# An Ultrametric Approach to Object Recognition

B. Caputo (*,+), Gy. Dorkó (+), H. Niemann (+)

(*) Smith-Kettlewell Eye Research Institute
2318 Fillmore Street, San Francisco
94115 California, USA
(+) Department of Computer Science
Chair for Pattern Recognition Erlangen-Nuremberg University
Martenstrasse 3, 91058 Erlangen, Germany

## Abstract

This paper presents a Bayes classifier with a hierarchical structure for appearance-based object recognition. It consists of a new kernel method, Ultrametric Spin Glass-Markov Random Fields, that integrates results of statistical physics with Gibbs distributions. Experiments show the effectiveness of our approach.

## 1 Introduction

Object recognition is an important part of our lives. We recognize objects in all our everyday activities: we recognize people when we talk to them, we recognize our cup on the breakfast table, our car in a parking lot, and so on. While this task is performed with great accuracy and apparent little effort by humans, it is still unclear how this performance is achieved. This has challenged the computer vision research community to build artificial systems able to reproduce the human performance. After 30 years of intensive research, the challenge is still open.

Most of work on object recognition tries to answer the following question: given a collection of objects, can we recognize correctly one of them among the others? This problem is faced in a wide variety of situations: in cluttered of heterogeneous background [6, 10], under different lighting conditions [12], in presence of noise or occlusion [12, 6, 3], and so on. The success of an algorithm is then measured in terms of recognition rates, as to say how many times the object was recognized successfully. Although the recognition rate is undoubtfully an important indicator of the performance of an algorithm, it cannot be the only one. In other words, not all errors can be considered in the same way. This is something we experience everyday: we know that not all errors have the same consequences, and that there are mistakes that we cannot do more that once in life. If we ask someone "Please give me my pen", and (s)he makes a mistake, it is not the same if we get a different pen, or a cup. Many times, a pen (even if not ours) will do, but cup won't. If we walk in the forest and hear a noise that makes us realize that there is an animal close to us, the point is not to recognize whether it is a lion, a tiger, an antelope or a rabbit, but to decide whether it is dangerous or not, and react consequently.

Many experiments [11] show that biological systems tackle this problem using a hierarchical organization of visual information, based on visual categories [1]. Coding by category is fundamental to mental life because it greatly reduces the demands on perceptual processes, storage space, and reasoning processes. This also induces a hierarchical classification system: a visual pattern is recognized first as a member of a visual category, then as a member of a relative sub-category, and so on until it is recognized as individual (if known).

Recently, there has been some interest in the recognition of visual categories [13, 14]. In this paper we concentrate the attention on the hierarchical organization of information, and we propose a probabilistic Bayes classifier with a hierarchical structure for appearance-based object recognition. To this purpose, we use a new kernel method, Spin Glass-Markov Random Fields (SG-MRF, [2, 3]). They are a new class of MRF that integrates results

[1] We are aware that some researchers in the computer vision community argue that such parallel are unneeded. Although we understand their motivations, it should be firmly kept in mind that "biological vision is currently the only indication we have that the general vision problem is even open to solution" [11]

of statistical physics of disordered systems with Gibbs probability distributions via nonlinear kernel mapping [16]. The resulting model, using a Hopfield energy function, has shown to be very effective for appearance-based object recognition [2] and to be remarkably robust to noise and occlusion [3]. Here we extend SG-MRF to a new SG-like energy function, inspired by the ultrametric properties of the SG phase space. We will show that this energy can be kernelized as the Hopfield one, thus, it can be used in SG-MRF modeling. This new class of SG--MRF, that we call Ultrametric Spin Glass-Markov Random Fields (USG-MRF) has shown to be very effective for combining color and shape information [5]. Here we show that the structure of this energy provides as well a natural framework for hierarchical appearance-based object recognition; we report experimental results that show the effectiveness of our approach.

The paper is organized as follows: Section 2 defines the general framework for appearance-based object recognition, and Section 3 review SG-MRF. Section 4 presents the new ultrametric energy function, shows how it can be used in a SG-MRF framework (Section 4.1) and applied for hierarchical appearance-based object recognition (Section 4.2). Experiments are presented in Section 5; the paper concludes with a summary discussion.

## 2 Probabilistic Appearance-based Object Recognition

Appearance-based object recognition methods consider images as feature vectors. Let $\mathbf{x} \equiv [x_{ij}], i = 1, \ldots \mathcal{N}, j = 1, \ldots \mathcal{M}$ be an $\mathcal{M} \times \mathcal{N}$ image. We will consider each image as a feature vector $\mathbf{x} \in G \equiv \Re^m, m = \mathcal{M}\mathcal{N}$. Assume we have $k$ different classes $\Omega_1, \Omega_2, \ldots, \Omega_k$ of objects, and that for each object is given a set of $n_j$ data samples, $d_j = \{\mathbf{x}_1^j, \mathbf{x}_2^j, \ldots, \mathbf{x}_{n_j}^j\}, j = 1, \ldots k$. We will assign each object to a pattern class $\Omega_1, \Omega_2, \ldots, \Omega_k$. The object classification procedure will be a discrete mapping that assigns a test image, showing one of the objects, to the pattern class the presented object corresponds to. Here we will concentrate on probabilistic appearance-based methods.

The probabilistic approach to appearance-based object recognition considers the image views of a given object $\Omega_j$ as random vectors. Thus, given the set of data samples $d_j$ and assuming they are a sufficient statistic for the pattern class $\Omega_j$, the goal will be to estimate the probability distribution $P_{\Omega_j}(\mathbf{x})$ that has generated them. Then, given a test image $\mathbf{x}$, the decision will be made using a Maximum A Posteriori (MAP) classifier:

$$j^* = \underset{j}{\operatorname{argmax}}\, P_{\Omega_j}(\mathbf{x}) = \underset{j}{\operatorname{argmax}}\, P(\Omega_j|\mathbf{x}),$$

and, using Bayes rule,

$$j^* = \underset{j}{\operatorname{argmax}}\, P(\mathbf{x}|\Omega_j)P(\Omega_j). \quad (1)$$

where $P(f|\Omega_j)$ are the Likelihood Functions (LFs) and $P(\Omega_j)$ are the prior probabilities of the classes. In the rest of the paper we will assume that the prior $P(\Omega_j)$ is the same for all object classes; thus the Bayes classifier (1) simplifies to

$$j^* = \underset{j}{\operatorname{argmax}}\, P(\mathbf{x}|\Omega_j). \quad (2)$$

Many probabilistic appearance-based methods do not model the pdf on raw pixel data, but on features extracted from the original views. The extension of equation (2) to this case is straightforward: consider a set of features $\{\mathbf{h}_1^j, \mathbf{h}_2^j, \ldots, \mathbf{h}_{n_j}^j\}, j = 1, \ldots k$, where each feature vector $\mathbf{h}_{n_j}^j$ is computed from the image $\mathbf{x}_{n_j}^j, \mathbf{h}_{n_j}^j = T(\mathbf{x}_{n_j}^j), \mathbf{h}_{n_j}^j \in G \equiv \Re^m$. The Bayes classifier (2) will be in this case

$$j^* = \underset{j}{\operatorname{argmax}}\, P(\mathbf{h}|\Omega_j). \quad (3)$$

## 3 Spin Glass-Markov Random Fields

A possible strategy for modeling the parametric form of the probability function is to use Gibbs distributions within a Markov Random Field framework. MRF considers each element of the random vector $\mathbf{h}$ as the result of a labeling of all the sites representing $\mathbf{h}$, with respect to a given label set. The MRF joint probability distribution is given by

$$P(\mathbf{h}) = \frac{1}{Z}\exp\left(-E(\mathbf{h})\right), Z = \sum_{\{\mathbf{h}\}}\exp\left(-E(\mathbf{h})\right). \quad (4)$$

The normalizing constant $Z$ is called the partition function, and $E(\mathbf{h})$ is the *energy function*. Thus, using MRF modeling for appearance-based object recognition, eq (2) will become

$$j^* = \underset{j}{\operatorname{argmax}}\, P(\mathbf{h}|\Omega_j) = \underset{j}{\operatorname{argmin}}\, E(\mathbf{h}|\Omega_j) \quad (5)$$

Only a few MRF approaches have been proposed for high level vision problems such as object recognition [15, 8], due to the modeling problem for MRF on irregular sites (for a detailed discussion about this point, we refer the reader to [2]). Spin Glass-Markov Random Fields overcome this limitation and can be effectively used for appearance-based object recognition [2].

The rest of this Section will review SG-MRFs (Section 3.1) and how they can be derived from results of statistical physics of disordered systems (Section 3.2). Section 4 will show how these results can be extended to a new class of energy function and how this extension makes it possible to use this approach for hierarchical appearance-based object recognition.

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

Spin Glass-Markov Random Fields (SG-MRFs) [2] are a new class of MRFs which connect SG-like energy functions (mainly the Hopfield one [1]) with Gibbs distributions via a non linear 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 $k$ object classes $\Omega_1, \Omega_2, \ldots, \Omega_k$, and for each object a set of $n_j$ data samples, $d_j = \{\mathbf{x}_1^j, \ldots \mathbf{x}_{n_j}^j\}, j = 1, \ldots k$. We will suppose to extract, from each data sample $d_j$ a set of features $\{\mathbf{h}_1^j, \ldots \mathbf{h}_{n_j}^j\}$. The SG-MRF probability distribution is given by

$$P_{SG}(\mathbf{h}|\Omega_j) = \frac{1}{Z} \exp\left[-E_{SG}(\mathbf{h}|\Omega_j)\right], \quad (6)$$

$$Z = \sum_{\{\mathbf{h}\}} \exp\left[-E_{SG}(\mathbf{h}|\Omega_j)\right],$$

with

$$E_{SG}(\mathbf{h}|\Omega_j) = -\sum_{\mu=1}^{p_j} \left[K(\mathbf{h}, \tilde{\mathbf{h}}^{(\mu_j)})\right]^2, \quad (7)$$

where the function $K(\mathbf{h}, \tilde{\mathbf{h}}^{(\mu_j)})$ is a Generalized Gaussian kernel [16]:

$$K(\mathbf{x}, \mathbf{y}) = \exp\{-\rho d_{a,b}(\mathbf{x}, \mathbf{y})\},$$

$$d_{a,b}(\mathbf{x}, \mathbf{y}) = \sum_i |x_i^a - y_i^a|^b. \quad (8)$$

$\{\tilde{\mathbf{h}}^{(\mu_j)}\}_{\mu=1}^{p_j}, j \in [1, k]$ are a set of vectors selected (according to a chosen ansatz, [2]) from the training data that we call *prototypes*. The number of prototypes per class must be finite, and they must satisfy the condition:

$$K(\tilde{\mathbf{h}}^{(i)}, \tilde{\mathbf{h}}^{(l)}) = 0, \quad (9)$$

for all $i, l = 1, \ldots p_j, i \neq l$ and $j = 1, \ldots k$. Note that SG-MRFs are defined on features rather than on raw pixels data. The sites are fully connected, which ends in learning the neighborhood system from the training data instead of choosing it heuristically. As we model the probability distribution on feature vectors and not on raw pixels, SG-MRF is not a generative model. Another key characteristic of the model is that in SG-MRF the functional form of the energy is given by construction. This is achieved using results for statistical physics of Spin Glasses. The next Section sketches the theoretical derivation of the model. The interested reader will find a more detailed discussion in [2].

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

Consider the following energy function:

$$E = -\sum_{(i,j)} J_{ij}\ s_i\ s_j \qquad i, j = 1, \ldots N, \quad (10)$$

where the $s_i$ are random variables taking values in $\{\pm 1\}$, $\mathbf{s} = (s_1, \ldots, s_N)$ is a configuration and $\mathbf{J} = [J_{ij}\ ], (i, j) = 1, \ldots, N$ is the connection matrix, $J_{ij} \in \{\pm 1\}$. Equation (10) is the most general Spin Glass (SG) energy function [1, 7].

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

$$J_{ij}\ = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)}\ \xi_j^{(\mu)}\ , \quad (11)$$

where the $p$ sets of $\{\xi^{(\mu)}\}_{\mu=1}^{p}$ are given configurations of the system (that we call *prototypes*) having the following properties: *(a)* $\xi^{(\mu)} \perp \xi^{(\nu)}, \forall \mu \neq \nu$; *(b)* $p = \alpha N, \alpha \leq 0.14, N \to \infty$. Under these assumptions it has been proved that the $\{\xi^{(\mu)}\}_{\mu=1}^{p}$

are the absolute minima of $E$ [1]; 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. $\mathbf{s} \in [-1, +1]^N$); note that this extension is crucial in the construction of the SG-MRF model.

Energy (10), with the prescription (11), can be written as:

$$E = -\frac{1}{N} \sum_{\mu} (\xi^{(\mu)} \cdot \mathbf{s})^2. \tag{12}$$

Equation (12) depends on the data through scalar products, thus it can be *kernelized*, as to say it can be written as

$$E = \frac{1}{N} \sum_{\mu} (\xi^{(\mu)} \cdot \mathbf{s})^2 =$$

$$-\frac{1}{N} \sum_{\mu} [\Phi(\tilde{\mathbf{h}}^{(\mu)}) \cdot \Phi(\mathbf{h})]^2 =$$

$$= -\frac{1}{N} \sum_{\mu} [K(\tilde{\mathbf{h}}^{\mu}, \mathbf{h})]^2 \tag{13}$$

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

$$\Phi : G \equiv \Re^m \rightarrow H \equiv [-1, +1]^N, N \rightarrow \infty,$$

that in terms of kernel means

$$K(\mathbf{h}, \mathbf{h}) = 1, \forall \quad \mathbf{h} \in \Re^m, dim(H) = N, N \rightarrow \infty. \tag{14}$$

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* [16], which was first used for Support Vector Machines (SVM). Conditions (14) are satisfied by generalized Gaussian kernels (8). Regarding the choice of prototypes, given a set of $n_k$ training examples $\{\mathbf{x}_1^{\kappa}, \mathbf{x}_2^{\kappa}, \ldots, \mathbf{x}_{n_\kappa}^{\kappa}\}$ relative to class $\Omega_\kappa$, the condition to be satisfied by the prototypes is $\xi^{(\mu)} \perp \xi^{(\nu)}, \forall (\mu \neq \nu)$ in the mapped space $H$, that becomes $\Phi(\tilde{\mathbf{h}}^{(\mu)}) \perp \Phi(\tilde{\mathbf{h}}^{(\nu)}), \forall \mu \neq \nu$ in the data 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, has the effect of orthogonalize the patterns in the space $H$. 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. This approximation is called the *naive ansatz*.

# 4 Ultrametric Spin Glass-Markov Random Fields

SG-MRF, with the Hopfield energy function (10)-(11), have been successfully applied to appearance-based object recognition, showing to be very effective, and presenting remarkable robustness properties [3]. A major drawback of the Hopfield energy function is the condition of orthogonality on the set of prototypes. Although the properties of generalized Gaussian kernels ensure theoretically that there exists at least one $\rho$ such that all the prototypes are orthogonal to each other, this can be not enough from the point of view of computer vision applications. In other words, the naive ansatz can turn out to be in some cases too rough an approximation.

The solution we propose consists in kernelizing a new SG energy function, that allows us to store non mutually orthogonal prototypes. As this energy was originally derived taking into account the ultrametric properties of the SG configuration space, we will refer to it as the *ultrametric energy*. The interested reader will find a complete description of ultrametricity and of the ultrametric energy in [1, 7]. In the rest of the Section we will present the ultrametric energy and we will show how it can be kernelized (Section 4.1); we will also show that, as a main feature of the ultrametric energy is that it induces a hierarchical organization of the data, it can be used for hierarchical appearance-based object recognition.

## 4.1 Ultrametric Spin Glass-Markov Random Fields: Model Derivation

Consider the energy function (10)

$$E = -\sum_{ij} J_{ij} s_i s_j$$

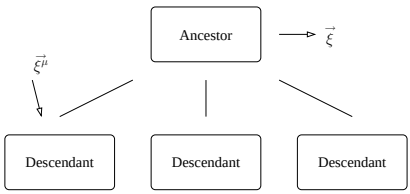


Figure 2: Hierarchical structure induced by the ultrametric energy function.

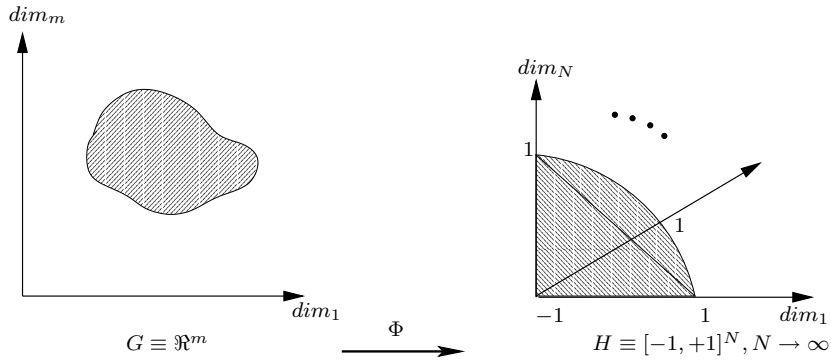


Figure 1: The kernel trick maps the data from a lower dimension space $G \equiv \Re^m$ to a higher dimension space $H \equiv [-1,+1]^N, N \to \infty$. This permits to use the H-L energy in a MRF framework.

with the following connection matrix:

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{(\mu)}\xi_j^{(\mu)} + \frac{1}{N\Delta(a_\mu)}\sum_{\mu=1}^{p} \xi_i^{(\mu)}\xi_j^{(\mu)} \cdot \sum_{\nu=1}^{q_\mu}(\eta_i^{(\mu\nu)} - a_\mu)(\eta_j^{(\mu\nu)} - a_\mu) \qquad (15)$$

with

$$\xi_i^{(\mu\nu)} = \xi_i^{(\mu)}\eta_i^{(\mu\nu)}, \qquad a_\mu^2 = \frac{1}{N}\sum_{i=1}^{N} \eta_i^{(\mu\nu)}\eta_i^{(\mu\lambda)}.$$

This energy induces a hierarchical organization of stored prototypes ([1], see Figure 3). The set of prototypes $\{\xi^{(\mu)}\}_{\mu=1}^{p}$ are stored at the first level of the hierarchy and are usually called the *ancestors*. Each of them will have $q$ *descendants* $\{\xi^{(\mu\nu)}\}_{\nu=1}^{q_\mu}$. The parameter $\eta_i^{(\mu\nu)}$ measures the similarity between ancestors and descendants; the parameter $a_\mu$ measures the similarity between descendants. $\Delta(a_\mu)$ is a normalizing parameter, that guarantees that the energy per site is finite. In the rest of the paper we will limit the discussion to the case[2]

$$a_\mu^2 = a^2.$$

The connection matrix thus becomes:

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{(\mu)}\xi_j^{(\mu)} + \frac{1}{N(1-a^2)}\sum_{\mu=1}^{p} \xi_i^{(\mu)}\xi_j^{(\mu)} \sum_{\nu=1}^{q_\mu}(\eta_i^{(\mu\nu)}-a)(\eta_j^{(\mu\nu)}-a) = \mathrm{Term1} + \mathrm{Term2}.$$

Term1 is the Hopfield energy (10)-(11); Term2 is a new term that allows us to store as prototypes patterns correlated with the $\{\xi^{(\mu)}\}_{\mu=1}^{p}$, and correlated between each other. This energy will have $p + \sum_{\mu=1}^{p} q^\mu$ minima, of which $p$ absolute (ancestor level) and $(\sum_{\mu=1}^{p} q^\mu)$ local (descendant level).

When $a \to 0$, the ultrametric energy reduces to a hierarchical organization of Hopfield energies; in this case the prototypes at each level of the hierarchy must be mutually orthogonal, but they can be correlated between different levels. Note also that we limited ourselves to two levels, but the energy can be easily extended to three or more. For a complete discussion on the properties of this energy, we refer the reader to [1].

Here we are interested in using this energy in the SG-MRF framework reviewed in Section 3. To this purpose, we show that the energy (10), with the connection matrix (15), can be written as a function of scalar product between configurations:

$$E = -\frac{1}{N}\sum_{\mu=1}^{p}(\xi^{(\mu)}\cdot\mathbf{s})^2 + \frac{1}{N(1-a^2)}\sum_{\mu=1}^{p}\sum_{\nu=1}^{q_\mu}(\xi^{(\mu\nu)}\cdot\mathbf{s})^2 - \frac{2a}{N(1-a^2)}\sum_{\mu=1}^{p}\sum_{\nu=1}^{q_\mu}(\xi^{(\mu)}\cdot\mathbf{s})(\xi^{(\mu)}\cdot\mathbf{s}) +$$

[2] Considering the general case would not add anything from the conceptual point of view and would make the notation even heavier.

$$\frac{a^2}{N(1-a^2)} \sum_{\mu=1}^{p} \sum_{\nu=1}^{q_\mu} (\xi^{(\mu)} \cdot \mathbf{s})^2. \qquad (16)$$

If we assume that $a \to 0$, as to say we impose orthogonality between prototypes at each level of the hierarchy, the energy reduces to

$$E = -\frac{1}{N^2} \sum_{\mu=1}^{p} (\xi^{(\mu)} \cdot \mathbf{s})^2 +$$

$$\sum_{\mu=1}^{p} \sum_{\nu=1}^{q_\mu} (\xi^{(\mu\nu)} \cdot \mathbf{s})^2. \qquad (17)$$

The *ultrametric energy*, in the general form (16) or in the simplified form (17) can be kernelized as done for the Hopfield energy and thus can be used in a MRF framework. We call the resulting new MRF model Ultrametric Spin Glass-Markov Random Fields (USG-MRF).

### 4.2 Ultrametric Spin Glass-Markov Random Fields: Model Application

The ultrametric energy (17), derived in Section 4.1, becomes in the SG-MRF framework

$$E_{USG} = -\sum_{\mu=1}^{p_j} [K_a(\tilde{\mathbf{h}}^{(\mu)}, \mathbf{h})]^2 -$$

$$\sum_{\mu=1}^{p_j} \sum_{\nu=1}^{q_\mu} [K_d(\tilde{\mathbf{h}}^{(\mu\nu)}, \mathbf{h})]^2, \qquad (18)$$

where $\{\mathbf{h}^\mu\}_{\mu=1}^{p_j}$ will be the set of prototypes relative to the ancestor level, and $\{\mathbf{h}^{\mu\nu}\}_{\nu=1}^{q_\mu}, \mu = 1, \ldots p_j$ the set of prototypes at the descendant level. $K_a$ is the generalized Gaussian kernel at the ancestor level, and $K_d$ is the generalized Gaussian kernel at the descendant level. It must be stressed that the kernel must be the same at each level of the hierarchy, but can be different between levels (as to say between ancestor and descendant).

The ultrametric energy (18) has been used for combining color and shape information for appearance-based object recognition, with excellent results [5]. Here we apply USG-MRF to hierarchical appearance-based object recognition, in a straightforward manner: the ancestor level will contain prototypes relative to the *visual category* the object $\Omega_j$ belongs to, while the descendant level will contain prototypes relative to the object class itself. The Bayes classifier based on USG-MRF is:

Figure 3: An example of 5 objects of the 59 contained into the used database. Views have different sizes for different objects and for different pose parameters.

$$j^* = \underset{j}{\operatorname{argmin}} \{ -\sum_{\mu=1}^{p_j} [K_a(\tilde{\mathbf{h}}^{(\mu)}, \mathbf{h})]^2 -$$

$$\sum_{\mu=1}^{p_j} \sum_{\nu=1}^{q_\mu} [K_d(\tilde{\mathbf{h}}^{(\mu\nu)}, \mathbf{h})]^2 \}. \qquad (19)$$

## 5 Experiments

We performed a series of experiments in order to test our model. We ran all the experiments on a database of 59 objects [10]: 11 cups, 5 dolls, 6 planes, 6 fighter jets, 9 lizards, 5 spoons, 8 snakes and 9 sport cars. Some examples are shown in Figure 4.

Each object is represented in the training set by a collection of views taken approximately every 20 degrees on a sphere; this amounts to 106 views for a full sphere, and 53 for a hemisphere. The test set consists of 53 (24) views, positioned in between the training views, and taken under the same conditions. Cups, dolls, fighters, planes, spoons are represented by 106 views in the training set and 53 views in the test set; lizards, snakes, sport cars are represented by 53 views in the training set and 24 views in the test set. As the views in the database are of different sizes, we decided to use a Multidimensional receptive Field Histogram (MFH) representation for all classes [12], that was already applied and successfully combined with SG-MRF [3]. We used three different kinds of MFH representation: the first consisted in Gaussian derivatives along $x$ and $y$ directions and with $\sigma = 1.0$, that we called $D_x D_y$. The second consisted in Laplacian Gaussian operator with $\sigma_1 = 1.6$ and $\sigma_2 = 3.2$, that we called $Lp2\sigma$. The third consisted in Laplacian Gaussian operator with $\sigma_1 = 1.6$, $\sigma_2 = 3.2$

and $\sigma_3 = 6.4$, that we called $Lp3\sigma$. For all these representations, the resolution for histogram axes was of 16 bins.

The structure of this database is "naturally" hierarchical: although it is composed by 59 objects, they can be "naturally" separated in subgroups with respect to the visual category they belong to (cup, doll, fighter, lizard, spoon, sport car). The goal of these experiments is to check whether the use of this hierarchical structure can lead to a) a higher recognition rate, b) a lower recognition time, c) a lower category error rate.

Thus, we ran a first set of experiments on the complete database of 59 objects, using SG-MRF in a MAP-Bayes classifier [2, 3], for all the three representations described above. Kernel parameters are learnt with a leave-one-out technique. These results constitute a benchmark for those obtained using USG-MRF, and are reported in Table 1.

| | $D_xD_y$ | $Lp2\sigma$ | $Lp3\sigma$ |
|---|---|---|---|
| Rec. Rate (%) | 90.82 | 97.74 | 98.23 |
| Rec. time (sec) | 0.67 | 1.35 | 6.77 |
| n. misclass. | 217 | 50 | 39 |
| category errors | 42 | 4 | 6 |

Table 1: Classification results using SG-MRF, for 59 objects

We see that the best recognition rate, obtained with the $Lp3\sigma$ representation, doesn't correspond to the lower category error. Moreover, the higher recognition rate is obtained with a 3D histogram representation, which leads to a remarkable increase of the computational time.

Then we ran a second set of experiments, consisting in the recognition of the 8 visual categories mentioned above. For each visual category, the training (test) set consisted of all the training (test) view of the object belonging to the category itself. We used a $D_xD_y$ and $Lp2\sigma$ representation, and SG-MRF as described in the previous experiment. Results are reported in Table 2.

We see that, using the $Lp2\sigma$ representation, we achieve a recognition rate of 100%, which of course corresponds in this case to 0 category errors. This inspired us to perform two hierarchical experiments. In the first experiment, we used the USG--MRF classifier as described in Section 4.2, and the $D_xD_y$ representation at the ancestor level (visual

| | $D_xD_y$ | $Lp2\sigma$ |
|---|---|---|
| Rec. Rate (%) | 98.23 | 100 |
| category errors | 39 | 0 |

Table 2: Classification results using SG-MRF, for 8 object categories

category level) and the $Lp3\sigma$ at the descendant level (object class level). Kernel parameters at each level were learnt with a leave-one-out technique. Results are reported in Table 3. In the second experiment, we used SG-MRF and $Lp2\sigma$ for the visual category classification; we considered this result prior knowledge and then we used again SG-MRF for the recognition of each object class. This procedure allows to use, for each group of objects belonging to the same visual category, a different kernel. Results are reported in Table 3.

| | $D_xD_y$-$Lp3\sigma$ | $Lp2\sigma$- $Lp3\sigma$ |
|---|---|---|
| Rec. Rate (%) | 98.28 | 98.59 |
| Rec. time (sec) | 3.18 | 4.94 |
| n. misclass. | 38 | 30 |
| category errors | 2 | 0 |

Table 3: Classification results using USG-MRF, for 59 objects

The best recognition rate is obtained with the second experiment, which gives also (not surprisingly) the lowest category error. Nevertheless, results obtained using USG-MRF are impressive: it is faster with respect to the second experiment, with just 8 more misclassifications, 2 of which category errors. This last result is particularly impressive, because we used the $D_xD_y$ representation at the ancestor level, which would give alone 39 category errors (see Table 2). This result, with the awareness that the second experiment relies heavily on the performance at the first level (that gives a 100% recognition rate in this case, but that by principle can be different for different databases), makes us conclude that USG-MRF is the most promising strategy for hierarchical appearance-based object recognition.

Finally, we compare USG-MRF results with those obtained with the first set of experiments. We see from Table 1 and Table 3 that USG-MRF gives the highest recognition rate and the lowest category

error; it pays a price in terms of computational time with respect to the $Lp2\sigma$ representation (second column, Table 1), but this is well compensated by the higher recognition rate and the lower category error. We can conclude from these experiments that USG-MRF is an effective probabilistic method for hierarchical appearance-based object recognition.

## 6 Summary

In this paper we presented a new kernel method for hierarchical appearance-based object recognition. This result is achieved using results of statistical mechanics of Spin Glasses combined with Markov Random Fields via kernel functions. The new model is an extension of Spin Glass-Markov Random Fields to a new class of SG-like energy functions, that use the ultrametric properties of the Spin Glass phase space; for this reason we call the new model Ultrametric Spin Glass-Markov Random Fields. Experiments confirm the effectiveness of the proposed approach. This work can be developed in many ways: first, we intend to develop new strategies for the recognition of visual categories, based on the choice of proper representations and the use of kernel properties. Second, we plan to use the Ultrametric Spin Glass-Markov Random Field in the Kernel-Class Specific Classifier framework [4], in order to fully use the power of kernels. Finally, we plan to benchmark these results with other kernel methods like Support Vector Machines.

### Acknowledgments

This work has been supported by the "Graduate Research Center of the University of Erlangen-Nuremberg for 3D Image Analysis and Synthesis", and by the Foundation BLANCEFLOR Boncompagni-Ludovisi. B. C. thanks A. Yuille for useful discussions.

## References

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

[2] B. Caputo, H. Niemann, "From Markov Random Fields to Associative Memories and Back: Spin Glass Markov Random Fields", SCTV2001.

[3] B. Caputo, S. Bouattour, H. Niemann, "Robust appearance-based Object Recognition using a Fully Connected Markov Random Field", ICPR02.

[4] B. Caputo, H. Niemann, "To Each According to its Need: Kernel Class Specific Classifier", ICPR02.

[5] B. Caputo, Gy. Dorkó, H. Niemann, "Combining Color and Shape Information for Appearance-based Object Recognition using Kernel Gibbs Distributions", ICPR02.

[6] A. Leonardis, H. Bischof, "Robust recognition using eigenimages", CVIU,78:99-118, 2000.

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

[8] J.W. Modestino, J. Zhang. "A Markov random field model–based approach to image interpretation". *PAMI*, 14(6),606–615,1992.

[9] H. Murase, S.K. Nayar, "Visual Learning and Recognition of 3D Objects from Appearance", *IJCV*,14(1):5-24, 1995.

[10] R. C. Nelson, A. Selinger, "A cubist approach to object recognition", ICCV98:614-621, 1998.

[11] D. N. Ohersen, E. E. Smith, *An invitation to cognitive science: Thinking*, the MIT press.

[12] B. Schiele, J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms", *IJCV*, 36(1),:31-52, 2000.

[13] M. Weber, M. Welling, P. Perona, "Towards Automatic Discovery of Objects Categories", *Proc. of CVPR2000*.

[14] M. Weber, M. Welling, P. Perona, "Unsupervised Learning of Models for Recognition", *Proc. of ECCV2000*.

[15] M. D. Wheeler, K. Ikeuchi. *Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. PAMI*, 17(3):252–265, 1995.

[16] B. Schölkopf, A. J. Smola, *Learning with kernels*, 2002, the MIT Press, Cambridge, MA.