

Robust Appearance-based Object Recognition using a Fully Connected Markov Random Field

Abstract

This paper presents a new kernel method for appearance-based object recognition, highly robust to noise and occlusion. It consists of a fully connected Markov Random Field that integrates results of Spin Glass theory with Gibbs probability distributions via nonlinear kernel mapping. We call the resulting model Spin Glass-Markov Random Field. We present theoretical analysis and several experiments that show its effectiveness and robustness to noise and occlusion. We obtain in both cases excellent results. Particularly, we achieve a recognition rate above 93 % with just 40 % of visible portion of the object.

1 Introduction

This paper presents a novel probabilistic approach to appearance-based object recognition, based on a fully connected Markov Random Field (MRF), which shows robustness to noise and occlusion. Object recognition is a challenging topic. Many approaches use appearance-based methods [6, 5, 4]. They consider the appearance of objects using just two-dimensional images representation. Robustness is an open point for these approaches. Several methods have been proposed in order to cope with noise, occlusion and cluttered background [8, 3, 7].

In this paper we present a novel strategy, consisting of a fully connected Markov Random Field that integrates results of statistical physics of disordered systems [1] with Gibbs probability distributions via non linear kernel mapping. The resulting model, that we call Spin Glass-Markov Random Field (SG-MRF, [2]), can be used for appearance-based object recognition; here we present several experiments that show its robustness. Robustness is explored with respect to noise and occlusion, obtaining in both cases excellent results. Particularly, we achieve a recognition rate above 93 % with just 40 % of visible portion of the object.

Next Section defines SG-MRFs and their properties; the rest of the paper reports the experimental results showing

the high robustness of the model. The paper concludes with a summary discussion.

2 Model Definition

Consider an object Ω_j and a set of k observations $\{\mathbf{x}^1 \dots \mathbf{x}^k\}$, $\mathbf{x} \in \mathfrak{R}^m$, that we consider random samples from the underlying, unknown, probability distribution $P(\mathbf{x})$ defined on \mathfrak{R}^m . Consider also \mathcal{K} different object classes $\Omega_j, j = \{1, \dots, \mathcal{K}\}$. Given an observation $\hat{\mathbf{x}}$, our goal is to classify $\hat{\mathbf{x}}$ as a sample from Ω_{j^*} , one of the Ω_j object classes. Using a Maximum A Posteriori (MAP) criterion we have

$$j^* = \operatorname{argmax}_j P(\Omega_j | \mathbf{x}) = \operatorname{argmax}_j \{P(\mathbf{x} | \Omega_j) P(\Omega_j)\};$$

using Bayes rule, where $P(\mathbf{x} | \Omega_j)$ are the Likelihood Functions (LFs) and $P(\Omega_j)$ are the prior probabilities of the classes. Assuming that $P(\Omega_j)$ are constant, the Bayes classifier simplifies to

$$j^* = \operatorname{argmax}_j P(\mathbf{x} | \Omega_j).$$

2.1 Spin Glass-Markov Random Fields

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 to use the power of kernels in a probabilistic framework. The SG-MRF probability distribution is given by

$$P_{SG-MRF}(\mathbf{x} | \Omega_j) = \frac{1}{Z} \exp[-E_{SG-MRF}(\mathbf{x} | \Omega_j)], \quad (1)$$

$$Z = \sum_{\{\mathbf{x}\}} \exp[-E_{SG-MRF}(\mathbf{x} | \Omega_j)],$$

with

$$E_{SG-MRF} = - \sum_{\mu=1}^{p_j} \left[K(\mathbf{x}, \tilde{\mathbf{x}}^{(\mu)}) \right]^2, \quad (2)$$

where the function $K(\mathbf{x}, \tilde{\mathbf{x}}^\mu)$ is a Generalized Gaussian kernel [9]:

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

$\{\tilde{\mathbf{x}}^\mu\}_{\mu=1}^{p_j}, j \in [1, \mathcal{K}]$ 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:

$$K(\tilde{\mathbf{x}}^i, \tilde{\mathbf{x}}^k) = 0, \quad (3)$$

for all $i, k = 1, \dots, p_j, i \neq k$ and $j = 0, \dots, \mathcal{K}$. The interested reader can find a detailed discussion regarding the derivation and properties of SG-MRF in [2].

2.2 Robustness of Spin Glass-Markov Random Fields

It can be shown [2] that equation (2) can be interpreted as a kernelized SG-like energy. This interpretation allows to use results of SG theory in order to understand the properties of the new model. Of particular interest for appearance-based object recognition applications are theoretical results which predict a high robustness of SG-MRF to noise and occlusion [1]. SG theory predicts that the model will be more robust to occlusion than to noise. Moreover, it predicts that the presence of degraded views into the training set will result in an increase of the performance and in a better generalization capability. It is impossible, due to space limitations, to enter here into the details of the derivation of these results. In the next Section, we will present experiments showing these behaviors. Here we prefer to give a more intuitive interpretation, based on the functional form of the energy (2).

First, a *kernel* is a function K such that, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$,

$$K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \cdot \Phi(\mathbf{y}),$$

where Φ is a mapping from \mathbb{R}^m to an (inner product) feature space F [9]. Thus, the SG-MRF energy function can be rewritten as:

$$\begin{aligned} E_{SG-MRF} &= - \sum_{\mu=1}^{p_j} [K(\mathbf{x}, \tilde{\mathbf{x}}^\mu)]^2 = - \sum_{\mu=1}^{p_j} [\Phi(\mathbf{x})^T \cdot \Phi(\tilde{\mathbf{x}}^\mu)]^2 = \\ &= - \sum_{\mu=1}^{p_j} \Phi(\mathbf{x})^T \cdot \Phi(\tilde{\mathbf{x}}^\mu) \Phi(\tilde{\mathbf{x}}^\mu)^T \cdot \Phi(\mathbf{x}); \\ \Rightarrow E_{SG-MRF} &= - \sum_{i,j=1}^m \Phi(x_i) C_{ij} \Phi(x_j), \end{aligned} \quad (4)$$

where

$$C_{ij} = \sum_{\mu=1}^{p_j} \Phi(\tilde{x}_i^\mu) \Phi(\tilde{x}_j^\mu)^T \quad (5)$$

is the *connection matrix*. Equation (4) shows that the SG-MRF energy function is given by the sum, on all the components of the vector, of all the possible coupled interactions between components, weighted by a factor given by the corresponding element of the connection matrix (5). In order to fix ideas, let us consider the case when the \mathbf{x} is the image view (and thus x_i is the pixel). The energy (4) models the pdf generating \mathbf{x} considering all the possible interaction between couples of pixels, weighted by a factor C_{ij} . Thus, the energy is fully connected, which makes the representation global. But this globality is achieved via the sum of all the possible local interactions, weighted by a coefficient (the matrix element C_{ij}) which is learned by the training data (note indeed that the connection matrix is determined by the set of prototypes, eq. (5)). This means that many of those contributions can be zeros. Thus we achieve a globality that is the result of the contribution of all the significative localities for the considered object. As a consequence of this property of SG-MRFs we expect them to be robust to noise (due to the globality) as well as to occlusion (due to the locality).

3 Experimental Results

In this Section we present experiments that show the robustness of SG-MRF to noise and occlusion. Experiments were done on the 100 objects of the Columbia [4] 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 [7]¹, that has been already used successfully combined with SG-MRF [2]. 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. For the classification step, we used SG-MRF in the MAP-MRF framework described in Section 2. For the choice of prototypes, we made a naive ansatz [2], which means that all training views are taken as prototypes, and the ρ in the Gaussian kernel is learned so to satisfy condition (3). The performance of SG-MRF was compared with a nearest neighbor classifier with χ^2 distance, which proved to be an effective comparison measurement for the MFH representation [7].

3.1 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 evaluation its robustness in the later experiments. Each training's

¹We gratefully thank B. Schiele who allowed us to use his software for the computation of MFH.

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 χ^2 and SG-MRF, the six chosen Gaussian kernels and the different databases are reported in Table 1. For unperturbed images, SG-MRF performs comparable to χ^2 , for some kernels even slightly better.

χ^2	SG-MRF			
99.52	$a = 1, b = 2$	94.43	$a = 0.5, b = 2$	99.58
	$a = 1, b = 1$	97.84	$a = 0.5, b = 1$	99.80
	$a = 1, b = 0.5$	99.38	$a = 0.5, b = 0.5$	99.84

Table 1. Classification results for χ^2 and SG-MRF.

3.2 Robustness to Noise

In order to test robustness of SG-MRF with respect to noise, we ran several series of experiments. We added to the views Independent Gaussian noise with $\sigma \in \{10, 50, 80, 120\}$; some examples of noisy views are shown in Figure 1. For



Figure 1. Examples of noisy images: from left to right, $\sigma \in \{10, 50, 80, 120\}$

each level of noise, we generated four 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 randomly². For each noise level, for each test set we ran a set of 5 experiments, one for each percentage of noise in the training set. For example, for the noise level $\sigma = 10$, we ran five experiments with 0% noisy views in the training set and respectively 0%, 25%, 50%, 75 % and 100% of noisy views in the test set. Then, we ran other 5 experiments with 25 % of noisy views in the training set; and so on. The views were represented using 2D-histograms; the kernel parameters were $a = 0.5$ and

²A pilot experiment with a uniform distribution of noisy views led to comparable outcomes.

$b = 1$. Due to space constraints, we present just the most significant results.

Table 2 shows the recognition rates obtained for the noise level of $\sigma = 10$, three different percentage of noisy views in the training set, three different percentage of noisy views in the test set, using SG-MRF and χ^2 . We see that robustness to noise improves dramatically when noisy views are included in the training data. For 50 % of noisy views the recognition rates, for all the different test sets, is above 99%; for 100 % of noisy views the recognition rate is above 97% for all the test sets. This last result, with the results obtained using χ^2 , show particularly that the increased robustness is due to the generalization capabilities of SG-MRF, and not to a “similarity effect” due to the presence of similarly degraded views in the training and test set. If this would be the case, we should observe in the last row of Table 2, SG-MRF results, more or less the same value we observe in the first row, in reversed order. This is instead observed in Table 2, χ^2 results. Figure 2 shows the recognition rates obtained with SG-MRF, for different levels of noise, with 50% of noisy views into the training set. We see that until a certain noise level ($\sigma = 50$), SG-MRF has a very high recognition rate and at the same time a great stability. As the level of noise increases, SG-MRF starts to lose stability.

SG-MRF	0 % te	50 % te	100 % te
0% tr	99.80	80.84	61.95
50 % tr	97.75	97.34	96.73
100 % tr	94.34	97.09	99.68

χ^2	0 % te	50 % te	100 % te
0% tr	99.52	58.88	18.04
50 % tr	95.66	95.76	95.36
100 % tr	33.13	66.36	99.45

Table 2. Classification results using SG-MRF (top) and χ^2 (bottom) for different percentage of noisy views in the training and testing set; noise level $\sigma = 10$.

3.3 Robustness to Occlusion

We tested robustness to occlusion of SG-MRF reducing gradually the amount of visible portion of the object, keeping it centered within the image. This corresponds to the ideal case that the location of the object is known. We computed the recognition rates for the following visible portions: {30%, 40%, 50%, 60%, 70%, 80%}; Figure 3 shows the resulting object portions.

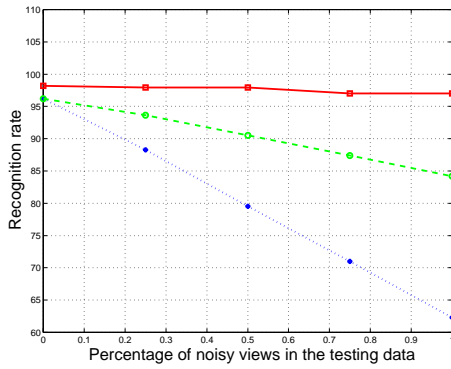


Figure 2. Robustness to noise: recognition rates as a function of the different percentages of noisy views in the test set. The training set consisted of 36 views, of which 50% were affected by increasing levels of noise: $\sigma = 50$ (-), $\sigma = 80$, (- -) and $\sigma = 120$ (:)



Figure 3. Examples of occluded views. From left to right: visible object portions {80%, 70%, 60%, 50%, 40%, 30%}

The views were represented using 3D-histograms. The kernel parameters were $a = 0.5$ and $b = 2$. Figure 4 shows the recognition rates as a function of the amount of occlusion in the testing set, for uncorrupted training views and for a training set containing 50% of occluded views. We see that including 50% of degraded views into the training data leads to a dramatical improvement of the recognition rate. We achieve a recognition rate above 93% with just 40% of visible object portion. On the basis of these results we expect SG-MRF to be robust to occlusions as well as to clutter background.

4 Conclusions

We presented in this paper a new probabilistic method for appearance-based object recognition that is highly robust to noise and occlusion. In particular, we obtain a recognition rate above 93% when just 40% of the object is visible, on a database of 100 objects. Future work will apply these results to the recognition of objects in cluttered background; we also plan to benchmark our technique with other strategies proposed in the literature.

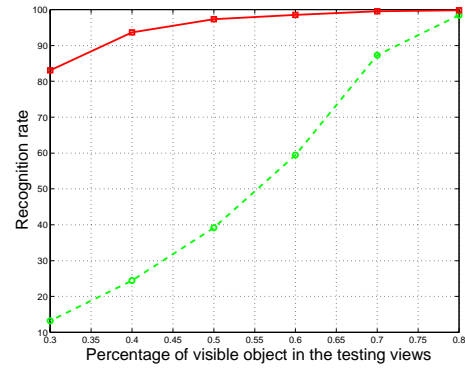


Figure 4. Robustness to occlusion: recognition rates as a function of the percentage of visible object in the testing set; the model is trained on ideal data (- -) and on 50% ideal, 50% degraded views (-) .

References

- [1] D. J. Amit, “*Modeling Brain Function*”, Cambridge University Press, Cambridge, USA, 1989.
- [2] B. Caputo, H. Niemann, “From Markov Random Fields to Associative Memories and Back: Spin Glass Markov Random Fields”, *SCTV01*.
- [3] A. Leonardis, H. Bischof, “Robust recognition using eigenimages”, *CVIU*,78:99-118, 2000.
- [4] H. Murase, S.K. Nayar, “Visual Learning and Recognition of 3D Objects from Appearance”, *IJCV*,14(1):5-24, 1995.
- [5] R. C. Nelson, A. Selinger, “A cubist approach to object recognition”, *ICCV98*:614-621, 1998.
- [6] K. Ohba, K. Ikeuchi, “Recognition of the multi specular objects for binpicking task”, *IROS96*, 3:1440-1448, 1996.
- [7] B. Schiele, J. L. Crowley, “Recognition without correspondence using multidimensional receptive field histograms”, *IJCV*, 36 (1), pp. 31- 52, 2000.
- [8] C. Schmid, R. Mohr, “Combining grayvalue invariants with local constraints for object recognition”, *CVPR96*:872-877, 1996.
- [9] B. Schölkopf, A. J. Smola, *Learning with kernels*, 2001, the MIT Press.