Pattern Recognition Letters 65 (2015) 8–14

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Dominant local binary patterns for texture classification: Labelled or unlabelled?☆

Francesco Bianconi [a,*], Elena González [b], Antonio Fernández [b]

[a] Università degli Studi di Perugia, Department of Engineering, Via G. Duranti 93, Perugia, 06125, Italy
[b] Universidade de Vigo, Department of Engineering Design, Vigo, 36310, Spain

ARTICLE INFO

Article history:
Received 17 October 2014
Available online 9 July 2015

Keywords:
Texture classification
Feature selection
Dominant local binary patterns

ABSTRACT

This paper investigates the problem of learning sets of discriminative patterns from local binary patterns (LBP). Such patterns are usually referred to as 'dominant local binary patterns' (DLBP). The strategies to obtain the dominant patterns may either keep knowledge of the patterns labels or discard it. It is the aim of this work to determine which is the best option. To this end the paper studies the effectiveness of different strategies in terms of accuracy, data compression ratio and time complexity. The results show that DLBP provides a significant compression rate with only a slight accuracy decrease with respect to LBP, and that retaining information about the patterns' labels improves the discrimination capability of DLBP. Theoretical analysis of time complexity revealed that the gain/loss provided by DLBP vs. LBP depends on the classification strategy: we show that, asymptotically, there is in principle no advantage when classification is based on computationally-cheap methods (such as nearest neighbour and nearest mean classifiers), because in this case determining the dominant patterns is computationally more expensive than classifying using the whole feature vector; by contrast, pattern selection can be beneficial with more complex classifiers such as support vector machines.

1. Introduction

LBP is a very popular approach to texture analysis with applications in a wide range of areas such as, among others, surface inspection, face recognition, biometrics and medical image analysis [2]. The method is much appreciated for its many desirable properties, such as ease of implementation, invariance to illumination changes, limited computational demand and high descriptive performance – especially when the level of noise is low [14]. LBP considers, as image features, the occurrence probability of the binary patterns that can be generated from an image patch of predefined shape and size when thresholded at the value of the central pixel. It is well known that the resulting probability distribution tends to be highly uneven: some patterns tend to occur much more frequently than others [24]. Many researchers have been concerned with the problem of reducing the dimensionality of LBP by determining the subsets of patterns that convey the largest amount of information. A common approach consists of reducing the number of features by using some *a priori* rules [20]: Ojala et al. for instance proposed to cluster patterns into rotationally-equivalent classes, an approach which generates the well-known family of rotation invariant descriptors (LBP^{ri}). They also suggested that further reduction could be obtained by considering the so called 'uniform patterns' (LBP^{riu2}), namely those patterns that have at most two bitwise transitions [24]. Experiments have shown that uniform patterns are the most common in natural images [24], a finding which was later on explained on a theoretical basis [1].

As an alternative, Liao et al. [17] and, more recently, Nanni et al. [22] and Guo et al. [12], proposed *a posteriori* strategies in which the patterns to retain are learnt from some training data. Liao et al. [17] for instance suggested to retain, as features, the probability of occurrence of the smallest set of patterns that, in any given image, represent a certain percent – 80% in their implementation – of the total population. The resulting dominant local binary patterns (DLBP) bear no information about the patterns' labels [17]; instead, they consider the relative patterns' frequency only. As a consequence this scheme does not guarantee that the *i*th element of the feature vector extracted from an image $\mathbf{I}_1$ and the *i*th element of the feature vector extracted from an image $\mathbf{I}_2$ refer to the same pattern. For this reason we refer to such selection strategy as an *unlabelled* model. A natural question arises whether comparing the probability of occurrence of different patterns makes sense altogether [6]. In [17] the authors affirm that omitting the pattern type information is not harmful; in

☆ This paper has been recommended for acceptance by A. Heyden.

* Corresponding author. Tel.: +39 75 5853703; fax: +39 75 5853703.
E-mail address: bianco@ieee.org (F. Bianconi).

http://dx.doi.org/10.1016/j.patrec.2015.06.025

this paper we endorse a diametric opposite view: that neglecting information about the patterns type has negative effects on the discrimination capability of the method. Our main claim is that feature selection schemes that keep knowledge of the patterns' type outperform the unlabelled approach. We refer to such reduction schemes as *labelled* methods.

In the remainder of the paper, after recalling the basics of LBP in Section 2, we discuss the unlabelled (Section 3) and labelled (Section 4) approach for determining dominant local binary patterns and perform an experimental comparison in Section 5. The results presented in Section 6 show that in no case the unlabelled model is superior to the labelled counterparts. We also evaluate the compression ratio that can be obtained with the various methods and study the effect of the different feature reduction schemes on the overall computing time. Section 7 concludes the paper with some final considerations.

2. Brief overview of LBP

The LBP operator characterizes images through the probability of occurrence of certain binary patterns that a neighbourhood of predefined shape and size can generate [24]. The typical configuration consists of a central pixel plus a set of peripheral points evenly spaced along a circle (see Fig. 1) – but other arrangements have been proposed as well [21]. The intensity values of those points that do not coincide with image pixels are estimated through interpolation. Such neighbourhoods are conventionally indicated in the form (m, R), where m represents the number of peripheral points and R the radius of the circle.

For each position of the neighbourhood a corresponding binary pattern is obtained by thresholding the intensity values of the peripheral points at the value of the central pixel. Each binary patterns is then assigned a unique label in the following way:

$$\mathrm{LBP}_{m,R} = \sum_{i=0}^{m-1} 2^i \xi(I_i - I_c) \tag{1}$$

where ξ is the binary thresholding function (Eq. 2).

$$\xi(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{2}$$

As a result, the $\mathrm{LBP}_{m,R}$ operator produces 2^m different binary patterns. Theoretically, when the input image rotates by angular steps of $\pm 2\pi/m$ radians, the binary sequence $\{\xi(I_i - I_c)\}$, $i \in \{1, \dots, m-1\}$ circularly shifts by one position to the left or to the right. To make the descriptor invariant against rotation, one can consider equivalent all the patterns that can be transformed into one another by a rotation of multiples of $\pm 2\pi/m$ radians. This approach gives rise to the rotation invariant operator, usually referred to as $\mathrm{LBP}^{ri}_{m,R}$. The number of rotationally-equivalent classes for a given m can be determined through group theory, as detailed in Ref. [8]. Table 1 shows the number of features generated by the $\mathrm{LBP}_{m,R}$ and $\mathrm{LBP}^{ri}_{m,R}$ operators for different values of m. Clearly the dimension of the descriptors grows quickly as m increases.

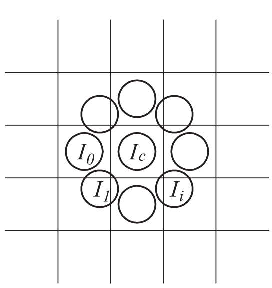

Fig. 1. Circular neighbourhood.

Table 1
Number of local binary patterns.

m	N	
	$\mathrm{LBP}_{m,R}$	$\mathrm{LBP}^{ri}_{m,R}$
4	16	6
8	256	36
16	65,536	4116
24	16,777,216	699,252

High dimensional data are in general difficult to handle due to the 'curse of dimensionality' [7]. Moreover, both experimental and theoretical studies have suggested that the probability of occurrence of local binary patterns may vary greatly from one pattern to another, and that certain patterns very seldom occur in practice [1,24]. As a result, some of them are likely to produce only noisy and irrelevant features that may mislead the classification [12]. The problem of determining the set of 'most discriminative' patterns is therefore a very actual and interesting one both from a theoretical and practical standopint.

3. Dominant local binary patterns: the unlabelled model (DLBP)

As we mentioned in Section 1, the unlabelled approach discards any information about the patterns' labels. The method consists of sorting the LBP histogram of each image in descending order and retaining a certain number of bins. Given a set of train images, the number of bins to retain is computed by determining, for each train image, the cardinality of the smallest set of patterns that accounts for a given fraction of the total occurrence probability and by averaging this value over the whole train set. Each histogram is sorted independently of the others in this scheme, therefore any information about the patterns' type is lost: the resulting DLBP features will only contain information about the patterns' frequencies. This strategy is based on the assumption that it is the relative probability distribution what really matters, not the occurrence probability of each specific pattern [17]. As for the fraction of the total occurrence to retain, throughout the paper we maintain the settings proposed in the above-cited reference, where the authors recommend the value 0.8. From a computational standpoint, the algorithm is dominated by the ordering of each vector of the train set, therefore executes in $\mathcal{O}(MN\log N)$ time, where M is the number of train patterns and N the dimension of the original descriptor.

4. Dominant local binary patterns: the labelled model

As opposed to the unlabelled model, the labelled model keeps knowledge of the patterns' labels. Different implementations of this approach have been proposed in the literature: we briefly recall them in the following subsections.

4.1. Labelled dominant local binary patterns (L-DLBP)

Labelled dominant local binary patterns have been described by Fu et al. [10] and, more recently, by González et al. [11]. In this scheme the original LBP histograms of the train images are first averaged column-wise (feature-by-feature) and the resulting vector (average patterns' frequencies) is sorted in descending order. Then the labels of the smallest set of co-occurrences that sum at least 0.8 are retained; the others are discarded. The labels this way obtained constitute the set of dominant patterns; the feature vector of any image is represented by the probabilities of occurrence of these patterns. From a