Kernel Methods for Melanoma Recognition

Elisabetta LA TORRE[a,1], Tatiana TOMMASI[a], Barbara CAPUTO[b],
Giovanni E. GIGANTE[a]
[a] *University of Rome La Sapienza, Physics Department, Italy*
[b] *NADA/CVAP, KTH, Stockholm, Sweden*

Abstract. Skin cancer is a spreading disease in the western world. Early detection and treatment are crucial for improving the patient survival rate. In this paper we present two algorithms for computer assisted diagnosis of melanomas. The first is the support vector machines algorithm, a state-of-the-art large margin classifier, which has shown remarkable performances on object recognition and categorization problems. The second method, spin glass-Markov random fields, combines results of statistical physics of spin glasses with Markov random fields. We compared the two approaches using color histograms as features. We benchmarked our methods with another algorithm presented in the literature, which uses a sophisticated segmentation technique and a set of features especially designed for melanoma recognition. To our knowledge, this algorithm represents the state of the art on skin lesions classification. We show with extensive experiments that the support vector machines approach outperforms the existing method and, on two classes out of three, it achieves performances comparable to those obtained by expert clinicians.
Keywords: Melanoma Recognition, Computer Assisted Diagnosis, Support Vector Machines, Kernel Methods.

1. Introduction

Malignant melanoma is a significant public health problem. Its incidence is rising faster than that of any other cancer in the US and in Europe [1, 2]. Early detection and treatment are critical and result in improved patient survival rates. The most used diagnostic technique is called Epiluminescence Microscopy (ELM). It is a non-invasive technique that allows for a detailed surface analysis of a suspicious skin lesion by using hand-held device emitting incident light from a light source penetrating the epidermal skin layer. The diagnosis of early melanoma is based on simple observation of images by dermatologists, who commonly use the ABCD (Asymmetry, Border, Color and Dimension) method as clinical guide. So, the diagnosis depends heavily on the physician's level of expertise.

An automatic system for melanoma recognition would constitute a valuable support for physicians in every day clinical practice and should reproduce the perceptual and cognitive strategy followed by doctors. The last years have witnessed numerous research on this topic; a key factor for the development and evaluation of these systems is the availability of a statistically significant database. To our knowledge the state of the art in melanoma recognition was presented by H. Ganster et

[1] Corresponding Author: Elisabetta La Torre, Università La Sapienza, p.le A. Moro 5, 00185, Rome, Italy (elisabetta.latorre@uniroma1.it).

al. [3]. That paper presented a large database, accompanied by: (a) a segmentation algorithm for isolating the potential melanoma from the surrounding skin [3]; (b) a set of features containing shape and radiometric features as well as local and global parameters, calculated to describe the malignancy of a lesion, from which significant features are selected by application of statistical feature subset selection methods [3]; (c) a nearest neighbor classification algorithm [3]. In that work the authors concentrated particularly on the segmentation techinque and the features selection process. Here we focus instead on the classification algorithm, and we propose the use of kernel methods for recognition of skin lesion images. Specifically, we focus our attention on two algorithms: Support Vector Machines (SVM) [4] and Spin Glass-Markov Random Fields (SG-MRF) [5]. SVM is a state-of-the-art large margin classifier, where the optimal separating surface is defined by a linear combination of scalar products between the view to be classified and some support vectors [6, 4]. SG-MRF is a fully connected MRF which integrates results of statistical mechanics with Gibbs probability distributions via non linear kernel mapping [5]. The database on which we will run our experiments is the same introduced by Ganster et al. [3]; our classification algorithms use binary masks determined by the segmentation algorithm developed in [3], and color histogram features. The choice of color histograms as feature types reproduces one of the criteria followed by dermatologists for diagnosis. We performed several series of experiments for selecting an optimal feature descriptor and we replicated the experimental setup used in [3] for a benchmark evaluation. Our results show that SVM obtains remarkably better performances than SG-MRF and Ganster's classification method. Furthermore, on two classes out of three, SVM achieves recognition results comparable to those obtained by skilled clinicians.

In summary, the contributions of this paper are: (1) The introduction of kernel methods for melanoma recognition, via two approaches: a probabilistic method, and a well known state-of-the-art classifier. For this second algorithm particularly, we studied in depth the classification performances with different kernel types. (2) The benchmark with a method presented in the literature [3], which to the best of our knowledge represents the state of the art in this field: on the same database and using the same segmentation masks, we had an improvement of more than 20% of the experimental results. Moreover our results are very stable and reliable because are obtained as mean value on five different partitions.

The rest of the paper is organized as follows: Section 2 describes some basic knowledge on SG-MRF theory and section 3 briefly explains the SVM algorithm. Section 4 reports on the experiments performed. The paper concludes with a summary discussion and some possible directions for future research.

2. Spin Glass - Markov Random Fields

In this section we describe the probabilistic method which constitutes one of the kernel methods proposed here for classification. This technique was introduced first for 3D object recognition [5], and was then applied to microcalcification detection with promising results [7].

Consider n visual classes Ω_j, $j = \{1, \ldots n\}$, and a set of k observations $\{x_j^1 \ldots x_j^k\}$, $x \in \Re^m$, random samples from the underlying, unknown, probability distribution $P(x)$ defined on $\Re^m$. Given an observation $\hat{x}$, our goal is to classify $\hat{x}$ as a sample from Ω_{j^*}

one of the Ω_j visual classes. Using a Maximum A Posteriori (MAP) criterion and Bayes rule we have $j = \text{argmax}_j P(\Omega_j / \hat{x}) = \text{argmax}_j P(\hat{x} / \Omega_j)$, where $P(\hat{x} / \Omega_j)$ are the Likelihood Functions (LFs) and $P(\Omega_j)$ are the prior probabilities of the classes which are assumed to be constant.

Spin Glass-Markov Random Fields (SG-MRFs) [5] are a new class of MRFs which connect SG-like energy functions (mainly the Hopfield one [8]) 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 kernels in a probabilistic framework. The SG-MRF probability distribution is given by $P_{SG\text{-}MRF}(x / \Omega_j) = (1/Z) \exp[-E_{SG\text{-}MRF}(x / \Omega_j)]$, $Z = \Sigma_x \exp[-E_{SG\text{-}MRF}(x / \Omega_j)]$ with $E_{SG\text{-}MRF} = -\sum_{\mu=1}^{p_j} [K(x,\tilde{x}^{(\mu)})]^2$, where the function $K(x,\tilde{x}^{(\mu)})$ is a generalized Gaussian kernel [6] $K(x,y) = \exp\{-\rho\, d_{a,b}(x,y)\}$, $d_{a,b}(x,y) = \Sigma_i |x_i^a - y_i^a|^b$ and $\{\tilde{x}^{(\mu)}\}_{\mu=1\ldots p_j}, j \in [1, n]$ are a set of vectors selected (according to a chosen ansatz [5]) from the training data that we call *prototypes*. The number of prototypes per class must be finite, and they must satisfy the condition $K(\tilde{x}^i, \tilde{x}^k) = 0$, for all $i, k = 1,\ldots p_j$, $i \neq j$ and $j = 0,\ldots n$ (the interested reader can find a detailed discussion regarding SG-MRF in [5]). Thus using SG-MRF modelling, the Bayes classifier (1) will become $j^* = \text{argmin}_j E_{SG\text{-}MRF}(\hat{x} / \Omega_j)$.

3. Support Vector Machines

In this section we briefly describe SVM in the two class case. For further details and the extension to multiclass settings we refer the reader to [4].

Consider the feature vector $x \in \Re^N$ and its class label $y \in \{-1, +1\}$. Let (x_1, y_1), (x_2, y_2), ... , (x_m, y_m) denote a given set of m training examples. If we assume that the two classes are linearly separable, there exists a linear function $f(x) = w \cdot x + b$ such that for each training example x_i, it yields $f(x_i) \geq 0$ for $y_i = +1$ and $f(x_i) \leq 0$ for $y_i = -1$. The optimal separating hyperplane is the one which has maximum distance to the closest points in the training set. Mathematically this hyperplane can be found by solving a constrained minimization problem using Lagrange multipliers α_i ($i = 1, \ldots, m$). It results in a classification function $f(x) = \text{sgn}(\sum_{i=1}^{i=m} \alpha_i y_i\, w \cdot x + b)$, where α_i and b are found by using an SVC learning algorithm [4]. It turns out that a small number of the $\alpha_i s$ are different from zero; their corresponding data x_i are called support vectors.

SVM can be extended to nonlinear problems by using a nonlinear operator $\Phi(\cdot)$ to map the input feature vectors x_i from the original $\Re^N$ into a higher dimensional feature space $\mathcal{H}$ by $x \rightarrow \Phi(x) \in \mathcal{H}$. Here the mapped data points of the two classes become linearly separable. Assuming there exists a kernel function K associated with the inner product of the desired nonlinear mapping such that $K(x,y) = \Phi(x) \cdot \Phi(y)$, then a non linear SVM can be obtained by replacing $x \cdot y$ by the kernel $K(x,y)$ in the decision function, obtaining then $f(x) = \text{sgn}(\sum_{i=1}^{i=m} \alpha_i y_i\, K(x_i, x) + b)$. This corresponds to constructing an optimal separating hyperplane in the feature space.

In this paper we consider four kernel types :

1. Polynomial kernel ("poly") $K(x, y) = (\gamma * x \cdot y)^d$
2. Generalized Gaussian kernel ("gengauss") $K(x, y) = \exp\{-\gamma * |x^a - y^a|^b\}$

3. Gaussian kernel ("gauss") $K(x, y) = \exp\{-\gamma * |x - y|^2\}$
4. Chi-squared kernel ("chi") $K(x, y) = \exp\{-\gamma * \chi^2(x, y)\}$.

4. Experiments

In this section we present experiments that show the effectiveness of kernel methods for melanoma recognition. To this purpose, in a preliminary step, we ran a first series of experiments for feature selection. Then we used the selected features for an extensive set of classification experiments. In the rest of the section we describe the database used, the experimental setup and our experimental findings.

Database: We performed our experiments on the database created by the Department of Dermatology of the Vienna General Hospital [3]. The whole database consists of 5380 skin lesion images, divided into three classes (these numbers are not perfectly coincident with those reported in [3], where the database is said to be of 5363 images, but this difference should not affect the comparison between the two algorithms): 4277 of these lesions are classified as clearly benign lesions (Class 1), 1002 are classified as dysplastic lesions (Class 2) and 101 lesions are classified as malignant melanomas (Class 3). The lesions of the classes 2 and 3 were all surgically excised and the ground truth was generated by means of histological diagnosis [3]. In order to have statistically significant results, we ran experiments with five different partitions, then we calculated the mean and the standard deviation of the obtained recognition rates. This procedure has been adopted for all the experiments which are reported in this paper.

Experimental Setup: The three key components for an automated melanoma recognition algorithm are: segmentation, features extraction and classification. We describe below the general approach followed in this paper for each of these steps:

Segmentation: We used the segmentation method developed by Ganster et al. [3] on this database. It consists of a binary mask determined by several segmentation algorithms combined together with a fusion strategy. This choice allows for a fair comparison between Ganster's technique and ours.

Feature Extraction: In the ABCD-rule of dermoscopy, the color distribution in the skin lesion is one of the discriminant features for clinical melanoma recognition, thus we used color histograms as features. The color histogram was computed by discretizing the colors within the image and counting the number of pixels for each color. We performed several experiments for selecting the best features, namely using hue, rg, RG, RB and GB color histograms. The resolution of the bin axes was varied for each representation, consisting of 8, 16, 32, 64 (for bidimensional histograms we chose the resolution of each axis with the same bin value). We found that the GB representation obtained the best results for all the bin values, thus we used it in all the following experiments.

Classification: We used SG-MRF and SVM algorithms (see section 2 and 3 respectively). For SG-MRF we learned the kernel parameters during the training stage using a leave-one-out strategy [5]. For SVM we used the four kernel types described in section 3. The kernel parameters were chosen via cross validation.

Classification Experiments: All the experiments were performed respecting the procedure reported by Ganster et al. [3]. The training set consisted of 270 images (90 for each class); the test set consisted of the whole database [3]. Note that training and test set are not disjoint; once again we underline that this follows the procedure

proposed in [3] which allows for benchmarking. We used the GB features and we ran experiments for 8, 16, 32 and 64 resolution of bins per axes, with five different partitions for training and test set, using SG-MRF and SVM with four different kernel types. Table 1 reports, for SG-MRF and SVM, the recognition rates for each class averaged on five partitions. We also report the average of the recognition rate obtained class by class ("Mean Class"), and the overall recognition rate ("Overall"). For sake of clarity we also report the results obtained in [3]; note that these results were obtained on a single run.

Table 1. Recognition results for the classification experiments on three classes of lesions obtained from Ganster et al. [3] and with SG-MRF and SVM methods with different kernels. We report the recognition rates for the three classes, the overall and the mean recognition rates. Results obtained with SG-MRF and SVM are mean values from five different runs with their standard deviations. Class 1, Class 2, Class 3 identify the benign, dysplastic and malignant lesions respectively.

	Ganster et al. [3] (%)	SG-MRF (%)	SVM (%)			
			poly	gauss	gengauss	chi
Class 1	59	48.6 ± 4.2	80.1 ± 13.0	71.9 ± 11.1	96.2 ± 4.0	68.6 ± 17.7
Class 2	53	38.8 ± 3.4	15.7 ±13.7	24.8 ± 12.7	11.0 ± 1.8	22.4 ± 7.5
Class 3	73	94.1 ± 3.4	29.5 ± 20.4	45.0 ± 28.5	89.5 ± 0.9	62.6 ± 19.7
Mean Class	61	60.5 ± 17.0	41.8 ± 19.6	47.2 ± 13.6	65.6 ± 27.4	51.2 ± 14.5
Overall	58	47.7 ± 2.9	67.1 ± 7.8	62.6 ± 6.2	80.2 ± 2.8	59.9 ± 12.9

A first comment is that SVM, with the generalized Gaussian kernel, obtains the best result with respect to Ganster's method and SG-MRF. The overall recognition rate is of 80.2% to be compared with a 58% obtained by Ganster and 47.7% obtained by SG-MRF. This proves the effectiveness of this technique for melanoma recognition. A second comment is that SVM performance varies considerably as the overall recognition rate goes from a minimum of 59.9% for the chi-squared kernel to a maximum of 80.2% for generalized Gaussian kernel. It is also interesting to note that, for the overall recognition rate, the kernels which obtains the worst performances tend to have the highest standard deviations, while the kernel with the best performance has the smallest one. This illustrates the importance of doing kernel selection in the training phase; the low standard deviation of the SVM's best result also shows the stability of our findings. A final remark should be made on the poor performance of SG-MRF. This might be due to the dimension of the training set for each class; it might be possible that the probabilistic method needs a higher statistic in order to estimate properly the energy function.

Table 2 reports the confusion matrix for SVM with generalized Gaussian kernel and the confusion matrices obtained by Ganster and the one obtained by dermatologists, both reported in [3]. We see that for class 1 and class 3 SVM outperforms Ganster's method and is comparable with the dermatologists' performances. It is very interesting to note that, in contrast, SVM performs poorly on class 2, which corresponds to dysplastic lesions. This might be explained considering that here we are using only color information, while Ganster used a selection of different features and dermatologists used the ABCD rule. It is thus possible that color information only is not discriminant enough in order to recognize correctly dysplastic

lesions, while it seems to be effective for separating benign and malignant lesions. In the future we will explore this issue by testing different types of informations.

Table 2. Confusion matrices for different classification methods: (a) Confusion matrix for the SVM results with the "gengauss" kernel. The number of images reported are mean value of the number obtained from 5 different partitions; (b) Confusion matrix obtained with the Ganster's method [3]; (c) Confusion matrix obtained from clinical diagnosis, performed from expert dermatologists of the Department of Dermatology at the Vienna General Hospital [3]. In the tables 1, 2, 3 identify the three classes corresponding to benign, dysplastic and malignant lesions respectively.

(a)	Assigned		
True	1	2	3
1	4112.6	112.6	50.8
2	874.8	110.0	17.2
3	10.4	0.2	90.4

(b)	Assigned		
True	1	2	3
1	2500	1347	410
2	324	531	155
3	14	12	70

(c)	Assigned		
True	1	2	3
1	4161	94	9
2	42	960	8
3	6	19	78

5. Conclusions

In this paper we proposed the use of kernel methods for melanoma recognition, with two approaches: SG-MRF and SVM. For this second algorithm particularly, we studied in depth the classification performances with different kernel types. The experiments showed that SVM, with the generalized Gaussian kernel, obtains an improvement of more than 20% with respect to the results presented in [3], which to the best of our knowledge represents the state of the art of the field. Moreover, on two classes out of three, SVM achieves recognition results comparable to those obtained by skilled clinicians. In the future we will conduct similar experiments with different descriptors, such as gray-level textural features and shape descriptors, in order to test the effectiveness of different types of information and to eventually reproduce the ABCD method followed by the dermatologists in every day clinical practice.

References

[1] D. S. Rigel and J. A. Carucci, Malignant Melanoma: Prevention, Early Detection, and Treatment in the 21st Century, CA Cancer J Clin 2000; 50:215-236.

[2] E. De Vries, F. I. Bray, J. W. W. Coebergh and D. M. Parkin, Changing Epidemiology of Malignant Cutaneous Melanoma in Europe 1953-1997: Rising Trends in Incidence and Mortality but Recent Stabilizations in Western Europe and Decreases in Scandinavia, Int. J. Cancer 107: 119-126; 2003 Wiley-Liss, Inc..

[3] H. Ganster, A. Pinz, R. Rhrer, E. Wildling, M. Binder and H. Kittler, Automated Melanoma Recognition. IEEE Transactions on Medical Imaging, vol. 20, n. 3, march 2001.

[4] V. Vapnik. Statistical learning theory. Wiley and Son, 1998.

[5] B. Caputo,A new kernel method for object recognition: spin glass Markov random fields, PhD thesis, Stockholm, November 2004. Available at http://www.nada.kth.se/~caputo

[6] B. Scholkopf, A. J. Smola, Learning with kernels, 2001, the MIT Press.

[7] B. Caputo, E. La Torre, S. Bouattour, G.E. Gigante, A New Kernel Method for Microcalcification Detection: Spin Glass- Markov Random Fields, Proc. of MIE02, Budapest, August 2002.

[8] D. J. Amit, Modeling Brain Function, Cambridge University Press, Cambridge, USA, 1989.