A Marxist Approach to Object Recognition: Kernel-Class Specific Classifiers

Barbara Caputo
NADA, CVAP
KTH SE-100 44 Stockholm
caputo@nada.kth.se

Henrik Christensen
NADA, CVAP/CAS
KTH SE-100 44 Stockholm
hic@nada.kth.se

Abstract

We present a new probabilistic classifier for object recognition that allows us to use separate feature vectors, selected specifically for each class. We obtain this result by extending previous work on Class Specific Classifiers and Spin Glass-Markov Random Fields. The resulting method, that we call Kernel-Class Specific Classifier, allows us to use a different kernel and a different reference hypothesis for each object class by learning them. We present promising experiments on object recognition in cluttered backgrounds, which show the power of our approach.

"From each according to his abilities, to each according to his needs".

Karl Marx, The Capital.

1 Introduction

Object recognition is a key topic in computer vision, and a consistent amount of research in this area has been devoted to the development of effective feature representations (see for instance [13, 17, 15, 4, 11] and many others). Thanks to these efforts, today it is possible to represent objects effectively by using color information [17, 11], global or local textural information [13, 15, 16], shape contour [4, 12], and so on.

Still, it is not clear which features should be used for a given task. Consider for example the objects in Figure 1. Color seems a good feature for tomatos and pears, but for artificial objects like cups and cars it is probably not very effective. For these objects shape contours appear more appropriate, but it is likely that apples and tomatos would get mixed with these features; and so on. The dilemma of how to choose the optimal features has been tackled mainly in two ways in the literature: (1) a very popular solution is to choose a single feature representation. This choice can be done with the help of some feature selection technique [18, 9], or (more often) on the basis of some prior knowledge or assumption on the nature of the task [17, 13, 15]; (2) another possibility is using multiple features, whether combined via a voting scheme [10], whether via a probabilistic model [6], whether defining a new feature representation that integrates visual information traditionally represented by different set of features [11]. Whatever strategy one chooses, it will result in the same feature(s) for all the objects to be recognized.

This paper presents a completely different approach. We describe a probabilistic classifier

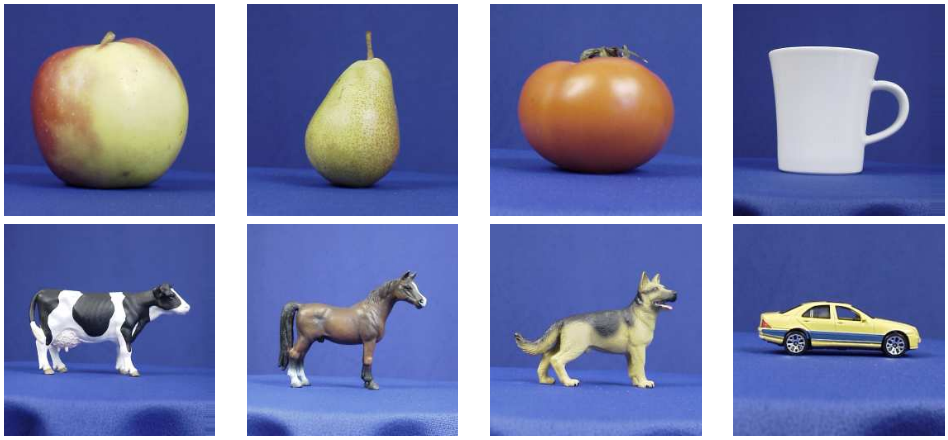

Figure 1: Objects from the CogVis-ETH database [10]. Which features should be used for optimal recognition?

that permits us to use different features for different objects, according to their specific characteristics. For example, in the case of the objects in Figure 1, it would allow us to use color features for tomatos, textural features for cows, shape contours for cups and so on: to each according to its needs. The algorithm we present consistutes an extension of previous work [5], and combines together Class Specific Classifiers (CSCs, [2, 3]) with Spin Glass-Markov Random Fields (SG-MRF, [6]). CSCs have been introduced recently, and they allow us the use of different features for different classes in a Bayes classifier. This is obtained by introducing a reference hypothesis class and using results of statistical theory [2, 3]. SG-MRF are a new class of MRF that use energy functions inspired by results of SG theory [1] via kernel functions (mainly Gaussian kernels, [14]) . Using SG-MRF in a CSC carries two main advantages:

- it does allow us to use the power of kernel functions for classification purposes, still leaving open the possibility to use different sets of features;
- the class of possible kernel functions is determined a priori by theoretical constraints. As the choice of kernels is limited, it does not become ungovernable. At the same time, it is wide enough to reasonably guarantee the possibility to tailor each kernel according to each class need.

For each class, kernel and reference hypothesis are *learned* from the training data; we call this method Kernel-Class Specific Classifier (K-CSC).

The paper is organized as follows: section 2 describes probabilistic appearance-based methods for object recognition and sets the notation for the rest of the paper. Then we introduce CSCs (section 3) and SG-MRFs (section 4). In section 5 we derive K-CSC, and section 6 presents experiments on object recognition in different backgrounds. A summary discussion concludes the paper.

2 Probabilistic Appearance-based Object Recognition

Probabilistic approaches to appearance-based object recognition considers the image views of a given object Ω_k as random vectors. Let $\boldsymbol{x} \equiv [x_{ij}], i = 1, \dots \mathcal{L}, j = 1, \dots \mathcal{M}$ be an $\mathcal{M} \times \mathcal{L}$ image. We will consider each image as a random feature vector $\boldsymbol{x} \in \Re^{\mathcal{ML}}$. Assume we have $\mathcal{K}$ different classes $\Omega_1, \Omega_2, \dots, \Omega_\mathcal{K}$ of objects, and that for each object is given a set of n_k data samples, $\omega_k = \{\boldsymbol{x}_1^k, \boldsymbol{x}_2^k, \ldots, \boldsymbol{x}_{n_k}^k\}, k = 1, \ldots \mathcal{K}$. We will assign each object to a pattern class $\Omega_1, \Omega_2, \ldots, \Omega_{\mathcal{K}}$ with a discrete mapping, associating a test image, showing one of the objects, to the pattern class the presented object corresponds to. How the object class Ω_k is represented, given a set of data samples ω_k, varies for different appearance-based approaches. Assuming that the data samples ω_k are a sufficient statistic for the pattern class Ω_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:

$$k^* = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} P_{\Omega_k}(\boldsymbol{x}) = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} P(\Omega_k|\boldsymbol{x}),$$

and, using Bayes rule,

$$k^* = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} P(\boldsymbol{x}|\Omega_k)P(\Omega_k). \tag{1}$$

where $P(\boldsymbol{x}|\Omega_k)$ are the Likelihood Functions (LFs) and $P(\Omega_k)$ are the prior probabilities of the classes. Assuming that the priors $P(\Omega_k)$ are constant and the same for all object classes, the Bayes classifier (1) simplifies to

$$k^* = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} P(\boldsymbol{x}|\Omega_k). \tag{2}$$

3 Class Specific Classifier

The LFs have to be learned from training data. An effective strategy consists in extracting a set of features (like Gaussian derivatives or color histograms [13], local descriptors [15] and so on) from the original images, and then learn the LFs on them. Let $\boldsymbol{z} = T(\boldsymbol{x})$ be such a set of features. The Bayesian classifier based on $\boldsymbol{z}$ is

$$k^* = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} P(\boldsymbol{z}|\Omega_k). \tag{3}$$

Thus, the features replace the raw data. The hidden implication in this procedure is that $\boldsymbol{z}$ is a sufficient statistics for the recognition problem:

$$P(\boldsymbol{x}|\Omega_k) = g(T(\boldsymbol{x})|\Omega_k)h(\boldsymbol{x}), k = 1, \ldots, \mathcal{K}.$$

This is the Neyman-Fisher factorization theorem [7]. The well known corollary of this theorem is that any likelihood ratio is invariant when written in terms of a sufficient statistic. Thus, if $\boldsymbol{z}$ is a sufficient statistic,

$$\frac{P(\boldsymbol{x}|\Omega_j)}{P(\boldsymbol{x}|\Omega_k)} = \frac{P(\boldsymbol{z}|\Omega_j)}{P(\boldsymbol{z}|\Omega_k)} \tag{4}$$

The $\mathcal{K}$-ary classifier (2) can be constructed by knowing only the likelihood ratios. Moreover, it is possible to use in the denominator an additional class, Ω_0:

$$k^* = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} \frac{P(\boldsymbol{x}|\Omega_k)}{P(\boldsymbol{x}|\Omega_0)}. \tag{5}$$

Combining together equations (4) -(5) leads to two main results:
Feature-based Class Specific Classifier [2]:

$$k^* = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} \frac{P(\boldsymbol{z}_k|\Omega_k)}{P(\boldsymbol{z}_k|\Omega_0)}, \tag{6}$$

where $\boldsymbol{z}_k = T_k(\boldsymbol{x}), 1 \leq k \leq \mathcal{K}$ are feature transformations that depend on the class being tested, thus they are *class specific* features;
PDF Projection Theorem [3]:

$$k^* = \operatorname*{argmax}_{k=1,\ldots\mathcal{K}} P(\boldsymbol{x}|\Omega_{0_k}) \frac{P(\boldsymbol{z}_k|\Omega_k)}{P(\boldsymbol{z}_k|\Omega_{0_k})}. \tag{7}$$

Equation (7) holds when the $\boldsymbol{z}_k$ are a sufficient statistics for the class Ω_k. When this is just approximately true $P(\boldsymbol{x}|\Omega_{0_k})$ and $P(\boldsymbol{z}_k|\Omega_{0_k})$ must be replaced by their approximations $\hat{P}(\boldsymbol{x}|\Omega_{0_k})$ and $\hat{P}(\boldsymbol{z}_k|\Omega_{0_k})$ [3].

We refer the reader to [2, 3] for a detailed derivation and discussion of the results listed above. Here we briefly point out three important issues for CSCs (6)-(7):

1. *The sufficiency requirement.* The reader can wonder how these results can be useful when the sufficient statistics are not known. The point is that, whenever one builds a Bayes classifier, it is never known whether sufficiency is achieved. It is worth stressing anyway that assuming
$$\frac{P(\boldsymbol{x}|\Omega_k)}{P(\boldsymbol{x}|\Omega_0)} = \frac{P(\boldsymbol{z}_k|\Omega_k)}{P(\boldsymbol{z}_k|\Omega_0)}, k = 1, \dots \mathcal{K}$$
is far milder than equation (4).
2. *The reference hypothesis.* There are two important issues concerning Ω_0. The first is that the denominator densities $P(\boldsymbol{z}_k|\Omega_0)$ in (6)-(7) need to be accurately evaluated in the tails, otherwise it could happen that for a given image $P(\boldsymbol{z}_k|\Omega_0) \to 0$ for all k and no reliable decision can be made. The second issue is that the implicit assumption in (6)-(7) is that Ω_0 is, loosely speaking, a member of each of the other object classes $\Omega_1, \dots, \Omega_{\mathcal{K}}$ (see [2, 3] for more details).
3. *The choice of the specific features.* CSCs give the possibility to use different features for different classes, but they actually don't say anything on how to choose these features.

4 Spin Glass-Markov Random Fields

There are many possible ways to estimate the pdfs for CSC. In this paper we use SG-MRFs; this choice will lead to the derivation of K-CSC. In this section we briefly describe SG-MRFs, and in section 5 we derive K-CSC. The interested reader can find more details on SG-MRFs in [6] and references therein.

The SG-MRF probability distribution is given by

$$P_{SG}(\boldsymbol{x}|\Omega_k) = \frac{1}{Z}\exp\left[-E_{SG}(\boldsymbol{x}|\Omega_k)\right], Z = \sum_{\{\boldsymbol{x}\}}\exp\left[-E_{SG}(\boldsymbol{x}|\Omega_k)\right], \qquad (8)$$

with

$$E_{SG} = -\sum_{\mu=1}^{p_k}\left[K(\boldsymbol{x}, \tilde{\boldsymbol{x}}^{(\mu)})\right]^2, \qquad (9)$$

where the function $K(\boldsymbol{x}, \tilde{\boldsymbol{x}}^{\mu})$ is a generalized Gaussian kernel [14]:

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

$\{\tilde{\boldsymbol{x}}^{\mu}\}_{\mu=1}^{p_k}, k \in [1, \mathcal{K}]$ are a set of vectors selected from the training data that we call *prototypes* [6]. The number of prototypes per class must be finite, and they must satisfy the condition $K(\tilde{\boldsymbol{x}}^i, \tilde{\boldsymbol{x}}^j) = 0$, for all $i, j = 1, \dots p_k, i \neq j$ and $k = 1, \dots \mathcal{K}$.

5 Kernel Class Specifi c Classifi er

In this section we show how the combination of CSC and SG-MRF leads to a new kernel classifier which fully uses the power of both ideas. First of all, a *kernel* is a function K

such that, for all $\boldsymbol{x}, \boldsymbol{y} \in X$,

$$K(\boldsymbol{x}, \boldsymbol{y}) = \Phi(\boldsymbol{x}) \cdot \Phi(\boldsymbol{y}),$$

where Φ is a mapping from X to an (inner product) feature space F [14]. Thus, the SG-MRF energy function can be rewritten as:

$$E_{SG} = -\sum_{\mu=1}^{p_k} [K(\boldsymbol{x}, \tilde{\boldsymbol{x}}^\mu)]^2 = -\sum_{\mu=1}^{p_k} \tilde{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}}^\mu) = -\sum_{\mu=1}^{p_k} \Phi(\boldsymbol{x}) \cdot \Phi(\tilde{\boldsymbol{x}}^\mu),$$

where $\tilde{K}$ represents a new kernel function [14]. The SG-MRF probability distribution becomes

$$P_{SG}(\boldsymbol{x}|\Omega_k) = \frac{1}{Z} \exp\left[\sum_{\mu=1}^{p_k} \Phi(\boldsymbol{x}) \cdot \Phi(\tilde{\boldsymbol{x}}^\mu)\right]. \qquad (10)$$

Equation (10) tells that P_{SG} depends on $\boldsymbol{x}$ via a mapping $\Phi(\boldsymbol{x}) = \boldsymbol{z}$. Thus, we can use this probability in the CSC classifiers (6)-(7), identifying the feature extraction operator $T_k(\boldsymbol{x})$ with the mapping $\Phi_k(\boldsymbol{x})$, as to say using a different mapping, and thus a *different kernel* for each class. We get:

$$\frac{P_{SG}(\Phi_k(\boldsymbol{x})|\Omega_k)}{P_{SG}(\Phi_k(\boldsymbol{x})|\Omega_0)} = \frac{\frac{1}{Z}\exp\left[\sum_{\mu_k} \Phi_k(\boldsymbol{x}) \cdot \Phi_k(\tilde{\boldsymbol{x}}^{\mu_k})\right]}{\frac{1}{Z}\exp\left[\sum_{\mu_0} \Phi_k(\boldsymbol{x}) \cdot \Phi_k(\tilde{\boldsymbol{x}}^{\mu_0})\right]}$$

where $\{\tilde{\boldsymbol{x}}^{\mu_k}\}, \mu_k = 1, \dots p_k$ are the set of prototypes of class Ω_k; $\{\tilde{\boldsymbol{x}}^{\mu_0}\}, \mu_0 = 1, \dots p_0$ are the set of prototypes of class Ω_0. As the mapping Φ_k is the same for the numerator and denominator, the constant Z is the same for both terms, thus it simplifies (see [6] and references therein). It follows:

$$\frac{P_{SG}(\Phi_k(\boldsymbol{x})|\Omega_k)}{P_{SG}(\Phi_k(\boldsymbol{x})|\Omega_0)} = \exp\left[\sum_{\mu_k} \Phi_k(\boldsymbol{x}) \cdot \Phi_k(\tilde{\boldsymbol{x}}^{\mu_k}) - \sum_{\mu_0} \Phi_k(\boldsymbol{x}) \cdot \Phi_k(\tilde{\boldsymbol{x}}^{\mu_0})\right].$$

So, the feature-based CSC (6) combined with SG-MRF becomes [5]:

$$k^* = \operatorname*{argmin}_k \left[-\sum_{\mu_k} [K_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}^{\mu_k})]^2 + \sum_{\mu_0} [K_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}^{\mu_0})]^2\right]. \qquad (11)$$

It is also possible to extend this result to the CSC (7), thus having the chance to use a class specific reference hypothesis. We obtain:

$$k^* = \operatorname*{argmax}_{k=1,\dots\mathcal{K}} P(\boldsymbol{x}|\Omega_{0_k}) \frac{P(\boldsymbol{z}_k|\Omega_k)}{P(\boldsymbol{z}_k|\Omega_{0_k})} = \operatorname*{argmax}_{k=1,\dots\mathcal{K}} \frac{1}{Z} \exp\left[\sum_{\mu=1}^{p_k} [K(\boldsymbol{x}, \tilde{\boldsymbol{x}}^\mu]^2\right] \cdot$$

$$\exp\left[\sum_{\mu_k} \Phi_k(\boldsymbol{x}) \cdot \Phi_k(\tilde{\boldsymbol{x}}^{\mu_k}) - \sum_{\mu_0} \Phi_k(\boldsymbol{x}) \cdot \Phi_k(\tilde{\boldsymbol{x}}^{\mu_0})\right].$$

Note that the first kernel does **not** have any index: here we are not exploiting the property of Mercer kernels of being scalar products with respect to a mapping Φ_k. Doing this, we impose that the implicit mapping in the first exponential is the same for all the object classes Ω_k. As a consequence, Z will be the same as well [6]. We get:

$$k^* = \operatorname*{argmin}_{k=1,\dots\mathcal{K}} \left[-\left[\sum_{\mu=1}^{p_k} [K(\boldsymbol{x}, \tilde{\boldsymbol{x}}^\mu)]^2 + \sum_{\mu_k} [K_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}^{\mu_k})]^2\right] + \sum_{\mu_0} [K_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}^{\mu_0})]^2\right]. \qquad (12)$$

We call the classifiers (11)-(12) Kernel-Class Specific Classifiers (K-CSCs). For both of them, the kernels K_k can be *learned*, for each class Ω_k, with a leave-one-out technique.

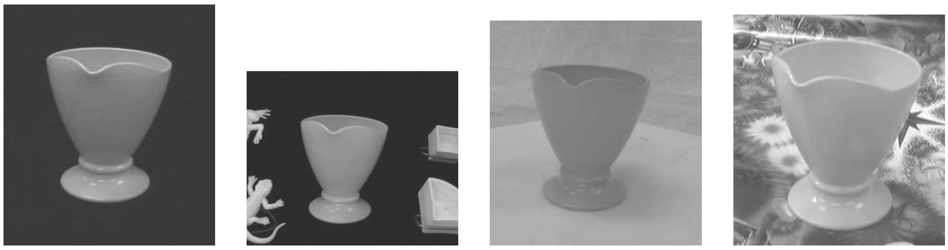

Figure 2: The cup from the Nelson6 database [12] in (from left to right) homogeneous background, with distracting objects, in white background and in textured background.

Thus, K-CSCs permit us to use for each class a different representation according to its needs, as CSC does. But as the K-CSC representation is bound to be a specific class of kernels, it solves the ambiguity of CSC regarding the choice of the representations and permits to learn them [1].

The structure of K-CSC (12) suggests us a method for choosing the reference hypothesis: we propose to take as $\Omega_{0_{k^*}}$, for the object class Ω_{k^*}, all the remaining object classes $\{\Omega_k\}, k = 1, \ldots \mathcal{K}, k \neq k^*$. Thus, the reference hypothesis is not chosen heuristically, but it is *learned* from the training data. This strategy addresses the issues regarding the choice of Ω_0 discussed in section 3. Experiments reported in section 6 show that it is also very effective from the experimental point of view. We conclude this section with a final consideration: the reader could wonder whether the $\boldsymbol{z}_k = \Phi_k(\boldsymbol{x})$ are a sufficient statistics for the class Ω_k, as required by CSCs. It could be argued that it is not, as the mapping Φ_k is a mapping in a higher dimensional space. The point is that, although Φ_k maps the data into a higher dimensional space, it can be proved that the mapped data are embedded in a subspace of the mapped space F , which will be of dimension lower or equal to the dimension of the data set [14]. Thus, if $\boldsymbol{x}$ is a sufficient statistic for the class Ω_k, so it will be $\Phi_k(\boldsymbol{x})$.

6 Experiments

We ran two series of experiments on a database of 6 objects [12]: a cup, a fighter, a plane, a car, a toy rabbit and a toy bear (see Figure 2). The database originally consisted of a training set of 106 views (53 for the car) taken on a sphere (hemisphere for the car), approximately every 20 degrees, in black homogeneous background. There are four different test sets: one taken in black homogeneous background, one in black homogeneous background with distracting objects around, one on a white marble table and one on a poster with Christmas decorations (Figure 2). Each of these sets has 53 (24 for the car) views, positioned in between the training views. The test views are taken at the same scale of those in the training set, but the illumination conditions change from background to background.

In the first series of experiments, we used the training and test set described above. As images are of different sizes, we first represented each view with Gaussian derivative histograms [13] using Gaussian derivative filters along x and y, variance $\sigma = 1.0$ and resolution for histogram axis of 16 bins. ($D_x D_y$). From the point of view of specific features, this is equivalent to do, for each image, $\boldsymbol{x} \to D_x D_y(\boldsymbol{x}) \to \Phi(D_x D_y(\boldsymbol{x}))$. We performed classification using (a) K-CSC with a common reference hypothesis ($\mathrm{K-CSC}_c$, equation (11)), consisting in the homogeneous background of the training set; (b) K-CSC with

[1]This procedure leaves open the chance of using a different kernel and a different set of features for each class. In that case, the problem of how to select features would still be open.

specific reference hypothesis ($\mathrm{K}-\mathrm{CSC}_s$, equation (12)); (c) SG-MRF. For all these three classifiers, kernel (one for SG-MRF, 6 for both K-CSCs) were learned from the training data with a leave-one-out technique. We also ran experiments using a nearest neighbor classifier with χ^2 and $\cap$ similarity measure; they both have been shown to be very effective with Gaussian derivatives histograms. Results are reported in Table 1.

	homo	**heter1**	**heter2**	**heter3**
χ^2	98.86 %	81.06 %	33.33 %	34.47 %
$\cap$	90.53 %	57.95 %	28.03 %	28.79 %
SG-MRF	99.24 %	80.68 %	32.95 %	35.98 %
$\mathrm{K}-\mathrm{CSC}_c$	99.30 %	80.56 %	37.15 %	40.61 %
$\mathrm{K}-\mathrm{CSC}_s$	**99.65 %**	**87.16 %**	**55.21 %**	**49.30 %**

Table 1: Results for the first series of experiments. Training was done on object views taken in homogeneous background. Test was done on object views taken in four different backgrounds.

We see that for all experiments, $\mathrm{K}-\mathrm{CSC}_s$ using a specific reference hypothesis gives the best performance. In three cases out of four, K-CSC with common reference hypothesis ($\mathrm{K}-\mathrm{CSC}_c$) obtains the second best performance. The only case in which this doesn't happen is for the test set in white heterogeneous background. A possible explanation is that this test set presents less contrast between the objects and the background than all the others. Thus, the choice of homogeneous background as reference hypothesis can be not optimal.

Table 1 shows a strong improvement in performance between SG-MRF and K-CSC, for all test sets. This is an important result, because in both cases we are using the same method for evaluating the pdfs. Thus, the better performance depends by the CSC approach. Nevertheless, the performance for images in heterogeneous background is generally poor. This mostly depends from the fact that we are using as starting data $\boldsymbol{x}$ a feature representation (Gaussian derivative histograms) which is global, and it is well known that global representations suffer for changes of background with respect to the training set. But it is also known that SG-MRFs are quite robust to degradation in the test set, provided that a reasonable amount of degradation is introduced as well in the training set (see [6] and references therein) [2]. Thus, we ran a second series of experiment using as training set the old training set plus 1/4 of views taken from all the previous test sets. Views were chosen randomly, and test sets consisted of the remaining views. Results are reported in Table 2, and confirm the previous results as well the robustness properties of SG-MRF. We can conclude that K-CSCs provides an effective method for object recognition.

	homo	**heter1**	**heter2**	**heter3**
χ^2	98.86 %	91.67 %	74.24 %	85.60 %
$\cap$	91.29 %	74.62 %	63.64 %	75.76 %
SG-MRF	99.24 %	91.29 %	85.98 %	85.98 %
$\mathrm{K}-\mathrm{CSC}_c$	99.30 %	90.28 %	88.78 %	89.29 %
$\mathrm{K}-\mathrm{CSC}_s$	**99.65 %**	**93.40 %**	**91.29 %**	**91.67 %**

Table 2: Results for the second series of experiments. Training was done on object views taken in homogeneous and heterogeneous background. Test was done on object views taken in four different backgrounds. Training and test set are disjoint.

[2] This is similar in spirit to the idea behind Virtual Support Vector Machines, [8].

7 Summary

We presented Kernel-Class Specific Classifiers, a new method for appearance-based object recognition that allows us to use class specific recognition strategies. This is obtained by combining together two existing methods, Class Specific Classifiers [2] and Spin Glass-Markov Random Fields [6]. We report promising experimental results, showing the power of our idea. In the future we plan to apply this method to object categorization, to integrate in K-CSC results of SG-MRFs (such as [6]) and to investigate the classifier performance with increasing number of objects.

Acknowledgments

We thank www.jmlg.org that inspired the title. This work has been supported by the EU project " Cognitive Vision Systems" -IST-2000-29375 CogVis.

References

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

[2] P. M. Baggenstoss, H. Niemann, "A theoretically optimal probabilistic classifier using class-specific features", *ICPR00*, pp 767-772.

[3] P. M. Baggenstoss, "Class-specific features in classification", *IEEE TRans. SP*, 3428 -3432, 1999.

[4] S. Belongie, J. Malik, J. Puchiza, "Matching Shapes", *Proc ICCV'01*.

[5] B. Caputo, H. Niemann, "to each according to its needs: kernel class specific classifiers", *Proc of ICPR02*.

[6] B. Caputo, G. Dorko, "How to combine color and shape information for object recognition: kernels do the trick", *Proc of NIPS02*.

[7] D. R. Cox, D. V. Hinkley, *Theoretical Statistics*. Chapman and Hall, London, 1974.

[8] D. Decoste, B. Schoelkopf, "Training invariant support vector machines", *Machine Learning*, 46, 161-190, 2002.

[9] D. Koller, M. Sahami, "Toward optimal feature selection", *Proc ICML96*.

[10] B. Leibe, B. Schiele, "Analyzing appearance and contour based methods for object categorization", *Proc CVPR'03*.

[11] J. Matas, R. Marik, J. Kittler, "On representation and matching of multi-coloured objects", *Proc ICCV05*.

[12] R. Nelson and A. Selinger, "A cubist approach to object recognition", *ICCV98*, pp 614-621.

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

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

[15] C. Schmid, R. Mohr, C. Bauckhage, "Evaluation of Interest Point Detectors", *IJCV* 37(2): 151-172 (2000)

[16] H. Schneiderman , T. Kanade, "A statistical method of 3d object detection applied to faces and cars", *Proc CVPR'00*.

[17] M. J. Swain, D. H. Ballard, "Color Indexing", IJCV, 11-32, 1991.

[18] S. C. Zhu, Y. N. Wu and D.B. Mumford, "FRAME: Filters, Random field And Maximum Entropy: Towards a Unified Theory for Texture Modeling", *IJCV*, 1998.