2D Shape Recognition Using Information Theoretic Kernels

Manuele Bicego[1,2], André F. T. Martins[3], Vittorio Murino[1,2],
Pedro M. Q. Aguiar[4], Mário A. T. Figueiredo[3]

[1] *Computer Science Department, University of Verona - Verona, Italy*
[2] *Istituto Italiano di Tecnologia (IIT) - Genova, Italy*
[3] *Instituto de Telecomunicações, Instituto Superior Técnico, Lisboa, Portugal*
[4] *Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisboa, Portugal*

Abstract

In this paper, a novel approach for contour-based 2D shape recognition is proposed, using a recently introduced class of information theoretic kernels. This kind of kernels, based on a non-extensive generalization of the classical Shannon information theory, are defined on probability measures. In the proposed approach, chain code representations are first extracted from the contours; then n-gram statistics are computed and used as input to the information theoretic kernels. We tested different versions of such kernels, using support vector machine and nearest neighbor classifiers. An experimental evaluation on the chicken pieces dataset shows that the proposed approach outperforms the current state-of-the-art methods.

1. Introduction

Object recognition is undoubtely an important and still open research area in computer vision and pattern recognition. The classification of three-dimensional (3D) objects has been addressed using different approaches [6, 12], many of which are based on the analysis of two-dimensional (2D) aspects of objects, namely, 2D shapes. Many recognition tasks are addressed using only features of the boundary or the 2D shape. In this context, many contour representations have been proposed, like Fourier descriptors and chain code [12]. Actually, object contours have shown to be very expressive in many contexts, and they have been often exploited in several approaches proposed in the past, exhibiting different characteristics: robustness to noise and occlusions, invariance to translation, rotation, and scale, computational requirements, and accuracy [12, 18].

In this paper, a novel method for contour-based 2D shape recognition is proposed, using a class of information theoretic kernels recently introduced [13]. This type of kernels, based on a non-extensive generalization of the classical Shannon information theory, are defined on (possibly unnormalized) probability measures. In [13], these kernels were used in text categorization tasks, by being specifically applied to different types of multinomial representations of the texts: relative term frequencies (also known as bags of words) and n-gram statistics (relative frequencies of subsequences of n symbols).

In the approach herein presented, we begin by extracting chain code representations of the contours, which yields sequences of symbols. From these sequences, n-gram statistics are then computed, which are used as input to the information theoretic kernels.

The proposed approach has been tested on the *chicken pieces* dataset [1], a challenging testbed composed by silhouettes of different chicken parts (wings, backs, drumsticks, thighs, and breasts). We tested different information theoretic kernels, ranging from the classic Jensen-Shannon divergence kernel, to different versions of the recently introduced Jensen-Tsallis kernels [13]. These kernels are used in support vector machine and nearest neighbor classifiers. A study of the performance of these classifiers, as a function of the kernel parameters, has been carried out, as well as an evaluation of fully automatic versions, where the parameters are automatically computed from the training set. The results presented in Sect. 3 show that the classifiers based on the Jensen-Tsallis kernels significantly outperform the current state-of-the-art methods. Before presenting such results, the method characteristics are detailed in Sect. 2, and final remarks are summarized in Sect. 4.

2. The proposed method

In this section, the proposed methodology is presented. In particular, after describing how to obtain the probability measure from the contours, the information theoretic kernels are described.

2.1. From contours to probability measures

To extract a probability measure from a contour, we have different options, e.g., using hidden Markov models (HMM) [3]. Here, we adopt a simpler technique, based on statistics of n-grams extracted from chain code representations of the contours. For a given n, the n-gram statistics correspond to a multinomial distribution. This choice opens the door to the use of information theoretic kernels, which are defined on pairs of multinomial distributions. Although other, more complex models, such as HMM, could be considered, the corresponding kernels can not be computed in closed form. Because the information theoretic kernels defined in [13] can also be applied on unnormalized measures, they can be used either directly with the raw n-gram counts or with the corresponding normalized versions (multinomials).

2.2. Information theoretic kernels

Kernels on probability measures have been shown very effective in classification problems involving text, images, and other types of data [9, 8, 10, 11]. Given two probability measures p_1 and p_2, representing two objects, the following information theoretic kernels can be defined (for more details, see [13]):

- Jensen-Shannon kernel,

$$k^{\text{JS}}(p_1, p_2) = \ln(2) - JS(p_1, p_2), \qquad (1)$$

with $JS(p_1, p_2)$ being the Jensen-Shannon divergence

$$JS(p_1, p_2) = H\left(\frac{p_1 + p_2}{2}\right) - \frac{H(p_1) + H(p_2)}{2}, \qquad (2)$$

where $H(p)$ is the usual Shannon entropy.

- Jensen-Tsallis kernel,

$$k_q^{\text{JT}}(p_1, p_2) = \ln_q(2) - T_q(p_1, p_2), \qquad (3)$$

where $\ln_q(x) = (x^{1-p} - 1)/(1 - q)$ is the q-logarithm and $T_q(p_1, p_2)$ is the Jensen-Tsallis q-difference, defined as:

$$T_q(p_1, p_2) = S_q\left(\frac{p_1 + p_2}{2}\right) - \frac{S_q(p_1) + S_q(p_2)}{2^q}, \qquad (4)$$

and $S_q(r)$ is the Jensen-Tsallis entropy, defined, for a multinomial $r = (r_1, r_2, ..., r_L)$, with $r_i \geq 0$ and $\sum_i r_i = 1$, as

$$S_q(r_1, r_2, ..., r_L) = \frac{1}{q-1}\left(1 - \sum_{i=1}^{L} r_i^q\right).$$

q represents the free parameter of the kernel, which can be chosen for example by cross validation.

In [13], versions of these kernels applicable to unnormalized measures were also defined. Let $\mu_1 = \omega_1 p_1$ and $\mu_2 = \omega_2 p_2$ be two unnormalized measures, where p_1 and p_2 are the normalized counterparts (probability measures) and ω_1 and ω_2 arbitrary positive real numbers (weights). The weighted versions of the Jensen-Tsallis kernel are defined as follows:

- Weighted Jensen-Tsallis kernel (version 1),

$$k_q^{(1)}(\mu_1, \mu_2) = S_q(\pi) - T_q^{\pi}(p_1, p_2), \qquad (5)$$

where $\pi = (\pi_1, \pi_2) = \left(\frac{\omega_1}{\omega_1+\omega_2}, \frac{\omega_2}{\omega_1+\omega_2}\right)$ and

$$\begin{aligned} T_q^{\pi}(p_1, p_2) &= S_q\left(\pi_1 p_1 + \pi_2 p_2\right) \\ &\quad - \left(\pi_1^q S_q(p_1) + \pi_2^q S_q(p_2)\right). \end{aligned}$$

- Weighted Jensen-Tsallis kernel (version 2),

$$k_q^{(2)}(\mu_1, \mu_2) = \left(S_q(\pi) - T_q^{\pi}(p_1, p_2)\right)(\omega_1 + \omega_2)^q. \qquad (6)$$

It was shown in [13] that $k_q^{(1)}$ is a positive definite kernel for $q \in [0, 1]$, while $k_q^{(2)}$ is a positive definite kernel for $q \in [0, 2]$.

3. Experimental Results

We test the proposed approach on the publicly available chicken pieces dataset[1] [1]. This dataset contains 446 binary images (silhouettes) of chicken pieces, each belonging to one of five classes representing specific chicken parts: wing (117 samples), back (76), drumstick (96), thigh and back (61), and breast (96) – some examples may be found in Fig. 1. This constitutes a challenging classification task, which has been studied by several authors [1], [5], [15].

From binary silhouettes, contour descriptions are extracted: notice that these contour descriptions completely specify the underlying shape. These contours are then encoded using the (8 directions) chain code, leading to a sequence of symbols (on an alphabet of 8

[1] http://algoval.essex.ac.uk:8080/data/sequence/chicken/.

Figure 1. Examples from the chicken pieces dataset (rows 1 to 5: wings, backs, drumsticks, thighs, and breasts).

symbols) for each contour. From the sequence of symbols corresponding to each shape contour, we compute statistics of n-grams (for $n = 2$, $n = 3$, and $n = 4$), that is, we count how many times each possible subsequence of length n exists in each contour. This can be seen as a bag-of-words representations of the sequences, where the dictionary contains all the 8^n possible length-n sequences/"words". In summary, each shape is represented by a 8^n-dimensional vector of non-negative numbers (which may or may not be normalized to unit sum) that will serve as arguments for the kernels defined in the previous section subsequently used by a kernel-based classifier.

We consider two types of classifiers: support vector machines (SVMs) and K-nearest neighbors (K-NN). Parameter C of the SVM learning algorithm is optimized by 10-fold cross validation (CV). For the K-NN classifiers, we consider both the simplest version 1-NN as well as K chosen by 10-fold CV. Concerning parameter q of the Jensen-Tsallis kernels, we report results both with the best performing value and with the value selected by 10-fold CV. Finally, we found that the best choices of n were $n = 4$, in the case of the SVM, and $n = 3$ for the K-NN classifiers.

Table 1 reports the average accuracy results, obtained using 10 repetitions of holdout CV. The superiority of the information theoretic kernel over the linear kernel is evident, in particular for the K-NN classifiers.

Figure 2 plots the accuracies of the SVM on 4-grams, for different kernels, as a function of parameter q. The plot also shows the accuracy obtained with q chosen by cross-validation. In line with the results from [13], the best results are obtained with $q < 1$. Although we do not have, at this moment, a formal justification for this fact, it may be due to the following behavior of the Jensen-Tsallis kernels. For $q < 1$, the maximizer of

kernel	SVM	1-NN	K-NN
Linear	0.827(0.009)	0.305(0.014)	0.355(0.005)
k^{JS}	0.886(0.005)	0.775(0.008)	0.757(0.005)
k_q^{JT} (auto q)	0.882(0.009)	0.781(0.008)	0.783(0.008)
k_q^{JT} (best q)	0.890(0.042)	0.791(0.037)	0.805(0.038)
$k_q^{(1)}$ (auto q)	0.886(0.010)	0.636(0.006)	0.685(0.012)
$k_q^{(1)}$ (best q)	0.891(0.042)	0.636(0.030)	0.682(0.032)
$k_q^{(2)}$ (auto q)	0.884(0.006)	0.791(0.011)	0.787(0.008)
$k_q^{(2)}$ (best q)	0.895(0.042)	0.801(0.038)	0.814(0.039)

Table 1. Average accuracy rate, with standard deviation in parenthesis.

$k_q^{\mathrm{JT}}(p, v)$ with respect to p is not v, but another distribution closer to uniform. This is not the case for the Jensen-Shannon kernel k^{JS}, which coincides with J_1^{JT}, for which the minimizer of $k^{\mathrm{JS}}(p, v)$ with respect to p is precisely v. This behavior of k_q^{JT} plays the role of a regularizer (favoring uniform distributions) on the n-gram statistics.

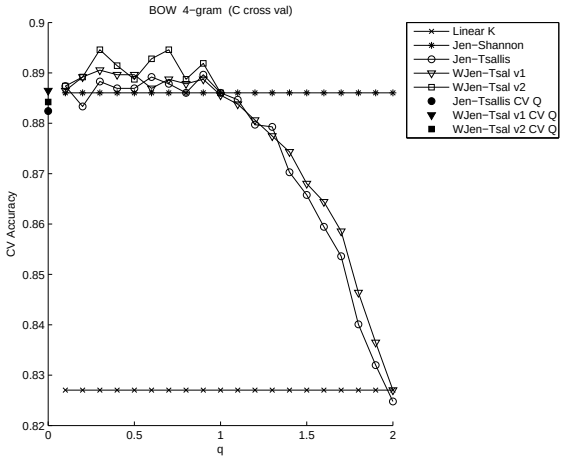

Figure 2. Accuracies of different kernels as a function of q.

Table 2 compares our best results against other published results on the same dataset [2], [4], [5], [7], [14], [15], [16], [17]. Although the experimental procedure is not the same in all those references, the results suggest that the proposed method performs better than the others.

4. Conclusions

In this paper, a novel approach for contour-based 2D shape recognition has been proposed. In the presented

Methodology	Accuracy	Ref.
K-NN Classifiers		
1-NN + Levenshtein edit distance	≈ 0.67	[14]
1-NN + approximated cyclic distance	≈ 0.78	[14]
K-NN + cyclic string edit distance	0.743	[15]
1-NN + mBm-based features	0.765	[5]
1-NN + HMM-based distance	0.738	[5]
Our best K-NN ($k_q^{(2)}$, 3-gram)	**0.814**	
SVM Classifiers		
Edit distance-based kernel	0.811	[15]
HMM-based entropic features	0.812	[16]
HMM + Fisher Kernel	0.817	[17]
HMM + Top Kernel	0.808	[17]
HMM + FESS-embedding + RBF	0.830	[17]
HMM + Trans embedding + RBF	0.811	[4]
HMM + Marginalized kernel	0.775	[7]
HMM + Cluster-based Fisher kernel	0.858	[2]
HMM + Non linear Marginalized kernel	0.855	[7]
Our best SVM ($k_q^{(2)}$, 4-gram)	**0.895**	

Table 2. Comparative results with other methods.

method, chain code representations are extracted from the contours and n-gram statistics are then computed. These statistics are used as argument for non-extensive information theoretic kernels, based on which kernel-based classifiers are built. We tested different versions of such kernels, using support vector machine and nearest neighbor classifiers. An experimental evaluation on the *chicken pieces* dataset showed that the proposed approach outperforms the current state-of-the-art methods.

5. Acknowledgements

We acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (Contract 213250) and from Fundação para a Ciência e Tecnologia (FCT) (grant PTDC/EEA-TEL/72572/2006).

References

[1] G. Andreu, A. Crespo, and J. Valiente. Selecting the toroidal self-organizing feature maps (TSOFM) best organized to object recognition. In *Proc. of IEEE ICNN97*, volume 2, pages 1341–1346, 1997.

[2] M. Bicego, M. Cristani, V. Murino, E. Pekalska, and R. Duin. Clustering-based construction of hidden Markov models for generative kernels. In D. Cremers, Y. Boykov, A. Blake, and F. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 5681 of *LNCS*, pages 466–479. Springer, 2009.

[3] M. Bicego and V. Murino. Investigating Hidden Markov Models' capabilities in 2D shape classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence - PAMI*, 26(2):281–286, 2004.

[4] M. Bicego, E. Pekalska, D. Tax, and R. Duin. Component-based discriminative classification for hidden markov models. *Pattern Recognition*, in press, 2009.

[5] M. Bicego and A. Trudda. 2D shape classification using multifractional Brownian motion. In *Proc. of Joint Int. Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, volume LNCS 5342, pages 906–916. Springer, 2008.

[6] R. Campbell and P. Flynn. A survey of free-form object representation and recognition techniques. *Computer Vision and Image Understanding*, 81:166–210, 2001.

[7] A. Carli, M. Bicego, S. Baldo, and V. Murino. Non-linear generative embeddings for kernels on latent variable models. In *Proc. ICCV09 Workshop on Subspace Methods*, 2009.

[8] M. Cuturi, K. Fukumizu, and J.-P. Vert. Semigroup kernels on measures. *Journal of Machine Learning Research*, 6:1169–1198, 2005.

[9] M. Hein and O. Bousquet. Hilbertian metrics and positive definite kernels on probability measures. In Z. Ghahramani and R. Cowell, editors, *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics (AISTATS)*. 2005.

[10] T. Jebara, R. Kondor, and A. Howard. Probability product kernels. *Journal of Machine Learning Research*, 5:819–844, 2004.

[11] J. Lafferty and G. Lebanon. Diffusion kernels on statistical manifolds. *Journal of Machine Learning Research*, 6:129–163, 2005.

[12] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1998.

[13] A. Martins, N. Smith, E. Xing, P. Aguiar, and M. Figueiredo. Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research*, 10:935–975, 2009.

[14] R. Mollineda, E. Vidal, and F. Casacuberta. Cyclic sequence alignments: Approximate versus optimal techniques. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 16(3):291–299, 2002.

[15] M. Neuhaus and H. Bunke. Edit distance-based kernel functions for structural pattern classification. *Pattern Recognition*, 39:1852–1863, 2006.

[16] A. Perina, M. Cristani, U. Castellani, and V. Murino. A new generative feature set based on entropy distance for discriminative classification. In *Proc. Int. Conf. on Image Analysis and Processing*, pages 199–208, 2009.

[17] A. Perina, M. Cristani, U. Castellani, V. Murino, and N.Jojic. A hybrid generative/discriminative classification framework based on free-energy terms. In *Proc. Int. Conf. on Computer Vision*, 2009.

[18] P. Suetens, P. Fua, and A. Hanson. Computational strategies for object recognition. *ACM Compuiting Surveys*, 24(1):5–61, 1992.