General framework for rotation invariant texture classification through co-occurrence of patterns

Elena González · Antonio Fernández · Francesco Bianconi

Abstract The use of co-occurrences of patterns in image analysis has been recently suggested as one of the possible strategies to improve on the bag-of-features model. The intrinsically high number of features of the method, however, is a potential limit to its widespread application. Its extension into rotation invariant versions also requires careful consideration. In this paper we present a general, rotation invariant framework for co-occurrences of patterns and investigate possible solutions to the dimensionality problem. Using Local Binary Patterns as bag-of-features model, we experimentally evaluate the potential advantages that co-occurrences can provide in comparison with bag-of-features. The results show that co-occurrences remarkably improve classification accuracy in some datasets, but in others the gain is negligible, or even negative. We found that this surprising outcome has an interesting explanation in terms of the degree of association between pairs of patterns in an image, and, in particular, that the higher the degree of association, the lower the gain provided by co-occurrences in comparison with bag-of-features.

Elena González* and Antonio Fernández
Universidade de Vigo
School of Industrial Engineering
Campus Universitario – 36310 Vigo (Spain)
Tel.: +34 986 818602
Fax: +34 986 812201
E-mail: elena,antfdez@uvigo.es
*Performed part of this work as a Visiting Professor in the Department of Industrial Engineering, Università degli Studi di Perugia, Italy

Francesco Bianconi
Università degli Studi di Perugia
Department of Engineering
Via G. Duranti, 67 – 06125 Perugia (Italy)
E-mail: bianco@ieee.org

1 Introduction

The bag-of-features model (BoF), in which images are represented through the probability distribution over a discrete vocabulary of local features, has proven effective in many image analysis tasks [8,47,29]. This representation bears no information about the spatial distribution of the local features. As a consequence, two images with the same distribution, but different spatial arrangement, are perfectly equivalent in this model, though they may be very unlike in fact – see Refs. [17,46] for examples. In order to overcome these somewhat paradoxical situations, researchers are actively seeking effective means to improve visual recognition by taking into account, in some way, information about the spatial interrelations of local features. Approaches so far proposed include spatial pyramids [20], spatial graphs [46], feature pooling [4], bag-of-strings [37] and co-occurrences of patterns [30,31]. For a review and a tentative categorization of these methods the interested reader may refer to the work of Wu *et al.* [46]. Theoretically, including structural information in image description should improve the discrimination capability.

This process, however, is by no means costless, since it inevitably produces a marked increase in the number of features through which images are described. Among the proposed strategies, co-occurrences of patterns are receiving increasing attention as a possible means to improve on the bag-of-feature model. Making a parallel to natural language processing, we can say that co-occurrences of patterns stand to patterns as bigrams stand to single words [42]. As we show in the following section, different implementations of this idea have already appeared in the literature. The transition from bag-of-features to co-occurrences, however,

is not straightforward, particularly when it comes to considering invariance against rotation. With regard to this, we believe that the theoretical implications have not been given the required attention and need further consideration. Another intrinsic difficulty is the high number of features that this strategy produces – N^2 in principle, where N is the number of features of the bag-of-features model. The higher dimensionality also pairs with a stronger computational demand, both in the feature extraction and classification stage. It is therefore of primary importance to assess the potential advantages and disadvantages of the co-occurrence model in comparison with the bag-of-features model. The contribution of this paper to the literature is in four points: first, we propose a general theoretical framework for rotation-invariant co-occurrence of patterns; second, we investigate possible solutions for reducing dimensionality; third, we carry out an experimental evaluation of the method on different datasets; fourth, we propose a model to explain the performance of co-occurrences and the extent of the gain they can provide against bag-of-features alone.

The remainder of the paper is organised as follows. After presenting a brief review of related literature (Sec. 2), we address the problem of formally defining rotation invariant co-occurrences of patterns (Sec. 3). In Sec. 4 we describe five strategies for dimensionality reduction. Experimental set-up, results and discussion are presented in Secs. 5–7, followed by some final considerations and directions for future research in Sec. 8.

2 Related research

Co-occurrences of local features are the extension of grey-level co-occurrence matrices to local features. They express the joint probability of pairs of patterns to occur in any two points of an image separated by a given displacement vector. The idea is general, and can be in theory applied to any image descriptors based on local features: Liu and Yang, for instance, based their model on 2×2 patterns computed from colour gradients [25]; Song, on 2×2 patterns computed from grey levels [39]; Lee *et al.*, on 2×2 motifs [21]; Kobayashi and Otsu, on Scale-invariant Image Features (SIFT) and Gradient Local Auto-Correlations (GLAC) [18]; Zou *et al.*, on quantized and rescaled intra- and inter-channel Gabor filter responses [48].

Among local texture features, Local Binary Patterns (LBP) have reached a prominent position, due to the ease of implementation, high discrimination capability and low computing demand. It is no surprise, then, that various researchers considered LBP as a possible basis for implementing co-occurrence strategies.

Here below we briefly review the approaches so far proposed. With reference to the rotation invariant properties of each method, we subdivide them into three groups: those that are not rotation invariant, those that are rotation invariant at the local feature level, and those that are rotation invariant at the co-occurrence level. In the round-up table (Tab. 1) the rotation invariance level of the three groups is indicated as *none*, *marginal* and *joint*, respectively. We also report, in the same table, the type of features used by each method: these can be the complete co-occurrences or some sets of global statistics from them extracted.

Nosaka *et al.* [30] define the co-occurrence of adjacent patterns separated by a given vector as the probability of occurrence of their combination in an image. If N is the number of possible patterns, the number of co-occurrences is N^2. The method has been later on extended into a rotation-invariant version [31] by considering equivalent all pairs of patterns that can be obtained from each other through a synchronous rotation of the two patterns which make up a pair. The authors also consider equivalent all pairs of patterns that can be obtained from each other through a reflection around an axis orthogonal to the displacement vector joining the two patterns, an operation that reduces the number of co-occurrences from N^2 to $N(N + 1)/2$. The methods is therefore invariant to two actions: a rotation and a reflection. To further reduce dimensionality, the authors split each binary pattern into sub-patterns of four pixels lying either on the vertical-horizontal directions (LBP$_+$) or on the two diagonal directions (LBP$_\times$), this way obtaining a feature vector of dimension $16(16 + 1)/2 = 136$.

Other authors dealt with the dimensionality problem by considering low-dimensional variations of local binary patterns. Sujatha *et al.* [40], for instance, as well as Li and Shi [22], used the reduced-dimension LBP variations LCLBP-OR$_{3\times3}$ and CLBP$_{3\times3}$, respectively. Sun *et al.* [41], along with Shadkam and Helfroush [38], employed rotation-invariant local binary patterns in the form of LBP$_{P,R}^{riu2}$. In this case the resulting descriptor is rotation invariant at the local feature level – but not at the co-occurrence level – as it will become clear in Sec. 3.1. The pairwise rotation invariant co-occurrence local binary patterns (PRI-CoLBP), recently proposed by Qi *et al.* [35] also consider co-occurrences of uniform local binary patterns, but, in addition, determine the intrinsic orientation of the co-occurrence through the image gradient in one of the two points, this way achieving invariance against rotation at the co-occurrence level. The magnitude of the gradient is then used to weight the co-occurrence. The potential problem with this approach, however, is that the gradient can be numerically

Table 1 Summary table of related methods.

Author(s)	Ref.	Year	LBP variants	Rotation invariance	Features
Ershad	[9]	2011	$LBP_{3\times3}$	None	Statistics
Nosaka *et al.*	[30]	2011	LBP_+; $LBP_\times$	None	Co-occurrences
Qi *et al.*	[35]	2012	$LBP_{8,1}^{ri}$	Joint	Co-occurrences
Shadkam and Helfroush	[38]	2012	$LBP_{8,1}^{ri}$; $LBP_{16,2}^{ri}$; $LBP_{24,3}^{ri}$	Marginal	Co-occurrences
Sujatha *et al.*	[40]	2012	$LCLBP\text{-}OR_{3\times3}$	None	Co-occurrences
Sun *et al.*	[41]	2012	$LBP_{8,1}^{ri}$; $LBP_{16,2}^{ri}$; $LBP_{24,3}^{ri}$	Marginal	Co-occurrences
Nosaka *et al.*	[31]	2012	LBP_+; $LBP_\times$	Joint	Co-occurrences
Wang *et al.*	[45]	2012	$LBP_{3\times3}$	None	Statistics
Li and Shi	[22]	2013	$CLBP_{3\times3}$	None	Co-occurrences

unstable in low-contrast regions, or even indeterminate in completely flat zones.

We finally mention that, instead of considering the complete set of co-occurrences as features, some authors (Wang *et al.* [45] and Ershad [9]) employed global statistical descriptors (i.e.: energy, contrast, homogeneity and entropy), which is the common practice when dealing with co-occurrences of grey-level intensities. But when it comes to co-occurrences of patterns, this solution should be considered with care, since there is in general no intrinsic order in the patterns' codes, as we discuss in Sec. 4.5.

In summary, the literature shows a fervid activity in the field, but the approaches thus far proposed lack, in our view, a formal conceptualisation of the problem in general terms. In the following section we therefore discuss a general rotation invariant framework for co-occurrences of patterns that can be applied to any bag-of-features model producing binary patterns.

3 Pairs of patterns and co-occurrences

Let p and q be the indices of two generic binary patterns generated by any bag-of-features operator, such as local binary patterns, improved local binary patterns and similar (see Ref. [11] for an up-to-date review of methods). For the sake of simplicity, let us assume that p, q are expressed as decimal codes, therefore $p, q \in \{0, \ldots, N-1\}$, where N is the number of binary pattern the descriptor generates. Now let (x_p, y_p) and (x_q, y_q) indicate the spatial positions of the two patterns within the image, with the convention that the origin of the reference system is the top-left pixel and that the y and x axes respectively point rightward and downward (see Fig. 1). Conventionally, we represent any oriented pair of patterns as $(p, q, \mathbf{v}_{pq})$, where $\mathbf{v}_{pq}$ is the displacement vector from (x_p, y_p) to (x_q, y_q).

Note that, in principle, $(p, q, \mathbf{v}_{pq}) \neq (p, q, \mathbf{v}_{qp})$, therefore oriented pairs of patterns are not symmetrical. The displacement vector can be decomposed into its
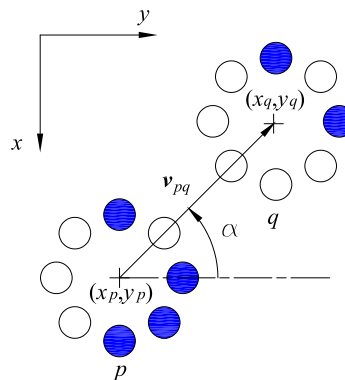

Fig. 1 An oriented pair of patterns.

length (d), orientation (α) and verse, and can be equivalently expressed as $\mathbf{v}_{\alpha,d}$, with the convention that α is measured positively counter-clockwise starting from the horizontal (y) direction (see Fig. 1).

For the sake of simplicity, we assume that neighbourhoods are either digital or interpolated circles (Fig. 2). In the first case, neighbourhoods are integral approximations of circles in the digital plane; in the second, the points lie on a circle and the value of those not coinciding with pixels' locations are determined through interpolation.
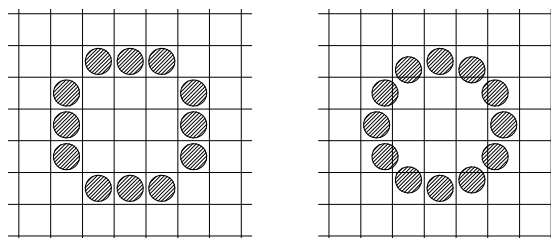

Fig. 2 Digital (left) and interpolated (right) circular neighbourhoods.

The neighbourhoods can be of two types (see Fig. 3): 'peripheral', which are made of peripheral pixels only

(so are, for instance, patterns generated by LBP), or 'full', which are composed of the peripheral pixels plus the central pixel (so are, for instance, patterns generated by ILBP).
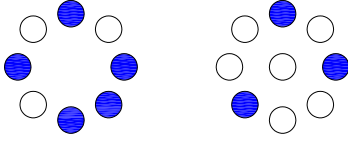

Fig. 3 Binary patterns of type 'peripheral' (left) and 'full' (right).

The number of discrete values that α can take depends on the number of pixels in the periphery of the neighbourhood. We define a co-occurrence of patterns in the following way:

Definition 1 A co-occurrence of patterns is an instance of an oriented pair of patterns.

3.1 Achieving rotation invariance

In order to achieve rotation invariance, let us establish, first, the concept of equivalence under rotation of two oriented pairs of patterns. Consider, to this end, the generic transform that maps an oriented pair of binary patterns into its rotated version by a discrete rotation angle $\theta_k = 2\pi k/n$, $k = \{0, \ldots, n-1\}$, where n is the number of pixels in the periphery of the neighbourhood. We can express such transform in the following way:

$$\varphi_{\theta_k} : (p, q, \mathbf{v}_{\alpha,d}) \rightarrow (p', q', \mathbf{v}_{\alpha',d}) \tag{1}$$

where:

$$\alpha' = (\alpha + \theta_k) \bmod 2\pi \tag{2}$$

$$p' = \{\xi(p_{\text{bin}}, k)\}_{\text{dec}} \qquad q' = \{\xi(q_{\text{bin}}, k)\}_{\text{dec}} \tag{3}$$

In the above equations subscripts 'bin' and 'dec' indicate conversion into binary and decimal format, respectively; $\xi(b, w)$ a circular shift by w position on the whole number b expressed in binary format. For instance, if $p = 25$, $n = 8$ and $k = 2$; we have:

$$p' = \{\xi(25_{\text{bin}}, 2)\}_{\text{dec}}$$
$$= \{\xi(00011001, 2)\}_{\text{dec}}$$
$$= \{01000110\}_{\text{dec}}$$
$$= 70$$

In practice, equations 1–3 describe, in mathematical form, what happens to a pair of patterns when the image it belongs to is rotated by an angle θ_k: as the joining vector rotates by θ_k, so do the two binary patterns that make up the pair (see Fig. 4). We can now define the concept of rotationally equivalent pairs of patterns:

Definition 2 Any two oriented pairs of patterns are rotationally equivalent if and only if there exists a transform of the type stated in Eqs. 1–3 that maps one pair into the other.

Any set of rotationally equivalent oriented pairs of patterns represents an *orbit* (see Ref. [33, p. 57]) of the transformation group defined by Eqs. 1–3. Any oriented pair of patterns belonging to an orbit is a *representative* of that orbit (see Fig. 4).
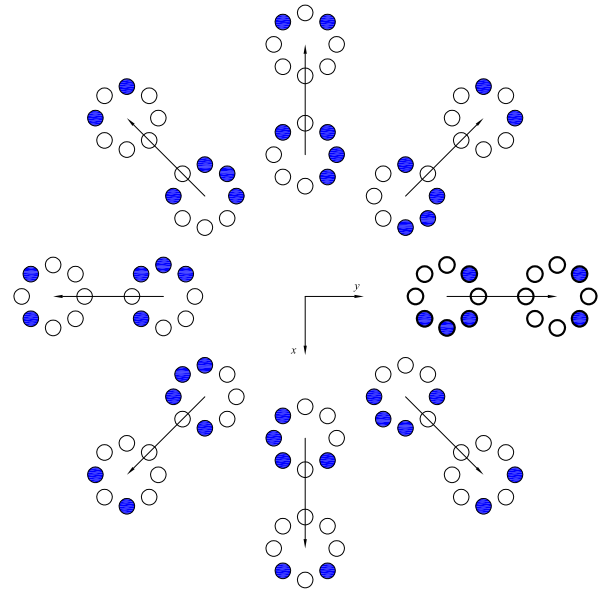

Fig. 4 An orbit of rotationally equivalent co-occurrences of patterns. The canonical form is depicted in bold line.

Algorithmically, in order to check if two pairs of patterns are rotationally equivalent, it is useful to establish a canonical orientation (reference) for the displacement vector $\mathbf{v}$. Herein we have chosen the y axis as the reference, so we can take, as the *canonical form* of a given pair of patterns, the representative of the orbit the pair belongs to that is collinear with the reference axis. Since two objects are equivalent, under a given transform, if and only if they have the same canonical form (see Ref. [33, p. 58-59]), in order to verify if two oriented pairs of patterns are rotationally equivalent it suffices to compare their canonical forms, as stated in the following lemma:

Lemma 1 *Any two pairs of patterns are rotationally equivalent if and only they have the same canonical form.*

As a consequence, the complete set of orbits can be identified with a complete list of canonical forms and vice versa. Elementary considerations let us easily figure out that the complete set of canonical forms coincides with all the possible (p, q) pairs, where $p, q \in \{0, \ldots, N-1\}$. Therefore, there are N^2 classes of rotationally invariant oriented pairs of patterns.

3.2 Computing co-occurence features

Computing rotation invariant co-occurrences of patterns means counting how many times each of the N^2 orbits comes about in an image. Operatively, our procedure works as follows. As a first step we compute the map of decimal codes that results from applying any bag-of-feature descriptor (i.e.: LBP, ILBP, BGC, etc.) to the input image. Then, for each $k \in \{0, \ldots, n-1\}$ we compute the co-occurrence matrix of each (p, q) pair at distance d and direction θ_k – let $\mathbf{M}_{d,\theta_k}$ indicate such matrix. Afterwards, we permute the elements of each $\mathbf{M}_{d,\theta_k}$ so that any (p, q) pairs in that matrix is reduced to its canonical form (note that, by definition, the permutation is an identity in the case $\mathbf{M}_{d,0°}$). Let us indicate any such matrices as $\hat{\mathbf{M}}_{d,\theta_k}$. The n matrices this way computed are averaged to give a probability distribution of the rotation invariant co-occurrence classes:

$$\bar{\mathbf{M}}_d = \frac{1}{n} \sum_{k=0}^{n-1} \hat{\mathbf{M}}_{d,\theta_k} \tag{4}$$

Finally, the result is serialized into a feature vector $\mathbf{f}$:

$$\mathbf{f} = \bar{\mathbf{M}}_d(:) \tag{5}$$

4 Strategies for dimensionality reduction

In the preceding section we have shown that rotation invariant co-occurrences of patterns generates N^2 features, where N is the dimension of the original bag-of-features model. Such number can be very large and may render the whole method impractical.

As we mentioned in Sec. 2, the most common approach to reduce dimensionality consists of using low-dimensional bag-of-features models. Nosaka *et al.* [31], for instance, split the LBP neighbourhood into sub-neighbourhoods of four pixels each, thus generating 16 patterns and $16 \times 16 = 256$ pairs. They further reduce this number to 136 by considering, along with rotation invariance, reflection invariance too. Shadkam and Helfroush [38] use rotation invariant uniform local binary patterns (LBP^{ri}) with neighbourhood radius one, two and three which give 10, 18 and 26 features, and 55,

71 and 351 co-occurrences, respectively. Here, too, the authors implicitly include reflection invariance.

Unfortunately, though the above solutions have been reported to be effective in a number of applications, they are not sufficiently general to be applied to the framework described in Sec. 3. In order to overcome the dimensionality problem, we considered five different strategies: the first four are *a posteriori* feature selection schemes based on the concept of *dominant patterns*; the fifth is a combination of the original features with global co-occurrence statistics.

4.1 Dominant co-occurrences

This approach derives from the dominant patterns scheme, proposed by Liao *et al.* [23] for reducing the dimensionality of local binary patterns. When applied to co-occurrences, the method simply consists of retaining, as features, the probabilities of occurrence of the pairs of patterns that, in any image, account for a given percent of the total co-occurrences. Following the settings proposed in the cited reference, we set this value to 80%. In practice, since the co-occurrence vector sums one by definition, for each image one has just to sort the feature vector (Eq. 5) in descending order and retain the smallest set of features that sum at least 0.8. We refer to the retained co-occurrences as the *dominant co-occurrences* (DC). It is important to remark that this is an *unlabelled* feature reduction scheme in which no information is retained about the co-occurrences' labels [23].

4.2 Labelled dominant co-occurrences

Discarding information about co-occurrences' labels may have negative effects on the discrimination capability of the method. A simple variation on the procedure described above enables retaining this important information. This solution, which we refer to as *labelled dominant co-occurrences* (LDC), is a feature selection scheme which learns the labels of the dominant co-occurrences from a set of training images. It works as follows. As a first step we compute the co-occurrence vectors of the training images. Then we average them, sort the resulting vector in descending order and retain the labels of the smallest set of co-occurrences that sum at least 0.8. Once we have established, with such a procedure, the set of dominant co-occurrences, the feature vector of any image is represented by the probabilities of the dominant co-occurrences. By definition this procedure generates as many features as DC.

4.3 Highest-variance co-occurrences

A simple variation on LDC consists of retaining the labels of the co-occurrences with the highest variance. This strategy has been suggested by Nanni *et al.* as a means for selecting the best performing rotation invariant patterns in local binary/ternary patterns [28]. To make their approach consistent with the others considered in this study, we implemented a modified version of the method described in the cited reference. We first compute the co-occurrence vectors of the training images, then calculate their variance, sort the resulting vector in descending order and finally retain the labels of the smallest set of co-occurrences that accounts for at least 80% of the total variance. We refer to this method as highest-variance co-occurrences (HVC).

4.4 Discriminative features

Another interesting learning framework for feature selection is represented by the so called *discriminative features* (DF in the remainder), recently introduced by Guo *et al.* [14]. The approach is very closely related to LDC, but the learning scheme is slightly different. The method determines, as a first step, a set of dominant co-occurrence labels for each class in the training set; then obtains the global set of dominant co-occurrence labels through the union of the labels of each class. This way it is ensured that any class in the training set is adequately represented within the selected co-occurrences' labels. Compared with the methods described in Secs. 4.1 and 4.2, the approach should in principle improve the representation capability. On the other hand, a potential drawback might be the higher number of retained features, which in principle tends to increase as the number of classes in the training set gets larger.

4.5 Combination of bag-of-features and global co-occurrence statistics

This strategy consists of combining suitable statistical descriptors computed from the co-occurrence distribution $\bar{\mathbf{M}}_d$ with the feature vector produced by the bag-of-features model. The idea is to indirectly capture some structural information while avoiding the burden of the whole co-occurrences. Since there is no intrinsic ordering in patterns, the statistics used to this purpose need to be invariant to permutations. Herein we used the following five:

Energy:
$$s_0 = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \bar{\mathbf{M}}_d(p,q) \tag{6}$$

Entropy:
$$s_1 = -\frac{1}{2\log_2(N)} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \bar{\mathbf{M}}_d(p,q) \log_2 \bar{\mathbf{M}}_d(p,q) \tag{7}$$

Normalized standard deviation:
$$s_2 = \sqrt{\frac{N^4}{(N^2-1)^3}} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \sqrt{\left[\bar{\mathbf{M}}_d(p,q) - \mu\right]^2} \tag{8}$$

Range:
$$s_3 = \max\left[\bar{\mathbf{M}}_d(p,q)\right] - \min\left[\bar{\mathbf{M}}_d(p,q)\right] \tag{9}$$

Trace:
$$s_4 = \sum_{p=0}^{N-1} \bar{\mathbf{M}}_d(p,p) \tag{10}$$

Another issue is the selection of a proper weighting scheme to combine the bag-of-features with the co-occurrence statistics. Since we used, in the experiments, a distance-based classifier (1-NN, L_2), we opted for a weighting strategy that guarantees that the maximum distance between any two points in the bag-of-feature space is the same as the maximum distance between any two points in the statistics' space. For a given image, let $\mathbf{b} = [b_0, \ldots, b_{N-1}]$ the corresponding bag-of-feature vector and $\mathbf{s} = [s_0, \ldots, s_4]$ the corresponding statistics. By definition, we have $0 \leq s_i \leq 1$, $\forall\, i \in \{0, \ldots, S-1\}$, and $0 \leq b_i \leq 1$, $\forall\, b \in \{0, \ldots, N-1\}$, but with the additional constraint that $\sum_{b=0}^{N-1} b_i = 1$. Under these constraints it is easy to see that the maximum distance between any two bag-of-feature vectors is $\sqrt{2}$, whereas the maximum distance between any two vectors of statistical descriptors is $\sqrt{S}$ ($S = 5$, in this case). We therefore apply a weighting factor $\beta = \sqrt{2/S}$ to the statistical descriptors to balance the contributions of the two parts. The resulting feature vector can be expressed in the following way:

$$\mathbf{f} = \mathbf{b} || \beta \mathbf{s} = [b_0, \ldots, b_{N-1}, \beta s_0, \ldots, \beta s_{S-1}] \tag{11}$$

5 Experiments

The main objective of the experiments is to assess the potential gain that we can obtain when switching from the rotation invariant bag-of-features model to the rotation invariant co-occurrence model. To this end we carried out a supervised image classification task based on six datasets.

5.1 Bag-of-features and co-occurrence models

As a test-bed for the experiments, we used the LBP operator in its simplest version. The baseline is therefore represented by the performance of $\mathrm{LBP}_{3\times3}^{ri}$ and $\mathrm{LBP}_{8,1}^{ri}$, bag-of-features models. Both operators produce 36 features [32]; the only one difference between the two is that the first uses a square neighbourhood, the second a circular (interpolated) one. The rotation invariant co-occurrence model is based on $\mathrm{LBP}_{3\times3}$. Since this operator produces 256 features, its co-occurrence version, which we indicate as $\mathrm{C}_d^{ri}(\mathrm{LBP}_{3\times3})$, returns $256 \times 256 = 65536$ features. We evaluated the performance of the full-dimensional co-occurrence model, as well as that of the different strategies for dimensionality reduction presented in Sec. 4. Co-occurrences have been computed at distance values $d = \{1, 2, 3\}$. For calibration purposes, we included the results obtained with RIC-LBP (Ref. [31]).

5.2 Datasets

The experimental test-bed comprehends six different datasets, each one containing hardware-rotated, stationary or nearly-stationary texture images. Rotation by hardware is required to avoid the artefacts and the consequently misleading results produced by software-rotated images (see Ref. [10] for a discussion on this topic). A summary list of their properties and the corresponding mosaics are reported in Tab. 2.

Dataset 1 contains a selection of 80 texture classes from the ALOT database, a project developed and maintained within the Intelligent Systems Lab at the University of Amsterdam, Holland [6, 1]. Dataset 2 is composed of 13 texture classes from the Brodatz's album. The hardware-rotated digital images have been acquired in our laboratory directly from the original reference [5]. Dataset 3 incorporates the entire Kylberg Sintorn Rotation dataset, provided by the Centre for Image Analysis at the Uppsala University, Sweden [19]. Dataset 4 includes the 12 granite classes of MondialMarmi (version 1.1), a granite image database for colour and texture analysis [3, 27]. Dataset 5 includes a selection of 45 texture classes from Outex as detailed in Ref. [2]. Finally, dataset 6 presents a set of 20 texture classes obtained from vectorial sources. To obtain rotated images in raster format, we preliminarily rotated the vectorial images, then raster-scanned the results.

5.3 Classification and accuracy estimation

We carried out supervised classification using two classification strategies: 1) nearest-neighbour (1-NN) rule with L_2 (Euclidean) distance and 2) Support Vector Machines (SVM). The 1-NN is well-suited for comparative purposes and has been used extensively for this task [44, 13, 7, 16], mainly on account of the absence of tuning parameters, ease of implementation and low computational demand. The last feature is particularly useful here, due to the high dimensionality of the methods involved. For the very same reason we opted for an SVM classifier with linear kernel. When number of features is large – as in the problem herein studied – one may not need to map data to a higher dimensional space and using the linear kernel is good enough (see Ref. [15] for a discussion on this topic). For fair play we used the same penalty parameter $C = 1$ for all classification problems. Our implementation of 1-NN and SVM is based on PRTOOLS [34], which in the case of SVM wraps around LIBSVM [24].

Accuracy estimation is based on split-half validation with stratified sampling. For each dataset we generate a set of 100 classification problems by splitting the dataset into two non-overlapping subsets, one for training and the other for validation. The stratified sampling constraint ensures that, for each class, half the samples are used for training and the other half for validation. For a given problem, the classification accuracy is estimated as the percentage of samples of the validation set classified correctly. To evaluate the effect of image rotation, we always train the classifier with non-rotated images ($\phi = 0°$) and test it using images rotated by ϕ_γ degrees, where γ is the index of one of the rotation angles available in the dataset (for dataset one, for instance, we have: $\phi_0 = 0°$, $\phi_1 = 60°$, $\phi_2 = 120°$ and $\phi_3 = 180°$ – see Tab. 2). The accuracy for the γ-th rotation angle can be expressed as follows:
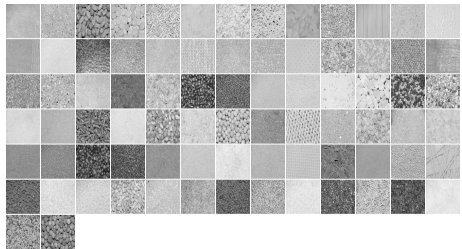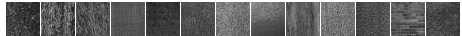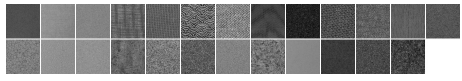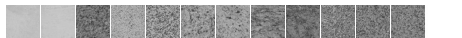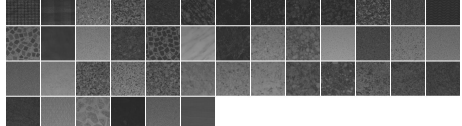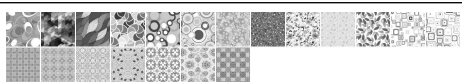
$$a_\gamma = \frac{1}{R} \sum_{r=1}^{R} \frac{W_{r,\gamma}}{V_{r,\gamma}} \tag{12}$$

where R is the total number of problems ($R = 100$, in this case); $V_{r,\gamma}$ and $W_{r,\gamma}$ respectively the number of validation samples and of correctly classified samples in the r-th problem. We also define the average accuracy (μ_a) and the standard deviation (σ_a) over the set of rotation angles in the following way:

$$\mu_a = \frac{1}{\Gamma} \sum_{\gamma=0}^{\Gamma-1} a_\gamma \tag{13}$$

$$\sigma_a = \frac{1}{\Gamma-1} \sum_{\gamma=0}^{\Gamma-1} \sqrt{(a_\gamma - \mu_a)^2} \tag{14}$$

Table 2 Summary list of the image datasets used in the experiments.

No.	Rotation angles	No. of classes	Samples / class	Image res.	Mosaic
1	0°, 60°, 120°, 180°	80	16	181 × 181	
2	0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, 90°	13	16	205 × 205	
3	0°, 40°, 80°, 120°, 160°, 200°, 240°, 280°, 320°	25	16	512 × 512	
4	0°, 5°, 10°, 15°, 30°, 45°, 60°, 75°, 90°	12	16	272 × 272	
5	0°, 5°, 10°, 15°, 30°, 45°, 60°, 75°, 90°	45	20	128 × 128	
6	0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, 90°	20	16	225 × 225	

where Γ is the number of rotation angles in a given image dataset.

5.4 Reproducible research

For reproducible research purposes, all the data needed to replicate the results reported in Tables 3 and 4 (i.e.: source code, images and subdivisions into train and validation sets) are available online [36].

6 Results

The overall results of the experiments are summarised in Tables 3 and 4. For each combination texture descriptor + feature selection method, the table reports the results in the form $\mu_a \pm \sigma_a$ (see Eqs. 13–14).

6.1 Bag-of-features vs. co-occurrences

The results show that switching from the bag-of-feature model (i.e.: $\text{LBP}_{3\times3}$ and $\text{LBP}_{8,1}^{ri}$) to co-occurrences may appreciably improve discrimination accuracy in most cases, but the trend is not peremptory. With the 1-NN classifier we observed, in four of the six image datasets included in this study – i.e: datasets no. one, three, four and five, an increase in classification accuracy of +12.4, +3.5, +10.3 and +12.3 percentage points, respectively. In dataset two, however, the difference is rather limited (+1.4), whereas in dataset six we even experience a marked decline (-5.1 percentage points). SVM returned very similar results: an increase of +12.7, +4.9, +13.4 and +8.5 percentage points with datasets one, three, four and five; a moderate increase of +3.2 with dataset two; a decrease of -4.1 percentage points with dataset six. Bag-of-features and co-occurrences give comparable results in terms of σ_a, which indicates that the two models are equally robust against rotation. No clear trend emerges as for the effect of parameter d.

6.2 Effects of dimensionality reduction

Tab. 5 reports, for each dataset, the average number of features retained by each of the strategies presented in Secs. 4.1–4.5. We notice that all the approaches provide significant dimensionality reduction from the baseline of $256 \times 256 = 65536$ features. On average, HVC tends to

Table 3 Overall results (1-NN).

Texture descriptor	Feature selection	Dataset					
		1	2	3	4	5	6
Bag-of-features (baseline)							
$\text{LBP}^{ri}_{8,1}$	NONE	70.7 ± 3.7	85.0 ± 7.5	$\mathbf{91.7} \pm 3.3$	83.0 ± 2.0	74.4 ± 1.0	72.4 ± 5.6
$\text{LBP}^{ri}_{3\times3}$	NONE	$\mathbf{74.8} \pm 3.3$	$\mathbf{87.5} \pm 5.8$	91.3 ± 2.8	$\mathbf{85.6} \pm 2.0$	$\mathbf{76.1} \pm 1.2$	$\mathbf{74.7} \pm 5.1$
Co-occurrences							
$\text{C}^{ri}_1 (\text{LBP}_{3\times3})$	NONE	84.4 ± 3.1	87.6 ± 6.4	$\mathbf{95.2} \pm 3.2$	95.0 ± 1.6	86.1 ± 0.7	63.0 ± 5.5
$\text{C}^{ri}_2 (\text{LBP}_{3\times3})$	NONE	85.7 ± 3.1	88.4 ± 7.0	94.9 ± 3.3	$\mathbf{95.9} \pm 1.8$	$\mathbf{88.4} \pm 1.8$	65.3 ± 5.5
$\text{C}^{ri}_3 (\text{LBP}_{3\times3})$	NONE	$\mathbf{87.2} \pm 2.9$	88.7 ± 7.8	94.4 ± 3.0	95.8 ± 1.8	87.2 ± 3.2	67.2 ± 5.2
$\text{C}^{ri}_1 (\text{LBP}_{3\times3})$	DC	69.2 ± 4.9	81.7 ± 6.9	85.9 ± 5.2	89.4 ± 0.7	76.8 ± 2.9	56.6 ± 3.6
$\text{C}^{ri}_2 (\text{LBP}_{3\times3})$	DC	65.5 ± 4.3	80.0 ± 8.6	82.3 ± 6.7	84.1 ± 1.7	72.5 ± 2.8	57.6 ± 5.1
$\text{C}^{ri}_3 (\text{LBP}_{3\times3})$	DC	63.8 ± 5.0	74.6 ± 10.8	79.9 ± 7.3	81.8 ± 3.3	64.7 ± 3.0	58.5 ± 5.2
$\text{C}^{ri}_1 (\text{LBP}_{3\times3})$	LDC	84.3 ± 3.1	87.6 ± 6.4	95.1 ± 3.2	94.9 ± 1.5	86.1 ± 0.7	57.4 ± 4.2
$\text{C}^{ri}_2 (\text{LBP}_{3\times3})$	LDC	85.7 ± 3.0	88.5 ± 7.0	94.9 ± 3.3	95.9 ± 1.7	88.4 ± 1.8	59.5 ± 5.3
$\text{C}^{ri}_3 (\text{LBP}_{3\times3})$	LDC	87.1 ± 2.8	$\mathbf{88.9} \pm 7.5$	94.4 ± 3.0	95.8 ± 1.8	87.2 ± 3.2	62.0 ± 5.6
$\text{C}^{ri}_1 (\text{LBP}_{3\times3})$	HVC	84.4 ± 3.1	87.6 ± 6.4	95.1 ± 3.2	95.0 ± 1.5	86.1 ± 0.7	62.4 ± 5.7
$\text{C}^{ri}_2 (\text{LBP}_{3\times3})$	HVC	85.7 ± 3.0	88.4 ± 7.0	94.9 ± 3.3	95.9 ± 1.7	88.4 ± 1.8	65.0 ± 5.5
$\text{C}^{ri}_3 (\text{LBP}_{3\times3})$	HVC	87.2 ± 2.9	88.7 ± 7.8	94.4 ± 3.0	95.8 ± 1.9	87.2 ± 3.2	67.1 ± 5.2
$\text{C}^{ri}_1 (\text{LBP}_{3\times3})$	DF	84.4 ± 3.1	87.6 ± 6.4	95.2 ± 3.2	94.9 ± 1.6	86.1 ± 0.7	61.3 ± 5.7
$\text{C}^{ri}_2 (\text{LBP}_{3\times3})$	DF	85.7 ± 3.0	88.4 ± 7.0	94.9 ± 3.3	95.9 ± 1.7	88.4 ± 1.9	65.0 ± 5.5
$\text{C}^{ri}_3 (\text{LBP}_{3\times3})$	DF	87.1 ± 2.8	88.7 ± 7.8	94.4 ± 3.0	95.8 ± 1.8	87.2 ± 3.2	67.1 ± 5.2
Bag of features + global co-occurrence statistics							
$\text{LBP}^{ri}_{3\times3} + \text{Stats}_{d=1}$	NONE	75.4 ± 3.4	88.0 ± 5.5	91.5 ± 3.2	86.8 ± 2.1	77.1 ± 1.2	$\mathbf{69.6} \pm 5.5$
$\text{LBP}^{ri}_{3\times3} + \text{Stats}_{d=2}$	NONE	77.1 ± 3.1	88.6 ± 5.0	90.8 ± 3.3	86.8 ± 2.1	77.8 ± 1.1	69.5 ± 5.4
$\text{LBP}^{ri}_{3\times3} + \text{Stats}_{d=3}$	NONE	77.3 ± 3.2	88.9 ± 4.9	91.0 ± 3.2	86.0 ± 2.1	77.1 ± 1.2	68.8 ± 5.4
Max. increment		12.4	1.4	3.5	10.3	12.3	-5.1
Other methods (Ref. [31])							
$\text{RIC-LBP}_{d=1}$	NONE	79.7 ± 2.4	91.6 ± 3.3	94.0 ± 3.7	91.3 ± 1.5	84.0 ± 1.1	72.8 ± 5.6
$\text{RIC-LBP}_{d=2}$	NONE	84.4 ± 2.0	90.9 ± 5.3	94.4 ± 3.6	92.3 ± 1.3	85.5 ± 2.0	75.3 ± 4.4
$\text{RIC-LBP}_{d=3}$	NONE	86.8 ± 2.3	91.4 ± 6.0	95.2 ± 3.6	94.8 ± 0.8	86.3 ± 2.1	74.8 ± 4.5

retain a higher number of features than the other methods, followed by DF and DC/LDC. Fig. 5 shows the average classification accuracy (over the three values of d) of the full co-occurrence model and of each feature reduction scheme (i.e.: DC, LDC, HVC and DF) obtained with the 1-NN and SVM classifiers. With the former we observe that LDC, HVC and DF virtually give the same accuracy as the original full-features model. By contrast, DC yields significantly worse accuracy, suggesting that discarding co-occurrences' labels has negative effects on classification. With the latter similar conclusions apply to DC, LDC and DF; moreover, it is worth noticing the remarkable performance of HVC, which in some cases even produces a slight increase in the accuracy with respect to the full-dimensional model (see Fig. 5 – right). Finally, it seems that combining bag-of-features and co-occurrence statistics can only provide a very slight improvement on the bag-of-features model.

7 Discussion

Related literature has consistently reported significant improvement in classification accuracy when switching from bag-of-features to co-occurrences [31,30,45, 38,35]. Herein we found rather different results: as we mentioned in Sec. 6.1, co-occurrences work pretty well in most datasets, but not in *all* datasets.

Our claim is that the higher the association between the two patterns, the lower the gain we should expect when switching from the bag-of-features model to the co-occurrence model. To clarify this concept, let us think of matrix $\bar{\mathbf{M}}_d$ (see Eq. 4) as a *contingency table*. Contingency tables are commonly used to study the behaviour of populations cross-classified with respect to two or more sets of categorical attributes (also called *polytomies* – see Ref. [12]), and in particular to establish the degree of association between them. In A we briefly recall the basics of contingency tables and report the definition of three measures of association that are used here: Cramer's V, Goodman's λ and Theil's U. All these measures take value between 0 and 1, where zero means 'independence' and one 'complete associ-

Table 4 Overall results (SVM).

Texture descriptor	Feature selection	Dataset 1	2	3	4	5	6
Bag-of-features (baseline)							
$\mathrm{LBP}^{ri}_{8,1}$	NONE	52.9 ± 3.7	75.6 ± 7.6	$\mathbf{84.0} \pm 4.5$	73.7 ± 1.4	67.1 ± 0.7	53.4 ± 3.4
$\mathrm{LBP}^{ri}_{3\times3}$	NONE	$\mathbf{56.9} \pm 3.7$	$\mathbf{79.8} \pm 5.9$	82.0 ± 4.4	$\mathbf{75.0} \pm 1.7$	$\mathbf{69.5} \pm 1.0$	$\mathbf{55.9} \pm 3.4$
Co-occurrences							
C^{ri}_1 (LBP$_{3\times3}$)	NONE	66.9 ± 3.3	81.4 ± 6.8	$\mathbf{88.9} \pm 4.5$	87.5 ± 1.8	77.4 ± 1.5	49.6 ± 0.9
C^{ri}_2 (LBP$_{3\times3}$)	NONE	58.0 ± 2.4	76.4 ± 3.9	83.9 ± 4.9	82.1 ± 1.4	72.3 ± 1.3	$\mathbf{51.8} \pm 1.8$
C^{ri}_3 (LBP$_{3\times3}$)	NONE	$\mathbf{69.6} \pm 3.0$	82.9 ± 6.8	88.5 ± 4.9	$\mathbf{88.4} \pm 1.9$	76.4 ± 3.0	49.9 ± 1.1
C^{ri}_1 (LBP$_{3\times3}$)	DC	43.7 ± 2.7	68.3 ± 6.5	70.7 ± 8.6	73.5 ± 1.0	56.6 ± 2.0	47.0 ± 0.5
C^{ri}_2 (LBP$_{3\times3}$)	DC	47.7 ± 2.3	69.8 ± 4.5	77.6 ± 6.2	77.6 ± 1.9	66.8 ± 2.5	42.4 ± 2.1
C^{ri}_3 (LBP$_{3\times3}$)	DC	42.8 ± 3.3	64.4 ± 9.0	69.0 ± 9.4	74.4 ± 2.1	52.3 ± 2.3	46.8 ± 0.6
C^{ri}_1 (LBP$_{3\times3}$)	LDC	66.8 ± 3.3	81.4 ± 6.8	88.8 ± 4.5	87.4 ± 1.8	77.4 ± 1.5	47.7 ± 0.6
C^{ri}_2 (LBP$_{3\times3}$)	LDC	56.8 ± 2.4	75.5 ± 4.1	83.1 ± 5.0	81.4 ± 1.4	71.8 ± 1.4	43.1 ± 2.0
C^{ri}_3 (LBP$_{3\times3}$)	LDC	69.5 ± 3.0	$\mathbf{83.0} \pm 6.8$	88.5 ± 4.9	88.3 ± 1.9	76.4 ± 3.0	47.7 ± 0.8
C^{ri}_1 (LBP$_{3\times3}$)	HVC	65.1 ± 3.5	79.2 ± 6.7	86.9 ± 4.2	86.3 ± 2.1	$\mathbf{78.1} \pm 0.7$	47.8 ± 0.8
C^{ri}_2 (LBP$_{3\times3}$)	HVC	66.9 ± 3.3	81.4 ± 6.8	88.8 ± 4.5	87.5 ± 1.8	77.4 ± 1.5	49.5 ± 0.9
C^{ri}_3 (LBP$_{3\times3}$)	HVC	69.5 ± 3.0	82.9 ± 6.8	88.5 ± 4.9	88.4 ± 1.9	76.4 ± 3.0	49.8 ± 1.1
C^{ri}_1 (LBP$_{3\times3}$)	DF	66.9 ± 3.3	81.4 ± 6.8	88.8 ± 4.5	87.4 ± 1.8	77.4 ± 1.5	49.4 ± 0.9
C^{ri}_2 (LBP$_{3\times3}$)	DF	57.5 ± 2.4	75.8 ± 4.1	83.7 ± 4.9	81.7 ± 1.4	72.1 ± 1.3	49.7 ± 1.9
C^{ri}_3 (LBP$_{3\times3}$)	DF	69.5 ± 3.0	82.9 ± 6.8	88.5 ± 4.9	88.3 ± 1.9	76.4 ± 3.0	49.8 ± 1.1
Bag of features + global co-occurrence statistics							
$\mathrm{LBP}^{ri}_{3\times3}$ + Stats$_{d=1}$	NONE	55.7 ± 3.6	79.5 ± 5.5	81.4 ± 4.5	75.1 ± 1.7	69.9 ± 1.0	51.3 ± 2.1
$\mathrm{LBP}^{ri}_{3\times3}$ + Stats$_{d=2}$	NONE	57.3 ± 3.5	80.5 ± 5.1	81.5 ± 4.6	75.3 ± 1.6	70.3 ± 1.0	51.3 ± 1.5
$\mathrm{LBP}^{ri}_{3\times3}$ + Stats$_{d=3}$	NONE	65.1 ± 3.5	79.3 ± 6.6	87.0 ± 4.1	86.2 ± 2.1	78.1 ± 0.7	47.9 ± 0.8
Max. increment		12.7	3.2	4.9	13.4	8.5	-4.1
Other methods (Ref. [31])							
RIC-LBP$_{d=1}$	NONE	62.9 ± 2.0	79.4 ± 5.4	86.1 ± 5.1	83.5 ± 1.3	73.6 ± 2.2	55.9 ± 1.0
RIC-LBP$_{d=2}$	NONE	68.8 ± 2.1	85.8 ± 4.0	89.7 ± 4.4	86.8 ± 0.7	74.6 ± 2.7	54.4 ± 1.3
RIC-LBP$_{d=3}$	NONE	68.8 ± 2.1	85.8 ± 4.0	89.7 ± 4.4	86.8 ± 0.7	74.6 ± 2.7	54.4 ± 1.3

Table 5 Average number of features used for classification.

Texture descriptor	Feature selection	Dataset 1	2	3	4	5	6
C^{ri}_1 (LBP$_{3\times3}$)	DC, LDC	2212	632	1951	2134	2056	13
C^{ri}_2 (LBP$_{3\times3}$)	DC, LDC	4987	1473	4546	5118	4525	20
C^{ri}_3 (LBP$_{3\times3}$)	DC, LDC	5713	1938	5372	5970	5323	23
C^{ri}_1 (LBP$_{3\times3}$)	HVC	4530	1491	2593	3276	5428	225
C^{ri}_2 (LBP$_{3\times3}$)	HVC	12179	3225	6005	10604	13997	295
C^{ri}_3 (LBP$_{3\times3}$)	HVC	13847	4098	6811	12736	15703	284
C^{ri}_1 (LBP$_{3\times3}$)	DF	3886	1043	3598	2372	3203	181
C^{ri}_2 (LBP$_{3\times3}$)	DF	7585	2333	8239	5126	5513	293
C^{ri}_3 (LBP$_{3\times3}$)	DF	8013	3027	9425	5453	5832	333
$\mathrm{LBP}^{ri}_{3\times3}$ + Stats	None	41	41	41	41	41	41
RIC-LBP$_{d=\{1,2,3\}}$	None	136	136	136	136	136	136

ation'. When applied to co-occurrences of patterns, a value close to zero means weak association between the two patterns of a pair; a value close to one means strong association. More specifically, Cramer's V is related to the concept of independence in a bivariate distribution: in this context it tells us the extent to which we can consider the co-occurrence distribution as the product of two independent bag-of-feature models. Goodman's λ is a measure based on optimal prediction: it estimates how easily we can infer the label of one of the two patterns of a pair when we know the other's. Finally, Theil's U is a measure of the cross-information conveyed by one of the two patterns of a pair about the other and vice-versa. We have estimated the mean value of V, λ and U for all the datasets considered in the experiments by averaging the value for each image in the dataset, each
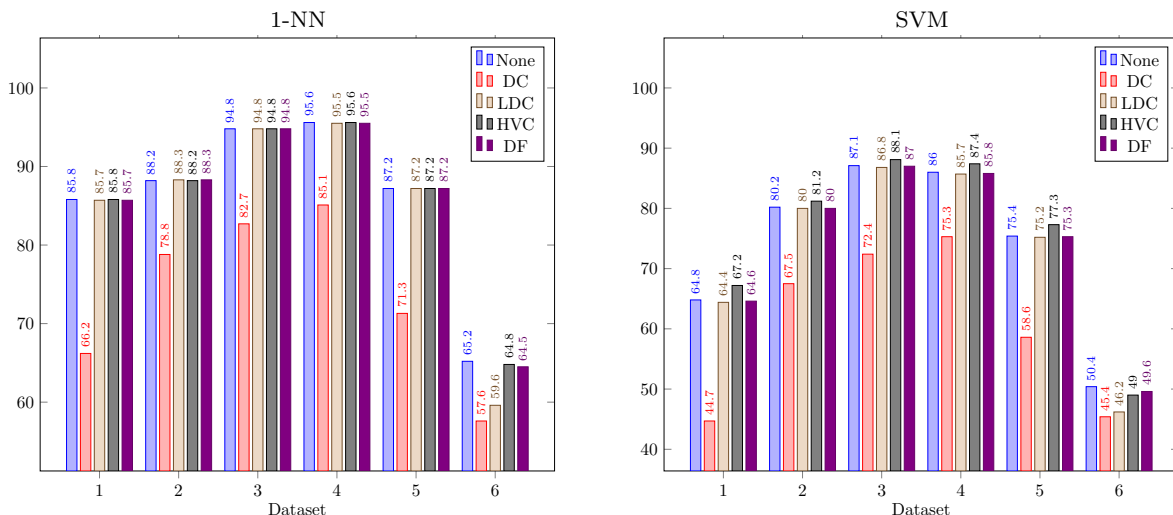
Fig. 5 Efficiency of the feature reduction schemes – average results of C_d^{ri} ($LBP_{3\times3}$) for $d = \{1, 2, 3\}$.

angle and each distance d. The results are reported in Tab. 6.

Table 6 Average measures of association for each dataset.

Measure	Dataset					
	1	2	3	4	5	6
Cramer's V	0,079	0,104	0,074	0,071	0,088	0,142
Goodman's λ	0,059	0,158	0,069	0,043	0,053	0,110
Theil's U	0,136	0,207	0,131	0,118	0,146	0,263

We notice that the measures of association are appreciably higher in datasets two and six; namely those datasets for which the worst (and even negative) is the gain we obtain when switching from bag-of-features to the co-occurrence model. It appears that the easier it is to infer the label of a patterns from that of the other pattern, the less the improvement we can obtain. Likewise, reversing the reasoning and thinking in terms of information, we could say that the higher the surprise of seeing the label of one of the two patterns given the other's, the higher the improvement we may expect from the co-occurrence model.

The question arises whether there is a perceptual correlate to the class of textures for which co-occurrences do or do not improve on the bag-of-features model. In our view, no significant difference can be appreciated between the images of datasets one to five. On the other hand, images of dataset six look significantly different from the others: these are composed of artificial textures – as contrasted with those of datasets one to five, which are natural. Remarkably, two things happen with

dataset six: 1) the number of retained features after any of the selection processes considered in the paper is significantly lower than in any other dataset; 2) the use of co-occurrences worsen the classification accuracy. It therefore seems there is, at least for dataset six, a perceptual correlation between the outcome and the image appearance. It is nonetheless difficult, at present, to extend this observation into a general statement.

8 Conclusions and future work

In this work we have described a general framework for rotation invariant co-occurrences of patterns for texture classification. We have also discussed five different strategies for reducing the high dimensionality of this model. Through a set of classification experiments we have evaluated the potential advantages that co-occurrences can provide in comparison with the bag-of-features approach. Our results show that co-occurrences remarkably improve classification accuracy in most cases, but in others the gain is negligible, or even negative. In our opinion this result is most probably related to the degree of association between the patterns that compose a pair, and, in particular, that a high degree of association correlates with poorer outcome.

Of the five strategies for dimensionality reduction, LDC and DF proved capable of reducing the overall number of features by approximately one order of magnitude – and even more in some datasets – with practically no accuracy loss. HVC performs at least as well, and in some cases even gives better accuracy than the full-dimensional model, but retains a higher number of features than the other reduction schemes. DC gave the

worst results in terms of classification accuracy. Concatenation of bag-of-features and co-occurrence statistics showed that the increment attainable with this method is limited.

In summary, the experimental results show that co-occurrences can, in many cases, improve the bag-of-feature model, but in other cases the improvement is limited or even negative. The number of features can be significantly reduced from the full-dimensional model, yet it might remain too high for some practical applications. Theoretically, this number could be further reduced by adjusting the 80% rule (see Sec. 4.1) to the specific application needs. These considerations suggest that, in any case, the use of co-occurrences of patterns in practical applications should be considered with care, and the benefit versus the bag-of-features alone carefully evaluated.

It is worth recalling, as a final remark, that we focussed the present paper on the analysis of images containing stationary textures. An extension of this study to non-stationary textures, for scene gist or crude object categorization – for instance, could be an interesting subject for future work. We cannot in principle exclude that things could be different in that case.

Aknowledgements

A Contingency tables and measures of association

Given a population and two sets of unordered, categorical attributes (polytomies) $\mathcal{A} = \{A_0, A_1, \ldots, A_{\alpha-1}\}$ and $\mathcal{B} = \{B_0, B_1, \ldots, B_{\beta-1}\}$ a contingency table is an $\alpha \times \beta$ matrix of which each (ρ_a, ρ_b) entry ($a \in \{0, \ldots, \alpha-1\}$; $b \in \{0, \ldots, \beta-1\}$), reports the fraction of the population that is classified as both A_a and B_b [12]. The degree of association between the attributes can be estimated through a number of parameters, some of which are breifly recalled here below.

A.1 Pearson's chi-squared coefficient and Cramér's V

A traditional measure of association is Pearson's chi-squared coefficient, which estimates the bias of cross-classification from statistical independence [26]. For a population of ν members it is defined as follows:

$$\chi^2 = \nu \sum_{a=0}^{\alpha-1} \sum_{b=0}^{\beta-1} \frac{(\rho_{ab} - \rho_a \cdot \rho_{\cdot b})^2}{\rho_a \cdot \rho_{\cdot b}} \tag{15}$$

where ρ_a. and $\rho_{\cdot b}$ are the proportions of the population classified as A_a and B_b, respectively.

A normalized version of Pearson's χ^2 is Cramér's V, which takes value between 0 and 1 (where 0 means 'independence' and 1 'complete association') and is defined as follows:

$$V = \sqrt{\frac{\chi^2}{\nu \left[\max\left(\alpha-1, \beta-1\right)\right]}} \tag{16}$$

A.2 Goodman's λ

Suppose a situation in which we had to guess the B-class (or the A-class) of an individual chosen at random from the population and were given: 1) no further information about the individual, or 2) its A class (or its B class). Goodman's λ estimates the relative decrease in the error probability that we experience when switching from case 1) to case 2):

$$\lambda = \frac{\frac{1}{2}\left(\sum_{a=0}^{\alpha-1} \rho_{am} + \sum_{b=0}^{\beta-1} \rho_{bm} - \rho_{\cdot m} - \rho_{m \cdot}\right)}{1 - \frac{1}{2}\left(\rho_{\cdot m} + \rho_{m \cdot}\right)} \tag{17}$$

where

$$\rho_{\cdot m} = \max_{b}\left(\rho_{\cdot b}\right), \quad \rho_{am} = \max_{b}\left(\rho_{ab}\right) \tag{18}$$

$$\rho_{m \cdot} = \max_{a}\left(\rho_{\cdot a}\right), \quad \rho_{mb} = \max_{a}\left(\rho_{ab}\right) \tag{19}$$

A.3 Theil's U

The last measure of association considered here is based on the concept of cross-information and was discussed by Theil [43]. It considers on the amount of information conveyed about A by B and vice-versa. This is properly scaled to give a value between 0 and 1. In formulas:

$$U^2 = \frac{2I}{H(A) + H(B)} \tag{20}$$

where

$$I = \sum_{a=0}^{\alpha-1} \sum_{b=0}^{\beta-1} (\rho_{ab}) \log\left(\frac{\rho_{ab}}{\rho_a \cdot \rho_{\cdot b}}\right) \tag{21}$$

$$H(A) = -\sum_{a=0}^{\alpha-1} (\rho_a \cdot) \log\left(\rho_a \cdot\right) \tag{22}$$

$$H(B) = -\sum_{b=0}^{\beta-1} (\rho_{\cdot b}) \log\left(\rho_{\cdot b}\right) \tag{23}$$

References

1. Amsterdam library of textures (ALOT) (2009). Available online at http://staff.science.uva.nl/~aloi/public_alot/
2. Bianconi, F., Fernández, A., González, E., Armesto, J.: Robust colour texture features based on ranklets and discrete Fourier transform. Journal of Electronic Imaging **18**, 043,012–1–8 (2009). DOI 10.1117/1.3273946

3. Bianconi, F., González, E., Fernández, A., Saetta, S.A.: Automatic classification of granite tiles through colour and texture features. Expert Systems with Applications **39**(12), 11212–11218 (2012) DOI 10.1016/j.eswa.2012.03.052

4. Boureau, Y.L., Ponce, J., LeCun, Y.: A theoretical analysis of feature pooling in visual recognition. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 111–118. Haifa, Israel (2010)

5. Brodatz, P.: Textures: a photographic album for artists and designers. Dover Publications, New York, USA (1966)

6. Burghouts, G.J., Geusebroek, J.M.: Material-specific adaptation of color invariant features. Pattern Recognition Letters **30**(3), 306–313 (2009)

7. Crosier, M., Griffin, L.D.: Using basic image features for texture classification. International Journal of Computer Vision **88**(3), 447–460 (2010)

8. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Adapted vocabularies for generic visual categorization. In: Proceedings of Workshop on Statistical Learning in Computer Vision, 8th European Conference on Computer Vision. Prague, Czech Republic (2004)

9. Ershad, S.: Texture classification approach based on combination of edge & co-occurrence and local binary pattern. In: Proceedings of International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), vol. 2, pp. 626–629. Nevada, USA (2011)

10. Fernández, A., Ghita, O., González, E., Bianconi, F., Whelan, P.F.: Evaluation of robustness against rotation of LBP, CCR and ILBP features in granite texture classification. Machine Vision and Applications **22**(6), 913–926 (2011). DOI 10.1007/s00138-010-0253-4

11. Fernández, A., Álvarez, M.X., Bianconi, F.: Texture description through histograms of equivalent patterns. Journal of Mathematical Imaging and Vision **45**(1), 76–102 (2013) DOI 10.1007/s10851-012-0349-8

12. Goodman, L., Kruskal, W.: Measures of association for cross classifications. Journal of the American Statistical Association **49**(268), 732–764 (1954)

13. Guo, Z., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. IEEE Transactions on Image Processing **19**(6), 1657–1663 (2010)

14. Guo, Y., Zhao, G., Pietikäinen, M.: Discriminant features for texture description. Pattern Recognition **45**(10), 3825–3843 (2012)

15. Hsu, C., Chang, C., Lin, C.: A practical guide to support vector classification (2010). Available online at `http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf`. Last accessed on December 27, 2013

16. Kandaswamy, U., Schuckers, S.A., Adjeroh, D.: Comparison of texture analysis schemes under nonideal conditions. IEEE Transactions on Image Processing **20**(8), 2260–2275 (2011)

17. Khan, R., Barat, C., Muselet, D., Ducottet, C.: Spatial orientations of visual word pairs to improve bag-of-visual-words model. In: Proceedings of British Machine Vision Conference (BMVC 2012). Guildford, UK (2012)

18. Kobayashi, T., Otsu, N.: Bag of hierarchical co-occurrence features for image classification. In: Proceedings of the 20th International Conference on Pattern Recognition (ICPR), pp. 3882–3885. Istanbul, Turkey (2010)

19. Kylberg Sintorn Rotation dataset (2013). Available online at `http://www.cb.uu.se/~gustaf/KylbergSintornRotation/`. Last accessed on October 24, 2013

20. Lazebnik, S., Schmid, C., , J., P.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178. New York, USA (2006)

21. Lee, Y.S., Hao, S.S., Lin, S.W., Li, S.Y.: Image retrieval by region of interest motif co-occurrence matrix. In: IEEE International Symposium on Intelligent Signal Processing and Communication System (ISPACS), pp. 270–274. New Taipei, Taiwan (2012)

22. Li, Q., Shi, Z.: A high order contextual descriptor for image retrieval using generalized texton co-occurrence matrix. Information **16**(1A), 155–174 (2013)

23. Liao, S., Law, M.W.K., Chung, A.C.S.: Dominant local binary patterns for texture classification. IEEE Transactions on Image Processing **18**(5), 1107–1118 (2009)

24. LIBSVM – A Library for Support Vector Machines (2014). Available online at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`. Last accessed on January 9, 2014

25. Liu, G.H., Yang, J.Y.: Image retrieval based on the texton co-occurrence matrix. Pattern Recognition **41**(12), 3521–3527 (2008)

26. Mirkin, B.: Eleven ways to look at the chi-squared coefficient for contingency tables. The American Statistician **55**(2), 111–120 (2001)

27. Mondial Marmi: a granite image database for colour and texture analysis. v1.1. (2011). Available online at `http://dismac.dii.unipg.it/mm`. Last accessed on October 24, 2013

28. Nanni, L., Brahnam, S., Lumini, A.: Selecting the best performing rotation invariant patterns in local binary/ternary patterns. In: Proceedings of the 2010 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'10), pp. 369–375. Las Vegas, USA (2010)

29. Nanni, L., Brahnam, S., Lumini, A.: Random interest regions for object recognition based on texture descriptors and bag of features. Expert Systems with Applications **39**(1), 973–977 (2012)

30. Nosaka, R., Ohkawa, Y., Fukui, K.: Feature extraction based on co-occurrence of adjacent local binary patterns. In: Advances in Image and Video Technology. Proceedings of the 5th Pacific Rim Symposium on Image and Video Technology, 2011, *Lecture Notes in Computer Science*, vol. 7088, pp. 82–91. Springer, Gwangju, South Korea (2012)

31. Nosaka, R., Suryanto, C., Fukui, K.: Rotation invariant co-occurrence among adjacent LBPs. In: Computer Vision – ACCV 2012 International Workshops, *Lecture Notes in Computer Science*, vol. 7728, pp. 15–25. Springer, Daejon, South Korea (2013)

32. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7), 971–987 (2002)

33. Olver, P.: Classical invariant theory, *London Mathematical Society Student Texts*, vol. 44. Cambridge University Press, Cambridge, UK (1999)

34. PRTools: A Matlab toolbox for pattern recognition (2014). Available online at `http://prtools.org/`. Last accessed on January 9, 2014

35. Qi, X., Xiao, R., Guo, J., Zhang, L.: Pairwise rotation invariant co-occurrence local binary pattern. In: Proceedings of the European Computer Vision Conference,

Lecture Notes in Computer Science, vol. 7577, pp. 158–171. Springer, Florence, Italy (2012)

36. General framework for rotation invariant texture classification through co-occurrence of patterns (2013). Code, data and results of this paper. Available online at `http://webs.uvigo.es/antfdez/downloads.html`. Last updated on Jan 14, 2013

37. Ros, J., Laurent, C., Jolion, J.: A bag of strings representation for image categorizations. Journal of Mathematical Imaging and Vision **35**(1), 51–67 (2009)

38. Shadkam, N., Helfroush, M.: Texture classification by using co-occurrences of local binary patterns. In: Proceedings of the 20th Iranian Conference on Electrical Engineering, pp. 1442–1446. Tehran, Iran (2012)

39. Song, Q.: Illumination invariant texture classification with pattern co-occurrence matrix. In: G. Shen, X. Huang (eds.) Advanced Research on Computer Science and Information Engineering, *Communications in Computer and Information Science*, vol. 152, pp. 67–72. Springer (2011)

40. Sujatha, B., Vijayakumar, V., Harini, P.: A new logical compact LBP co-occurrence matrix for texture analysis. International Journal of Scientific & Engineering Research **3**(2), 1–5 (2012)

41. Sun, X., Wang, J., Chen, R., She, M., Kong, L.: Multiscale local pattern co-occurrence matrix for textural image classification. In: Proceedings of the IEEE World Congress on Computational Intelligence, pp. 1–7. Brisbane, Australia (2012)

42. Tan, C.M., Wang, Y.F., Lee, C.D.: The use of bigrams to enhance text categorization. Information Processing and Management **38**(4), 529–546 (2002)

43. Theil, H.: Statistical Decomposition Analysis: With Applications in the Social and Administrative Sciences, *Studies in Mathematical and Managerial Economics*, vol. 14. North-Holland (1972)

44. Varma, M., Zisserman, A.: A statistical approach to material classification using image patch exemplars. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(11), 2032–2047 (2009)

45. Wang, Z., Liu, H., Xu, T.: Crowd density estimation based on local binary pattern co-occurrence matrix. In: Proceedings of the IEEE International Conference on Multimedia and Expo Workshops, pp. 372–377. Melbourne, Australia (2012)

46. Wu, Z., Huang, Y., Wang, L., Tan, T.: Spatial graph for image classification. Lecture Notes in Computer Science **7724**, 716–729 (2013)

47. Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. International Journal of Computer Vision **73**(2), 213–238 (2007)

48. Zou, J., Liu, C.C., Zhang, Y., Lu, G.F.: Object recognition using Gabor co-occurrence similarity. Pattern Recognition **46**(1), 434–448 (2013)