A Novel Probabilistic Model for 3D Object Recognition: Spin-Glass Markov Random Fields

B. Caputo, S. Bouattour, D. Paulus

Computer Science Department, Chair for Pattern Recognition,
University of Erlangen-Nuremberg,
Martensstrasse 3, D-91058, Erlangen, Germany
Phone +49 9131 8527824; fax: +49 9131 303811
E-Mail: {caputo, sabouatt, paulus}@informatik.uni-erlangen.de

Abstract

This contribution presents a new class of MRF, that is inspired by methods of statistical physics. The new energy function assumes full-connectivity in the neighborhood system and thus solves the modeling problems of MRF in the case of irregular sites. Moreover, it is defined independently from the considered application; this makes it possible to avoid the use of a search algorithm for the energy minima, since those and their analytical properties are provided by theory. Experiments were performed on a database of 12 Objects using multidimensional receptive field histograms. We achieved a recognition rate of 98.33%.

1 Introduction

Object recognition is one of the most researched area of computer vision; many papers have proposed to tackle this problem using several frameworks for the statistical representation of the object (see for instance [15], [10]). This contribution describes a new model that allows the use of Spin--Glass Theory (SGT, [12]) results in a Maximum A Posteriori-Markov Random Field (MAP-MRF, [10]) framework for 3-D object recognition. Many vision problems can be posed as labeling problems; labeling is also a natural representation for the study of MRFs [10]. Two major tasks when modeling MRFs are how to define the neighborhood system for irregular sites, and how to choose the energy function for a proper encoding of constraints. The neighbor relations between sites is related to their regularity; in the irregular case [10], the neighborhood system is mostly defined by means of a heuristic distance that is feature-dependent. If the application problem is object recognition, we have additional problems: if the chosen features are not invariant to pose, we should incorporate the pose parameters into the energy formulation and in the neighbor relations definition, possibly with a dramatic increase in complexity; moreover, due to mutual occlusion, neighborhoods change with pose parameters. The energy function is a quantitative cost measure of the quality of a solution, where the best solution is the minimum. In the case of irregular sites, the energy function's formulation can become something of an art, as it is generally done manually. The problem of the neighborhood definition can be avoided in a fully connected MRF: full connectivity eliminates the need to define distances between sites, but it does not solve the problem of increased complexity; on the contrary, it increases it.

SGT provides sophisticated techniques and knowledge which can be used to deal with MRFs modeling problems in an elegant manner: full connectivity makes the neighborhood definition irrelevant, and the energy function is defined independently of the considered application; this makes it possible to find the analytical properties of the minima and may make it unnecessary to construct fast algorithms for searching for absolute minima, thus avoiding the explosion of the search space. A basic assumption in SGT is the infinite dimension of the configuration space where the energy lives [12]. This condition cannot be satisfied for a generic pattern recognition problem, due to the curse of dimensionality [2]. The choice of a particular energy, which is a function of the scalar product between configurations [7, 1], allows us to use a kernel function [16] in the energy formulation; this solves the problem of the high dimensionality and makes it possible to use SGT results for MRF modeling purposes: we call this new model Spin Glass-Markov Random Field (SG-MRF). SG-MRF are independent from the neighborhood system definition, due to the full connectivity of the sites; this property allows us to use this new model for probabilistic object recognition. To the best of our knowledge, this model is the first attempt to connect SGT results in a MRF framework.

The paper is organized as follows: Section 2 discusses related work, Section 3 and 4 review MRF and SGT; the new model is presented in Section 5. Experiments are reported in Section 6; the paper concludes with a summary discussion of the presented model.

2 Related Work

Statistical approaches and especially Markov Random Fields are widely used and extensively studied in image processing and computer vision. A detailed discussion of MRF applications, for instance, can be found in the monographs [20, 10]. It is important to note that approaches which use MRF based modeling solve mostly low level image processing problems [21, 20]. Only a few authors consider the high level vision task of object recognition using MRF [19, 11, 13]. And there is still the open question of defining an appropriate neighborhood system and energy function automatically; also the issue of self occlusion is not solved yet. The first approach to generate probabilistic models based on MRF automatically from observations was published in [21]. The proposed MRF model and the resulting recognition system allows the robust recognition of 3D objects in 2D images; the novelty of our contribution is that many problems which usually appear in the context of MRF based image processing and computer vision are avoided by using SGT.

3 Markov Random Fields

Markov Random Field (MRF) [20, 10] is a branch of probability theory which provides a foundation for modeling spatial interactions on lattice systems or, more generally, of interacting features. Labeling is a natural representation for the study of MRFs; furthermore, many vision problems can be posed as labeling problems in which the solution to is a set of labels assigned to image pixels or features.

Let $\mathcal{S} = \{1, \ldots, m\}$ index a discrete set of m sites, and let $\mathcal{L}$ be a set of continuous ($\mathcal{L}_c = \Re^{a \times b \times \cdots}$, $(a, b, \ldots)$ dimensions) or discrete ($\mathcal{L}_d = \{1, \ldots, M\}$, M number of labels) labels; the labeling problem is to assign a label from the label set $\mathcal{L}$ to each of the sites in $\mathcal{S}$. The set $f = \{f_1, \ldots, f_m\}$ is called a labeling of the sites in $\mathcal{S}$ in terms of the labels in $\mathcal{L}$. In the terminology of random fields, a labeling is called a configuration. The sites in $\mathcal{S}$ are related to one another via a neighborhood system. A neighborhood system for $\mathcal{S}$ is defined as

$$\mathcal{N} = \{\mathcal{N}_i | \forall i \in \mathcal{S}\} : i \notin \mathcal{N}_i, i \in \mathcal{N}_t \Longleftrightarrow t \in \mathcal{N}_i \quad (1)$$

where $\mathcal{N}_i$ is the set of sites neighboring i. Let $\boldsymbol{F} = \{F_1, \ldots, F_m\}$ be a family of random variables defined on the set $\mathcal{S}$. For a discrete label set $\mathcal{L}$, the probability that random variable F_i takes the value f_i is denoted $P(F_i = f_i)$, and the joint probability is denoted $P(\boldsymbol{F} = \boldsymbol{f}) = P(F_1 = f_1, \ldots, F_m = f_m)$. $\boldsymbol{F}$ is defined as a MRF on $\mathcal{S}$ with respect to a neighborhood system $\mathcal{N}$ if

$$P(f_i | f_{\mathcal{S}-\{i\}}) = P(f_i | f_{\mathcal{N}_i}),$$

where $\mathcal{S} - \{i\}$ is the set difference, $f_{\mathcal{S}-\{i\}}$ denotes the set of labels at the sites in $\mathcal{S} - \{i\}$ and $f_{\mathcal{N}_i} = \{f_{i'} | i' \in \mathcal{N}_i\}$stands for the set of labels at the sites neighboring i. Note that every random field is a MRF when all different sites are neighbors.

A set of random variables $\boldsymbol{F}$ is said to be a Gibbs Random Field (GRF) on $\mathcal{S}$ with respect to $\mathcal{N}$ if its configurations obeys a Gibbs distribution:

$$P(\boldsymbol{f}) = \frac{1}{Z} \exp\left(-\frac{1}{T} E(\boldsymbol{f})\right), \quad (2)$$

$$Z = \sum_{\{\boldsymbol{f}\}} \exp\left(-\frac{1}{T} E(\boldsymbol{f})\right).$$

The normalizing constant Z is called partition function, T is a constant called temperature, and $E(\boldsymbol{f})$ is the *energy function.* The Hammersley-Clifford theorem establishes the equivalence between MRF and the Gibbs distribution ([20, 10]): for a given $\mathcal{N}_i$, $P(\boldsymbol{f})$ is a MRF distribution if and only if $P(\boldsymbol{f})$ is a Gibbs distribution.

4 Spin Glass Theory and Beyond

As already mentioned in Section 1, SGT was born to describe magnetic materials in which the interactions between the spins are random and conflicting [12]. The attempt to understand the cooperative physics of such systems has led to the development of concepts and techniques which have been finding applications and extensions in many areas such as attractor neural networks [1], combinatorial optimization problems, prebiotic evolution, and so on [12]. Two basic properties of SG are *quenched disorder* and *frustration*: these features are readily visualized in the following energy function:

$$E = -\sum_{(i,j)} J_{ij}\, s_i\, s_j \qquad i,j = 1,\dots N, \tag{3}$$

where the s_i are random variables taking values in $\{\pm 1\}$, $\boldsymbol{s} = (s_1,\dots,s_N)$ is a configuration and $\boldsymbol{J} = [J_{ij}\ ], (i,j) = 1,\dots,N$ is the connection matrix, $J_{ij} \in \{\pm 1\}$. The problem of finding, for a given $\boldsymbol{J}$, the set s_i that minimizes (3) is an NP complete problem [9]. SGT approach makes use of the tools of equilibrium statistical mechanics, finding out the analytical properties of the energy's absolute minima; it is obvious that such knowledge may makes the algorithm unnecessary. The results of these investigations are that frustrated interactions lead to a low temperature energy surface with many valleys [12]. This suggests that similar energy functions can be used for pattern recognition purposes: the valleys will be configuration states related to the classes to be recognized; these valleys should have reasonably large basins of attraction, subject to the constraints imposed by having many classes. In addition, these properties should be extended to configuration spaces of finite dimension in order to avoid the curse of dimensionality problem.

The simplest model in which many different attractors do exist has been proposed by Hopfield [7]. Assuming that the system is fully connected, and that the effect of the interactions from the $N-1$ sites on the i-th is given by

$$h_i = \sum_{j=1}^{N} J_{ij} s_j, \tag{4}$$

and that $J_{ij} = J_{ji}$, the stable states of the system will be those configurations which are the minima of the energy function (3) [1]. Following Hopfield's proposal [7], we choose J_{ij} to be given by

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)}\, \xi_j^{(\mu)}\ , \tag{5}$$

where the p sets of $\{\boldsymbol{\xi}^{(\mu)}\ \}_{\mu=1}^{p}$ are certain particular configurations of the system (that we call *prototypes*) having the following properties:

- *a)* $\boldsymbol{\xi}^{(\mu)} \perp \boldsymbol{\xi}^{(\nu)} \qquad \forall \mu \neq \nu$;
- *aa)* $p = \alpha N, \qquad \alpha \leq 0.14, \qquad N \to \infty$.

Under these assumptions it has been proved that ([1], chapter 4-6) the $\{\boldsymbol{\xi}^{(\mu)}\ \}_{\mu=1}^{p}$ are the absolute minima of E; for $\alpha > 0.14$ there is a phase transition [1], the $\{\boldsymbol{\xi}^{(\mu)}\ \}_{\mu=1}^{p}$ are no more the absolute minima and the system looses its capability to store certain particular configurations. These results can be extended from the discrete to the continuous case (i.e. $\boldsymbol{s} \in [-1,+1]^N$, see [8, 6]).

5 Spin-Glass Markov Random Fields

The same analogies which led to use SGT for modeling brain functions make attractive the idea to use SG-like energy functions in a MRF framework for pattern recognition purposes; this time the valleys would be interpreted as configuration states related to the different classes to be recognized. Moreover, the use of eq. (3)-(5) would solve the modeling problem for MRF illustrated in Section 3 for irregular sites and energy choice: full connectivity would make the neighborhood definition irrelevant, and the energy function would be defined independently of the considered application. The detailed analytical knowledge of the energy function should also make it possible to avoid the NP complete problem (see Section 4). Consider a given pattern recognition problem: let $\mathcal{S} = \{1,\dots,m\}$ be a set of m sites, and $\mathcal{L}_x = [x_l, x_h] \subset \Re$ a continuous label set; then $\boldsymbol{f} = \{f_1,\dots,f_m\}$ will be a labeling configuration in the configuration space G; let there also be K different classes $\Omega_\kappa, \kappa = \{1,\dots,K\}$. Given a configuration $\hat{\boldsymbol{f}}$, our goal is to classify $\hat{\boldsymbol{f}}$ as a sample from Ω_{κ^*}, one of the Ω_κ classes. For example, $\mathcal{S}$ could be viewed as the pixels in an image, $\mathcal{L}_x$ the gray-intensity value, and the considered problem could be that of classifying n classes of objects from single views: in such an example, a particular configuration will be one of the possible images. Using a

Maximum A Posteriori (MAP) criteria we have

$$\kappa^* = \underset{\kappa}{\mathrm{argmax}} P(\Omega_\kappa|\hat{\boldsymbol{f}}\,),$$

and using Bayes rule, the MAP classifier can be rewritten as

$$\kappa^* = \underset{\kappa}{\mathrm{argmax}}\{P(\hat{\boldsymbol{f}}\,|\Omega_\kappa)p(\Omega_\kappa)\}.$$

If we assume that all classes are equiprobable, the a priori probability will become $P(\Omega_\kappa) = 1/K$; the likelihood $P(\hat{\boldsymbol{f}}\,|\Omega_\kappa)$ can be evaluated using MRF modeling:

$$P(\hat{\boldsymbol{f}}\,|\Omega_\kappa) \propto \exp\{-E(\hat{\boldsymbol{f}}\,|\Omega_\kappa)\}$$

Thus, the MAP classifier becomes a ML classifier

$$\kappa^* = \underset{\kappa}{\mathrm{argmin}}\{E(\hat{\boldsymbol{f}}\,|\Omega_\kappa)\}. \tag{6}$$

If we want to model $E(\hat{\boldsymbol{f}}\,|\Omega_\kappa)$ with eq. (3)-(5) in the configuration space G, we must first be able to determine, for each class Ω_κ, a set of prototypes $\{\boldsymbol{\phi}^{(\mu)}\}_{\mu=1}^{\mu_\kappa}, \sum_{\kappa=1}^{K}\mu_\kappa = p$; then, it must hold :

- *i)* $\boldsymbol{\phi}^{(\mu)} \perp \boldsymbol{\phi}^{(\nu)}, \quad \forall \mu,\nu = 1,\ldots,p, \quad \mu \neq \nu$;
- *ii)* $\boldsymbol{f} \in [-1,+1]^N, N \to \infty$.

The orthogonality condition can be relaxed to linear independence between the prototypes; orthogonality will be then obtained with a change of basis. This implies that the relevant information is not contained in the norm, but in the normalized configuration vectors; this would also satisfy the second condition, although with severe limitations on the kind of features and of pattern recognition problems that could be solved with such an approach. The third condition implies that we should have a large number of features (approaching infinity); however, due to the curse of dimensionality [2], this condition cannot be satisfied. These considerations suggest that it is generally not possible to use the energy function (3)-(5) in the configuration space G. Thus we find ourselves in a sort of dichotomic situation: if we want to work on real-life applications we need a finite dimension data space; if we want to use SG--like energy function we need an infinite dimension space. The solution to this dilemma is to actually take two different spaces, one for the data and one for the energy, and to go from one space to another with a non-linear mapping. The data space G will be determined by the chosen features for the particular application under consideration; without any loss of generality we can assume $G \equiv \Re^m$. The energy space is determined by SGT requirements and is given by $H \equiv [-1,+1]^N, N \to \infty$.

If we were able to find a mapping

$$\Phi : \Re^m \to [-1,+1]^N, N \to \infty$$

we could use energy (3)-(5) for MRF modeling purposes: condition *ii)* would be automatically satisfied, and condition *i)* would become

- *i)* $\Phi(\boldsymbol{\phi}^{(\mu)}) \perp \Phi(\boldsymbol{\phi}^{(\nu)}), \forall \mu,\nu = 1,\ldots,p, \mu \neq \nu$;

As a matter of fact, we don't need to find the mapping Φ; things are much easier. First, notice that the energy function (3), due to the choice of the connection matrix (5), can be rewritten as a function of the scalar product between two configuration states:

$$E = -\frac{1}{N}\sum_{i,j}\sum_{\mu}\xi_i^{(\mu)}\xi_j^{(\mu)}s_i s_j =$$

$$= -\frac{1}{N}\sum_{\mu}\sum_{i}(\xi_i^{(\mu)}s_i)\sum_{j}(\xi_j^{(\mu)}s_j) =$$

$$E = -\frac{1}{N}\sum_{\mu}(\boldsymbol{\xi}^{(\mu)}\cdot\boldsymbol{s})^2. \tag{7}$$

Equation (7) depends on the data through scalar products in the space H, that is, on functions of the form $\Phi(\boldsymbol{f}_1)\cdot\Phi(\boldsymbol{f}_2)$. If we can find a *kernel function* K such that

$$K(\boldsymbol{f}_1,\boldsymbol{f}_2) = \Phi(\boldsymbol{f}_1)\cdot\Phi(\boldsymbol{f}_2), \tag{8}$$

which satisfies the conditions

- *j)* $K(\boldsymbol{f},\boldsymbol{f}) = 1, \qquad \forall \boldsymbol{f} \in G$;
- *jj)* $\dim(H) = N, N \to \infty$,

we could substitute equation (8) in equation (7) and use $K(\cdot,\cdot)$ instead of Φ. A theoretical result of Mercer [16, 18] actually allows us to find the kernel function K satisfying equation (8) without explicitly know the mapping Φ. Mercer's condition tells us for which kernels there exists a pair $\{H,\Phi\}$ with the properties described above: there exist a mapping Φ and an expansion

$$K(x,y) = \sum_i \Phi(x)_i\Phi(y)_i$$

if and only if, for any $g(x)$ such that

$$\int g(x)^2 dx$$

is finite, then

$$\int K(x,y)g(x)g(y)dxdy \geq 0$$

The idea to substitute a kernel function, representing the scalar product in a higher dimension space, in algorithms depending just from the scalar products between data is the so called *kernel trick* [18, 16] 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 [16]. This is exactly what we do here: the kernel trick thus allow to use SGT results in a MRF-MAP framework: the energy function (7) becomes

$$E_\kappa = -\frac{1}{N}\sum_{\mu=1}^{\mu_\kappa} K(\boldsymbol{f},\boldsymbol{\phi}_\kappa^{(\mu)})^2, \qquad (9)$$

where to each class Ω_κ will be associated a subset of prototypes $\{\boldsymbol{\phi}_\kappa^{(\mu)}\}_{\mu=1}^{\mu_\kappa}, \sum_{\kappa=1}^{K}\mu_\kappa = p$. The MRF-MAP classifier (6) will become:

$$\kappa^* = \operatorname*{argmin}_\kappa \; -\frac{1}{N}\sum_{\mu=1}^{\mu_\kappa} K(\boldsymbol{f},\boldsymbol{\phi}_\kappa^{(\mu)})^2 \; . \qquad (10)$$

In this sense, SG-MRF can be seen as a new kernel method for probability density estimation and classification. It is important to note that, using the kernel trick, the conditions to be satisfied are *a), aa)* and not *i), ii)*: this means that the energy is defined in a space H rather than the space G where the data lives. Regarding the condition of orthogonality between the $\{\boldsymbol{\phi}^{(\mu)}\}_{\mu=1}^{p}$, from the signal to noise analysis on the stability of the stored prototypes [1] it turns out that, if p is fixed, the condition $N \to \infty$ dominates the noise term (see [1]). Thus, the stability of the prototypes is guaranteed if the chosen kernel satisfies the condition *jj)*, as the kernel function is the scalar product in the space H. Another way to see it is to say that with the mapping we "orthogonalize" the prototypes. We want to stress that the kernelization of the energy function (3) is due to the fact that $\boldsymbol{J}$ is given by the (5); indeed it is this choice that allows us to write the energy as a function of the scalar product. Thus SG-MRF can be seen as a kernel associative memory.

Many algorithms which make use of the kernel trick do not provide criteria in order to choose the kernel type, in spite of the fact that the choice of a certain kernel instead of another may lead to a poor performance of the algorithm; this is the case for example of SVM [5, 16]. On the contrary, SG-MRF kernel's choice must satisfy criteria *(j, (jj*. These conditions are satisfied by the Gaussian Radial Basis Function kernel (G-RBF) [18, 16]:

$$K_{G-RBF}(\boldsymbol{x},\boldsymbol{y}) = \exp\{-\rho||\boldsymbol{x}-\boldsymbol{y}||^2\} \qquad (11)$$

which is proved to satisfy Mercer's condition. This kernel can be seen as a particular case of a generalized Gaussian kernel [5]:

$$K_{d-RBF}(\boldsymbol{x},\boldsymbol{y}) = \exp\{-\rho d(\boldsymbol{x},\boldsymbol{y})\} \qquad (12)$$

where $d(\boldsymbol{x},\boldsymbol{y})$ can be chosen to be any distance in the input space. With this kernel's formulation it is possible to define a generalized distance measure $d_{a,b}$ [5]:

$$d_{a,b}(\boldsymbol{x},\boldsymbol{y}) = \sum_i |x_i^a - y_i^a|^b \qquad (13)$$

For $a = 1$ and $b = 1,2$, equation (13) becomes an L_1 ($|x_i - y_i|$) and L_2 ($|x_i - y_i|^2$) distance measure, respectively. It is demonstrated that equation (13) satisfies Mercer's condition if and only if $0 \le b \le 2$ [18]; the exponentiation of x_i by a does not affect the validity of the Mercer's condition, as it can be seen as a remapping of the input variables. Kernel (12), with generalized distance (13) satisfies also conditions *(j, (jj* [18]. The generalized kernels (12)-(13) have been used for color histograms image based classification, proving to be very effective [5].

Regarding the choice of prototypes, let us suppose to be given a set of m training examples $\{\boldsymbol{f}_{\kappa 1},\boldsymbol{f}_{\kappa 2},\dots,\boldsymbol{f}_{\kappa m}\}$ relative to class Ω_κ, the condition to be satisfied by the prototypes is

$$\boldsymbol{\xi}^{(\mu)} \perp \boldsymbol{\xi}^{(\nu)} \qquad \forall \mu \neq \nu$$

in the mapped space H, that becomes

$$\Phi(\boldsymbol{f}_\kappa^{(\mu)}) \perp \Phi(\boldsymbol{f}_\kappa^{(\nu)}), \qquad \forall \mu \neq \nu \qquad (14)$$

in the representation space G. The measure of the orthogonality of the mapped patterns is the kernel function (8) that, due to the particular properties of Gaussian Kernels *j), jj)*, has the effect of orthogonalize the patterns in the space H. Thus, the condition (14) does not really give us any constraint as it is satisfied by default: if we don't want to introduce further criteria for the choice of prototypes, the natural conclusion is to take all the training samples as prototypes:

$$\{\boldsymbol{f}_{\kappa 1},\boldsymbol{f}_{\kappa 2},\dots,\boldsymbol{f}_{\kappa m}\} = \{\boldsymbol{\phi}^{(\mu)}\}_{\mu=1}^{\mu_\kappa}, \sum_{\kappa=1}^{K}\mu_\kappa = p$$

In this case the energy function will become

$$E = \sum_{\kappa} E_{\kappa}(\boldsymbol{f}) = \sum_{\mu=1}^{p} (K(\boldsymbol{\phi}^{(\mu)}, \boldsymbol{f}))^2, \qquad (15)$$

$$E_{\kappa}(\boldsymbol{f}) = \sum_{\mu=1}^{\mu_\kappa} (K(\boldsymbol{f}_{\kappa}^{(\mu)}, \boldsymbol{f}))^2, \qquad (16)$$

where $E_{\kappa}(\boldsymbol{f})$ represents the contribution to the energy given by the prototypes relative to class Ω_{κ}.

The choice of taking all the training patterns as prototypes can become a problem when a large number of training examples is given for every class. Indeed, it is possible that some of the training patterns are very similar to each other; in such a case the orthogonalization of patterns made by the kernel can be insufficient to guarantee the stability of the solutions: in other words, if we take all the training examples as prototypes we risk to store prototypes for which hold $\boldsymbol{\phi}^{(\mu_i)} \simeq \boldsymbol{\phi}^{(\mu_j)}$; this risk will be higher as the number of training samples increases.

6 Experiments

In this section we present experiments which show the effectiveness of SG-MRF for 3D object recognition applications. To this purpose we tested the proposed approach on a subset of the DIROKOL database [14], which is composed by 13 objects that can be found in office or hospital environments. In the original database each object is represented by 3720 views, uniformly distributed over a hemisphere, with a sampling rate of 3°. Here we used a subset of the original database, given by 12 objects (see Figure 1), 240 views per object (corresponding to a sampling rate of 12°); each view was a gray level image with homogeneous background and resolution of 256×256. Half of them were used as training set, the other half as test set.

Features were extracted using a Multidimensional receptive Field Histogram (MFH) representation [15][1]. This approach was developed by Schiele to extend the color histogram method of Swain and Ballard [17]. The underlying idea is to compute for each image pixel a multidimensional vector of receptive fields and calculate the histograms of these responses. A MFH is determined once we chose the local property measurements (i.e., the receptive field functions), which determine the dimension of the histogram, and the resolution of each axis. On the basis of the results reported in [15], we chose three local characteristics based on Gaussian derivatives:

$$D_x = -\frac{x}{\sigma^2} G(x,y), \qquad D_y = -\frac{y}{\sigma^2} G(x,y)$$

$$Lap = G_{xx}(x,y) + G_{yy}(x,y),$$

where

$$G(x,y) = \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

is the Gaussian distribution and

$$G_{xx}(x,y) = \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right) G(x,y),$$
$$G_{yy}(x,y) = \left(\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2}\right) G(x,y)$$

are the second order derivatives with respect to x and y. Each view was represented by three different combinations of local characteristics: D_x-D_y with $\sigma = 1.0$, that we call 2D-histogram, D_x-D_y--Lap with $\sigma = 1.0$, called 3D-histogram and D_x--D_y-Lap with two different scales $\sigma_1 = \sigma, \sigma_2 = 2\sigma$, called 6D-histogram. All the obtained representations are discrete histograms with a resolution of 16 bins per axis.

For the classification task, we used SG-MRF in the MAP-MRF framework described in Section 5. The number of prototypes corresponds to the number of complete training set. As kernels we chose generalized Gaussian kernels with $a = 1, 0.5$ and $b = 2, 1, 0.5$. We run the experiments with $\rho \in [10^{-5}, 10^{-10}]$, which can be considered as a significant window, since it delivers stable recognition rates. The obtained results for the six chosen Gaussian kernels and the three different histogram representations are summarized in Table 1, Table 2 and Table 3.

$10^{-5} \leq \rho \leq 10^{-10}$	$b = 2$	$b = 1$	$b = 0.5$
$a = 1$	91.59	97.84	49.65
$a = 0.5$	98.05	71.59	22.36

Table 1: Recognition rates for 6D histograms

[1] We gratefully thank B. Schiele which allowed us to use his software fro the computation of MFH.

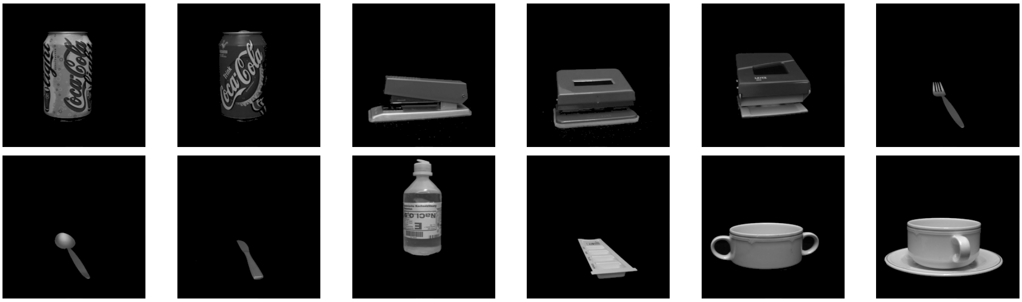

Figure 1: Objects from the DIROKOL database.

$10^{-5} \leq \rho \leq 10^{-10}$	$b = 2$	$b = 1$	$b = 0.5$
$a = 1$	94.09	97.98	96.87
$a = 0.5$	98.33	97.43	92.29

Table 2: Recognition rates for 3D histograms

$10^{-5} \leq \rho \leq 10^{-10}$	$b = 2$	$b = 1$	$b = 0.5$
$a = 1$	87.36	92.98	96.04
$a = 0.5$	95.55	97.36	96.04

Table 3: Recognition rates for 2D histograms

6.1 Discussion

Results reported in Tables 1, 2 and 3 show that for all the performed experiments there is at least one kernel with which is obtained a recognition rate $> 97\%$; actually, for $a = 0.5$ and $b = 2$ the obtained recognition rates for 3D and 6D histograms are $> 98\%$. These results show the effectiveness of the proposed method and are in agreement with results obtained with the same approach on different databases [3] and with appearance based representation [4]. According to theoretical properties of generalized Gaussian kernels [5], SG-MRF's performance is expected to improve when the product $ab \rightarrow 0$. This behavior can be explained as follows: suppose that a π-pixel bin in the histogram represents a single uniform grey level region in the image represented by histogram $\boldsymbol{H}_1$. A small variation of intensity value in that region can have the effect to move the π-pixel in a neighboring bin; the result will be a different histogram $\boldsymbol{H}_2$. Assuming for simplicity that the neighboring bin was empty in $\boldsymbol{H}_1$, we have [5, 3]:

$$K_{G-RBF}(\boldsymbol{H}_1, \boldsymbol{H}_2) = \exp\{-2\rho\pi^2\},$$

$$K_{L-RBF}(\boldsymbol{H}_1, \boldsymbol{H}_2) = \exp\{-2\rho\pi\},$$

$$K_{d_{ab}-RBF}(\boldsymbol{H}_1, \boldsymbol{H}_2) = \exp\{-2\rho\pi^{ab}\},$$

where $L - RBF$ stands for Laplacian distances. It is clear that the exponential decay will be faster and faster as $ab \rightarrow 0$. This behavior is observed in Table 3 and partially in Table 2; on the contrary, Table 1 shows, for the 6D-histogram representation, very significant degradations of the recognition rate. We interpret this performance as a "curse of dimensionality" effect [2]; this hypothesis is confirmed by results obtained for 3D- and 2D - histogram representations; as the dimension of the chosen representation decreases, performances agree more and more with the theoretically expected behavior.

Discrepancies from this behavior presented also in Table 2 and 3 have to be interpreted as a side effect of the background. Indeed from Figure 1 it appears that the images have a lot of background; this "common information" between all the views of all the objects becomes relevant when $ab = 0.25$, in agreement with theoretical expectations. In the future we plan to perform experiments on segmented views of the DIROKOLL database, in order to confirm out interpretation of the experimental results.

7 Summary

In this paper we present a new class of MRF which is inspired by models of physics of disordered systems. It uses an energy function presenting two main advantages: it can be very easily applied to problems modeled by irregular sites because it considers the neighborhood system as fully connected; it does not require an algorithm for searching the absolute minima because those and their analytical properties are given by theory. A basic assumption in SGT is the infinite dimension of the configuration space; the choice of a particular energy, which depends on the scalar product between configurations, allow us to use the kernel trick and make it possible to use SGT results in a MRFs framework. Experiments on objects database show the effectiveness of the proposed approach.

Acknowledgments

This work has been partially supported by the "Graduate Research Center of the University of Erlangen-Nuremberg for 3D Image Analysis and Synthesis".

References

[1] D. J. Amit, *"Modeling Brain Function"*, Cambridge University Press, Cambridge, USA, 1989.

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

[3] B. Caputo, H. Niemann, "From Markov Random Fields to Associative Memories and Back: Spin-Glass Markov Random Fields ", *Proc. of IEEE workshop on Statistical and Computational Theories of Vision*, Vancouver, CA, July 2001, to appear.

[4] B. Caputo, J. Hornegger, D. Paulus, H. Niemann, "A spin-glass markov random field for 3D object recognition", Technical Report LME-TR-2001-01, January 2001.

[5] O. Chapelle, P. Haffner and V. Vapnik, "Svms for histogram-based image classification", *IEEE Trans. on Neural Networks*, 9, 1999.

[6] T. Fukai, S. Shiino, "Large suppression of spurious states in neural networks of nonlinear analog neurons", *Physical Review A*, Vol. 42, N. 12, pp. 7459-7466 Dec. 1990.

[7] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", *Proc. Natl. Acad. Sci. USA*, Vol. 79, pp 2554-2558, April 1982.

[8] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons", *Proc. Natl. Acad. Sci. USA*, Vol. 81, pp 3088- 3092, May 1984.

[9] Kirkpatric, S and Sherringtong D., "Infinite--ranged models of spin-glasses", *Physical Review B*, Vol. 17, pp. 4384-4404 1978.

[10] S. Z. Li, *"Markov Random Field Modeling in Computer Vision"*, Computer Science Workbench, Springer-Verlag, Tokyo, 1995.

[11] S. Z. Li and J. Hornegger, " A two-stage probabilistic approach for object recognition", In H. Burkhard and B. Neumann, editors, *Computer Vision — ECCV '98*, volume II of *Lecture Notes in Computer Science*, pages 733–747, Heidelberg, 1998. Springer.

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

[13] J.W. Modestino and J. Zhang. A Markov random field model–based approach to image interpretation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):606–615, June 1992.

[14] M. Reinhold, C. Drexler, H. Niemann, "Image Database for 3D Object Recognition", Technical Report LME-TR-2001-02, Chair for Pattern Recognition, 2001.

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

[16] B. Schölkopf, A. J. Smola, *Learning with kernels*, to appear, 2001, the MIT Press.

[17] M. Swain, D. Ballard, "Color indexing",*International Journal of Computer Vision*, 7, pp. 11-32, 1991.

[18] V. Vapnik, *Statistical learning theory*, J. Wiley, New York, 1998.

[19] M. D. Wheeler and K. Ikeuchi. Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):252–265, 1995.

[20] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, volume 27 of *Applications of Mathematics*. Springer, Heidelberg, 1995.

[21] S. C. Zhu, Y. Wu, D. Munford, "Filters, Random fields And Maximum Entropy (FRAME)", *International Journal of Computer Vision*, 27 (2), pp. 1- 20, 1998.