturi, D. Laurendean, C. Suen Editors, *Proc of International Conference on Pattern Recognition (ICPR02)*, Quebec City, 2002, Vol 3, pp 565-568.

[6] J. Hornegger and H. Niemann, "Statistical learning, localization and identification of objects", *International Conference on Computer Vision (ICCV95)*, Cambridge, Massachusset, 1995, pp 914-919.

[7] A. Leonardis and H. Bischof, "Robust recognition using eigenimages", *Computer Vision and Image Understanding*, 78(1), 2000, pp 99-118.

[8] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer, Tokyo, 1995.

[9] M. Mezard, G. Parisi, M. Virasoro, *" Spin Glass Theory and Beyond"*, World Scientific, Singapore, 1987.

[10] H. Murase and S. Nayar, "Visual learning and recognition of 3-d objects from appearance", *International Journal of Computer Vision*, 14(1), 1995, pp 5-24.

[11] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-100)", Technical Report No. CUCS–006–96, Dept. Comp. Science, Columbia University, 1996.

[12] K. Ohba, K. Ikeuchi, "Recognition of the multi specularity objects for binpicking task", *International Conference on Intelligent Robots and Systems (IROS96)*, Osaka, Vol 3, pages 1440-1448, 1996.

[13] E. Osuna, R. Freund, F. Girosi, "Training support vector machines: An application to face detection", *International Conference of Computer Vision and Pattern Recognition (CVPR97)*, Puerto Rico, 1997, pp 130-136.

[14] M. Pontil and A. Verri, "Support vector machines for 3-d object recognition", *IEEE Transanction on Pattern Aanalysis and Machine Intelligence*, 20, 1998, pp 637-646.

[15] R. Rao and D. Ballard, "An active vision architecture based on iconic representations", *Artifi cial Intelligence*, 78, 1995, pp 461-505.

[16] D. Roobaert, M. Zillich, J. O. Eklundh, "A pure learning approach to background- invariant object recognition using pedagogical support vector learning", *Proc of International Conference on Computer Vision and Pattern Recognition*, Kauwai, 2001, pp 351-357.

[17] B. Schiele, J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms", *International Journal of Computer Vision*, 36(1), 2000, pp 31-52.

[18] C. Schmid, R. Mohr, "Combining grayvalue invariants with local constraints for object recognition", *International Conference of Computer Vision and Pattern Recognition*, San Francisco, pp 872-877, 1996.

[19] B. Schölkopf, A. J. Smola, *Learning with kernels*, MIT Press, 2002.

[20] M. Swain and D. Ballard, "Color indexing", *International Journal of Computer Vision*, 7(1), 1991, pp 11-32.

[21] M. Turk, A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, 3(1), 1991, pp 71-86.