Journal of Visual Communication and Image Representation 49 (2017) 447–458

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Color enhanced local binary patterns in covariance matrices descriptors (ELBCM)☆

Michèle Gouiffès[a,*], Andres Romero Mier y Terán[a], Lionel Lacassagne[b]

[a] LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay, France
[b] Sorbonne Universités, UPMC Univ. Paris 06, CNRS UMR 7606, LIP6, France

ARTICLE INFO

Keywords:
Covariance matrices
Region descriptors
LBP
Image matching
Texture
Color
Object retrieval
Person re-identification

ABSTRACT

This paper proposes a new version of LBP and its inclusion into covariance region descriptors for image matching and recognition. Starting from the non-rotation invariant uniform LBP (called nriLBP), the pattern is described by the cosine and sine values of the angular portion defined by the '1's. The use of this four-value vector leads to a better resilience of the feature to noise and small neighborhood rotations. Several color versions of this feature are proposed. For region description, these local features are included in covariance matrices, noted ELBCM for Enhanced-LBP Covariance Matrix. Experimental evaluations confirm the relevance of the proposed models on three databases designed for texture analysis, object retrieval and person re-identification. A study is also made on the impact of the colorspace included in the covariance descriptor and used for LBP definition. The experiments show that ELBCM has better recognition performance than the 12 other descriptors tested.

1. Introduction

In computer vision, many applications require to describe regions or objects in order to be recognized or matched: tracking, texture classification, human detection or object retrieval are just a few examples among others. A good descriptor should be distinctive *i.e.* it should capture the most representative characteristics of the region's appearance while being invariant to most of the following phenomena and artifacts such as geometric transformation, occlusion, viewpoint, lighting changes, blur, noise, especially for embedded applications. Ideally it should be compact so as to save memory and time as much as possible.

Regions can be described in various ways [1], by a set of feature points [2,3], by their contours, by color or gradient histograms [4]. However, the choice made on a descriptor generally depends on the application, on the acquisition conditions and on the object under consideration: whether it is colorful, textured, structured; whether it has a rigid motion or not. Covariance region descriptors proposed by Tuzel et al. [5] provide a compact model that embeds various image cues (e.g., intensity, directional gradients, spatial coordinates, texture representations, optical flow) using their correlations, leading to a Symmetric Positive Definite (SPD) matrix. To some extent, these descriptors are invariant to uniform illumination changes, and can offer a good resilience to scale and rotation variations. Their computation can be speeded up by using integral images, applying structure transforms or using specific instructions [6], then the computation time becomes the same whatever the region size.

Based on these qualities, covariance descriptors have proven their good behavior in various applications: object detection [5], texture classification [7–10], human detection [11,12], tracking [13,14] and re-identification [15], face recognition [16], action recognition using temporal derivatives [17,9] or fire and flame detection [18] where flames are considered as a dynamic texture of characteristic color.

Several tracks arise from these various works: the number and the nature of the features that feed the covariance descriptor as well as the way they are included; the spatial arrangement of the covariance matrices on the object to be described and possibly the weights given to each of them; finally, the way to handle metrics on the Riemannian manifold to which the covariance matrices belong.

Concerning the latter issue, the manifold can be viewed intuitively as a continuous surface lying in a high dimensional space. The question to address is then how to handle distances in such a complex space and therefore how to apply classification, comparison or machine learning techniques in a rigorous way. Several approaches have been proposed: using specific metrics based on Lie algebra [19,13], mapping the points on the manifold to the tangent space of the identity matrix [17], using the properties of Grassmann manifolds and their embedding in Reproducing Kernel Hilbert Spaces (RKHS) [9] or using a Log-Euclidean

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: michele.gouiffes@limsi.fr (M. Gouiffès), andres.romero@gmail.com (A. Romero Mier y Terán), lionel.lacassagne@lip6.fr (L. Lacassagne).

http://dx.doi.org/10.1016/j.jvcir.2017.09.012
Received 12 May 2017; Received in revised form 18 August 2017; Accepted 27 September 2017
Available online 19 October 2017

framework [8].

Concerning now the first track, the nature of features depends on the application. For region matching, the descriptors embed space, color and texture features. Any colorspace can be imagined, from RGB in most works [5] to invariant colorspaces [20] or can be chosen depending on the database. As for texture, let us mention spatial derivatives [5], Gabor filter responses [16] and Local Binary Patterns (LBP). The latter features have gain in interest due to their good performance and gave birth to LBCM [21] or GLRCD [22].

While many works deal with integrating texture in the region descriptor, fewer works are dedicated for color LBP features and even less concerning their use for covariance descriptors. However, this can increase the discriminant power of the descriptor and provide a better invariance to illumination changes. To that aim, color invariant features and Local Binary Patterns focused our attention. First, a LBP descriptor called Enhanced LBP (ELBP) is proposed for monochrome images and its marginal and vectorial color extensions are given. The previous work [23] which proposed gray ELBP with a far less thorough evaluation. After an analysis of their sensitivity to color appearance changes, the features are used for region description by covariance matrices. These descriptors are evaluated experimentally for different datasets designed for texture classification, object retrieval and person re-identification. A study is made on the choice of color cues to be included in the descriptor, which can have a big impact on the descriptiveness, the compactness and the invariance to distortions.

The remainder of this paper is organized as follows. Section 2 recalls the underlying theory of covariance matrix descriptors and discusses the features generally included. The proposed LBP features are presented in Section 3. Then, the experiments are detailed in Section 4. Finally our conclusions are given in Section 6.

2. Covariance descriptors

After a short explanation of the key principles of covariance region descriptors in 2.1, the existing spatial, color and texture features are introduced in 2.2.

2.1. Principles

Let I represent a luminance (grayscale) or a vectorial image (color, multispectral, infrared, depth images…), and let F be the $W \times H \times d$ dimensional feature image extracted from I

$$F(x,y) = \phi(I,x,y) \tag{1}$$

where ϕ is any d-dimensional mapping forming a feature vector for each pixel (see 2.2). Let now $\{\mathbf{z}_k\}_{k=1,\dots,n}$ be a set of d-dimensional feature vectors associated to each point inside the rectangular region $R \subset F$ of size n. This region is represented with the $d \times d$ covariance matrix

$$\mathbf{C}_R = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{z}_k - \boldsymbol{\mu})(\mathbf{z}_k - \boldsymbol{\mu})^T \tag{2}$$

where $\boldsymbol{\mu}$ is the mean vector of the $\mathbf{z}_k$ vectors.

The covariance matrix is a $d \times d$ square matrix which fuses multiple features naturally measuring their correlations. Elements in the main diagonal represent the variance of each feature, while elements outside this diagonal represent their correlations. Due to the averaging in the covariance computation, noisy pixels are largely filtered out which contrasts with raw-pixel methods (block matching typically). Covariance matrices are low-dimensional compared to other descriptors, and have only $(d^2 + d)/2$ different values due to symmetry, whatever the region size. When no information regarding neither the ordering of the pixels nor their number is used, the descriptor $\mathbf{C}_R$ is robust to scale and rotation changes. It ceases to be totally rotation invariant when orientation information is introduced in the feature vector such as the partial derivatives with respect to x and y directions.

Comparing and classifying covariance descriptors require basically to measure distances between two points on the manifold or to compute simple statistics such as the barycenter of all the samples coming from each class in order to learn a representative model. As mentioned in the introduction, this is an important research topic [13,17,9,8]. For interested readers, more information about SPD is available in [24].

Matching and tracking imply calculating multiple covariance descriptors in overlapping regions where a high redundancy of operations is expected. To solve the problem, integral images are used to store all the accumulations computed from the origin of the image to each possible point, then the covariance values are obtained fastly by only four readings in these images [25,3].

2.2. Space, color and texture features

Generally, the feature vector collects spatial, color (or intensity) and texture (or gradient) information, and also but less frequently dynamic information when motion analysis is targeted [26]. In the seminal work of [5] the features are the spatial coordinates (x,y), the luminance and its first and second order derivatives, yielding to a 7×7 descriptor. Since then, various extensions have been proposed either to improve the descriptive power of the representation, to reduce its size or to address new applications.

In terms of texture features, Pang et al. [16] proposed GRCM which embeds 40 Gabor filter responses, for 8 orientations at 5 different scales, in order to address the problem of face recognition which is particularly demanding in terms of discriminating ability. Due to their wide size Gabor-based covariance descriptors are expensive to compute and to match, that is why some attempts have been made to use only 12 or 24 Gabor filter bank responses [7], *i.e.* 4 (or 8) orientations with 3 radial center frequencies. Some works use LBP [27] as in the *Local Binary Covariance Matrix* (LBCM) [21] in order to obtain a more compact representation. In [22], spatial coordinates, RGB components, three Gabor features (one direction and three different scales) and the LBP decimal value are collected together to build the *Gabor-LBP based Region Covariance Descriptor* (GLRCD) applied to person re-identification. The specific problem of introducing LBP in covariance descriptors is discussed more thoroughly in Section 3.

Except for some applications which do not require any color cues such as face [12] or pedestrian detection, three color components are generally included in the descriptor. The colorspace can be RGB as in the seminal work [5], HSV for the photometric invariance [28], $L^*a^*b^*$ [29] or invariants [30]. The gradient vector is generally computed from the luminance information. The MRCG (Mean Riemannian Covariance Grid) is an exception [31] which concatenates the gradients magnitude and orientation of each RGB channel.

Optical flow vectors can directly be used as dynamic features [18]. To detect moving objects in the context of a static camera, the feature vector can include background subtraction features or better, a foreground probability value indicating the probability that the pixel belongs to the moving foreground [12], possibly with its spatial derivatives.

Thus, the variety of possible descriptors makes it difficult to choose the best configuration for a given application. To solve this issue, the COSMATI (COrrelation-based Selection of covariance MATrIces) feature selection approach [15] identifies from ten samples picked during the training stage, the four most descriptive features. For each class, a correlation-based technique selects the features that are highly correlated with the class and uncorrelated with the other classes.

For object retrieval and particularly for person re-identification, the priority is generally awarded to the descriptive power of the representation more than to its compactness, at the cost or a higher execution time. As an example, *Bio-Inspired Features* (BIF), used elegantly in [28] consist in producing 128 Gabor filter responses, that are then divided into small overlapping rectangular regions on which a *max pooling* operator is applied. The global region is then described by the

set of all the covariance matrices computed on these features.

The next section proposes several LBP features and their introduction in the covariance matrix.

3. Enhanced local binary covariance matrices (*ELBCM*)

After a brief explanation of LBP features in 3.1, the proposed monochrome ELBP feature is introduced in Section 3.2 as well as its color extensions in Section 3.3. To finish, section provides a summary of the descriptors that will be compared in the experiments.

3.1. Local Binary Patterns (LBP)

Signal processing methods based on filter banks are computationally too complex to meet the real-time requirements of many computer vision applications. Recently, more economical and still discriminative methods such as LBP have been proposed. An excellent source of information about the LBP operators and their applications in computer vision is the book [32]. The computation starts by the labeling of the neighborhood of each pixel as a binary number according to a predefined set of comparisons. One important advantage of LBP is that they are invariant against monotonic gray level changes caused by illumination variations. They are computationally simple and very discriminant, which makes them rivalize with filter banks in reaching high performance in texture classification. The original LBP features were introduced in [27] where the authors highlighted two important aspects of a texture: a pattern and its strength. They were generalized several years later in [27]. The first row of Fig. 1 illustrates the concept. The idea is to build a binary code of length P (*i.e.*, for a 8-pixel neighborhood there are $2^8 = 256$ possible labels). The central gray level g_c is used as a threshold for the neighboring values, then the resulting thresholded values are accumulated considering a weight related to their position, these weights increasing clockwise as powers of two: $\{2^0,2^1,...,2^{P-1}\}$.

The local texture of a monochrome image is characterized at each pixel (x_c,y_c) by the joint distribution of $P + 1$ gray values, that is the gray value of point g_c and of its P neighbor pixels g_p located on a circle of radius R around the central point. The distribution is approximated by the local neighborhood differences in order to make the operator invariant to changes of the mean gray value. For a higher invariance, only signs of the differences are considered $s(g_p-g_c)$. Accumulating the thresholded differences weighted by powers of two, the operator is finally defined as

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p-g_c)2^p, \text{ with } s(x) = \begin{cases} 1 & \text{when } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

which results in 2^P different labels, 256 for the example of Fig. 1. The rotation invariant version noted rorLBP corresponds to the rotation of the binary pattern until the minimum code is obtained. The variance of LBP [27] is simply defined as follows:

$$V_{P,R} = \frac{1}{P}\sum_{p=0}^{P-1} (g_p-u)^2 \tag{4}$$

where u is the average gray value of the P neighboring pixels.

Uniform patterns [27] are the result of an additional modification made to the original definition of LBP. Here, a uniformity measure that counts the number of bitwise transitions from 0 to 1 or vice versa is employed. A pattern is considered as *uniform* if its measure is $\leqslant 2$. For example, 00000000 and 01110000 are uniform but 11001001 and 10010011 are not. It has been proven experimentally that uniform LBP account for the majority of texture patterns in natural images, and that they are more stable or less prone to noise. They are also convenient because they are less numerous, which produces more compact and less sparse histogram distributions. There are two versions of the uniform LBP, whether they are normalized to be invariant to rotation (then they are noted uLBP) or not (nriLBP). The latter will be use to design the ELBP.

3.2. Enhanced local binary patterns (ELBP)

The right way to introduce LBP into the covariance descriptors is not obvious. GLRCD [22] directly uses LBP's decimal value in a very compact way but the resulting descriptor is unstable for some very common phenomena such as local neighborhood rotations. Additionally, carrying typical arithmetic operations with such values is meaningless *i.e.* adding or averaging two or more LBP decimal values does not mean anything in terms of texture. LBCM descriptor [21] uses bit-strings where each bit acts as an independent feature yielding to a more stable descriptor for which arithmetic operations are correctly defined. In return, this representation is far less compact, since the number of features is the number of bits forming the LBP pattern generally 8 or more. This has also a significant impact on the execution speed. The LBP variance (4) can also be introduced in the covariance descriptor [14]. It leads to a compact but less distinctive descriptor compared to LBCM. However, associated with color this is acceptable for tracking because temporal coherence helps in disambiguating matchings.

The proposed enhanced LBP-based covariance descriptor (denoted ELBCM) uses the angles formed by nriLBP mentioned in 3.1. In the case of the $LBP_{8,1}$ operator there are 58 possible uniform patterns, but the strings formed by all ones and all zeros are ignored because their angle representations are the same, which can be misleading. Fig. 2 shows the remaining 56 different patterns used to construct our ELBCM. The starting and ending angles θ_0 and θ_1 are marked by the curved arrows inside each pattern. Since computations have to be made on circular quantities, the angles are converted to points on the unit circle, e.g., θ_0 is converted to $(\cos(\theta_0),\sin(\theta_0))$. The following vector

$$\alpha(\theta_0,\theta_1) = [\cos(\theta_0)\sin(\theta_0)\cos(\theta_1)\sin(\theta_1)] \tag{5}$$

uniquely describes each nriLBP as a function of (θ_0,θ_1). An example for

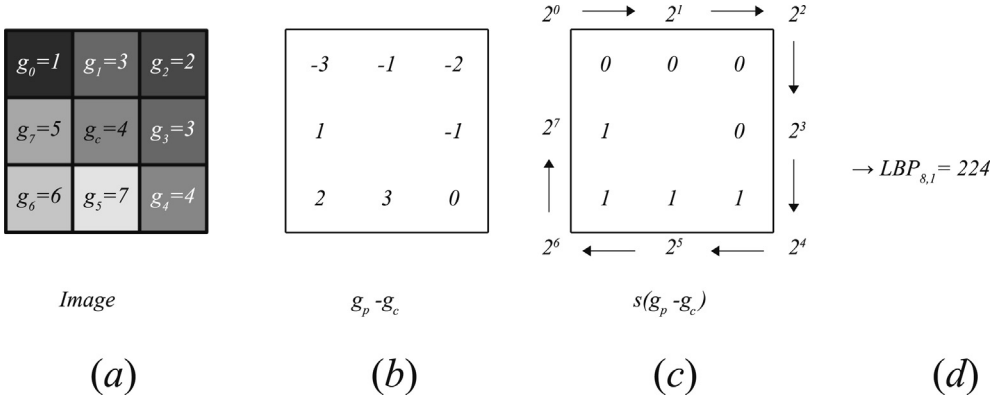


Fig. 1. An example of $LBP_{8,1}$ computed on the image (a). (b) The central pixel g_c is subtracted to each neighbor pixel g_p and (c) the sign of this difference leading to a binary code (11110000 here) and (d) the decimal value (224 here) after weighting.

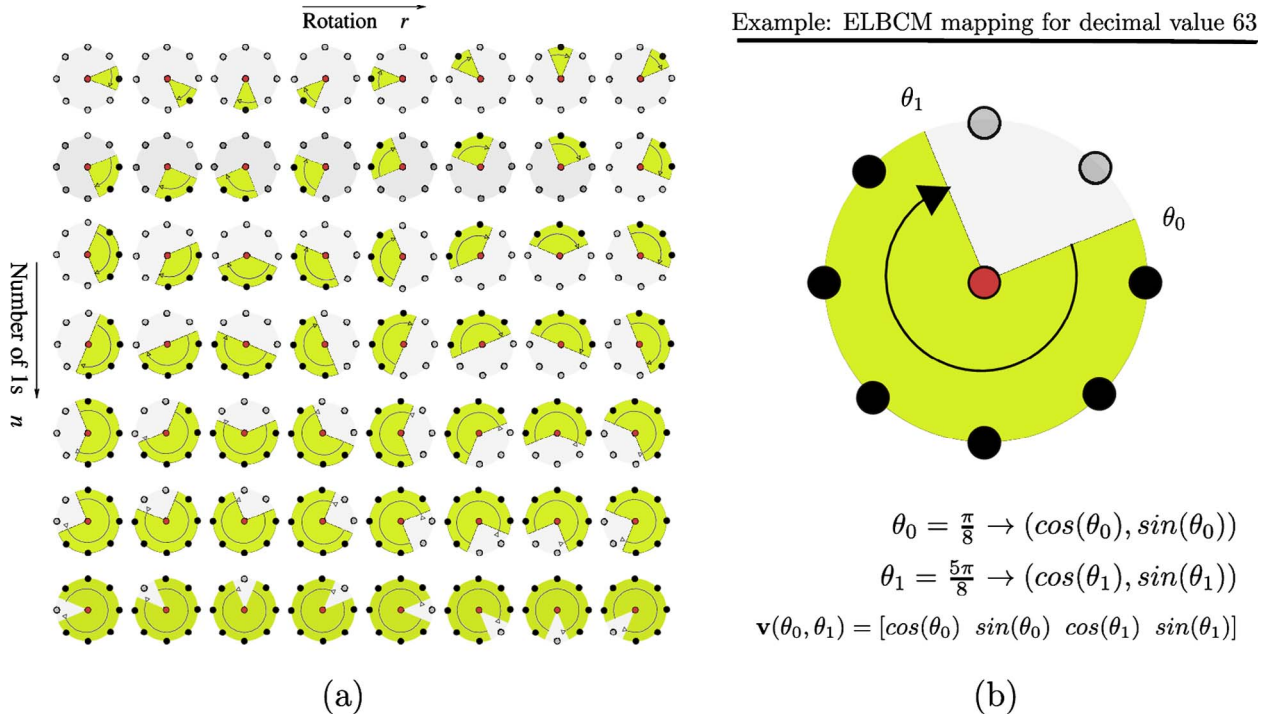


Fig. 2. (a) The set of 56 different uniform LBP patterns in a (8,*R*) neighborhood. Each *uLBP* is represented by a decimal value ∈ [1. .56]. For perfectly uniform patches of pixels, this value is 0. For each pattern a curved arrow indicates the start and end angles (θ_0 and θ_1). Note that the reference for the angular measure is the point located on the right of the central point. For uniform patches, $\theta_0 = 0$ and $\theta_1 = 0$. (b) In ELBCM, the uniform LBP pattern defined by two angles θ_0 and θ_1 is described by $\mathbf{v}(\theta_0,\theta_1) = [\cos(\theta_0)\sin(\theta_0)\cos(\theta_1)\sin(\theta_1)]$. The LBP pattern represented by decimal value 63 maps to $\mathbf{v}(\pi/8,5\pi/8) = [\cos(\pi/8)\sin(\pi/8)\cos(5\pi/8)\sin(5\pi/8)]$.

calculating $\mathbf{v}(\theta_0,\theta_1)$ is provided on Fig. 2(b) where the decimal value 63 (bit string 00111111) is mapped to the angles $\theta_0 = \pi/8$ and $\theta_1 = 5\pi/8$. The ELBCM descriptor is built using the following mapping function

$$\phi(I,x,y) = [\mathbf{S} \;\; \mathbf{C} \;\; \boldsymbol{\alpha}(LBP(x,y))]. \tag{6}$$

where $LBP(x,y)$ operator provides the pair of angles (θ_0,θ_1) at (x,y). **S** and **C** are the spatial and color features sub-vectors, classically $\mathbf{S} = [x,y]$ and **C** is the luminance $I(x,y)$. Local neighborhoods with non-uniform LBP patterns are ignored and their feature vectors do not contribute to the covariance matrix computation. Because $\boldsymbol{\alpha}$ is a four dimensional vector, the total dimension of $\phi(I,x,y)$ is classically $d = 7$, hence 7×7 for the covariance matrix.

The use of trigonometric formulas in (6) should make a naive implementation of ELBCM need much more computations than the other LBP-based methods, but this problem is completely overcome using a look-up table that maps directly the 56 nriLBP decimal numbers to the sinus and cosinus values of θ_0 and θ_1.

ELBCM has multiple advantages compared to previous covariance-based descriptors. It is more compact (7 components of ELBCM versus 11 components of LBCM), more stable *i.e.* it is less impacted by small rotations that reflect as changes on the angles θ_0 and θ_1 described by the nriLBP while for LBCM the same rotations irregularly affect the value of the bits in $LBP(x,y)$ depending on their position (see the example of Fig. 3). On the contrary, the impact of noise on ELBP does not depend on the impacted pixel. An additional problem of LBCM is observed in practice where non-positive-definite matrices appear more frequently. A possible explanation for this phenomenon is that bit-strings in the $LBP(x,y)$ operator can be very sparse in smooth regions. Finally, the size of ELBCM is completely independent of the number of *P* neighbors used in the LBP operator.

3.3. Color LBP and color ELBP

To some extent, covariance descriptors provide invariance to uniform luminance changes due to the subtraction of the mean feature. However, illumination variations are generally non-uniform [33] particularly when the region to be described is large. This can be improved by choosing color and texture features that are intrinsically insensitive to these phenomena, for example HSV (where H and S are invariant under white light). When a high descriptive power has to be reached, the region can be decomposed into overlapping blocks, each of them yielding a covariance descriptor [12,15,28] on which a photometric normalization can be performed [12]. Choi [34] proposes color LBP, where a classical LBP is extracted independently on each color channel, and [30] uses a similar approach using common colorspaces and invariants: RGB and its normalized version nRGB, HSV, Opponent, in a multiscale approach with dimensionnality reduction [30]. In order to better take into account the vectorial nature of color, the LCVBP for *Local Color Vector Binary Patterns* [35] proposes to describe the color LBP using two components (norm and angle) while the recent TCLBP [36] encodes the inter-channel information. In the following, four extensions of the Enhanced LBP are proposed.

ELBPa. The ELBP quartet defined in 3.2 is extracted independently on each color channel leading to a vector of 12 values.

ELBPb. Let be $\boldsymbol{\mu}$ the mean color vector of the region to be described, and $\boldsymbol{c}$ the color of a pixel. The vector going from $\boldsymbol{\mu}$ to $\boldsymbol{c}$ is extracted and its orthogonal projection is computed:

$$p(\boldsymbol{c}) = (\boldsymbol{c}-\boldsymbol{\mu})\cdot\boldsymbol{\mu} \tag{7}$$

p is a signed value: it is negative when $\boldsymbol{c}$ is located under the line passing through the color point $\boldsymbol{\mu}$ and perpendicular to the vector $\boldsymbol{\mu}$, it is positive otherwise. The nriLBP defined from p, noted pLBPb, is first computed, then the ELBPb quartet is deduced.

ELBPc. Here colors are ranked using the reduced ordering [37] which is a sub-ordering that uses one vector of the space as the reference to compute the distances. For ELBPc, the origin of the colorspace is selected as the reference color point. The LBP computed after reduced ordering is noted pLBPc.

ELBPd. This method is similar to ELBPb but the mean $\boldsymbol{\mu}$ is replaced by the color of the central pixel g_c of each LBP support. The

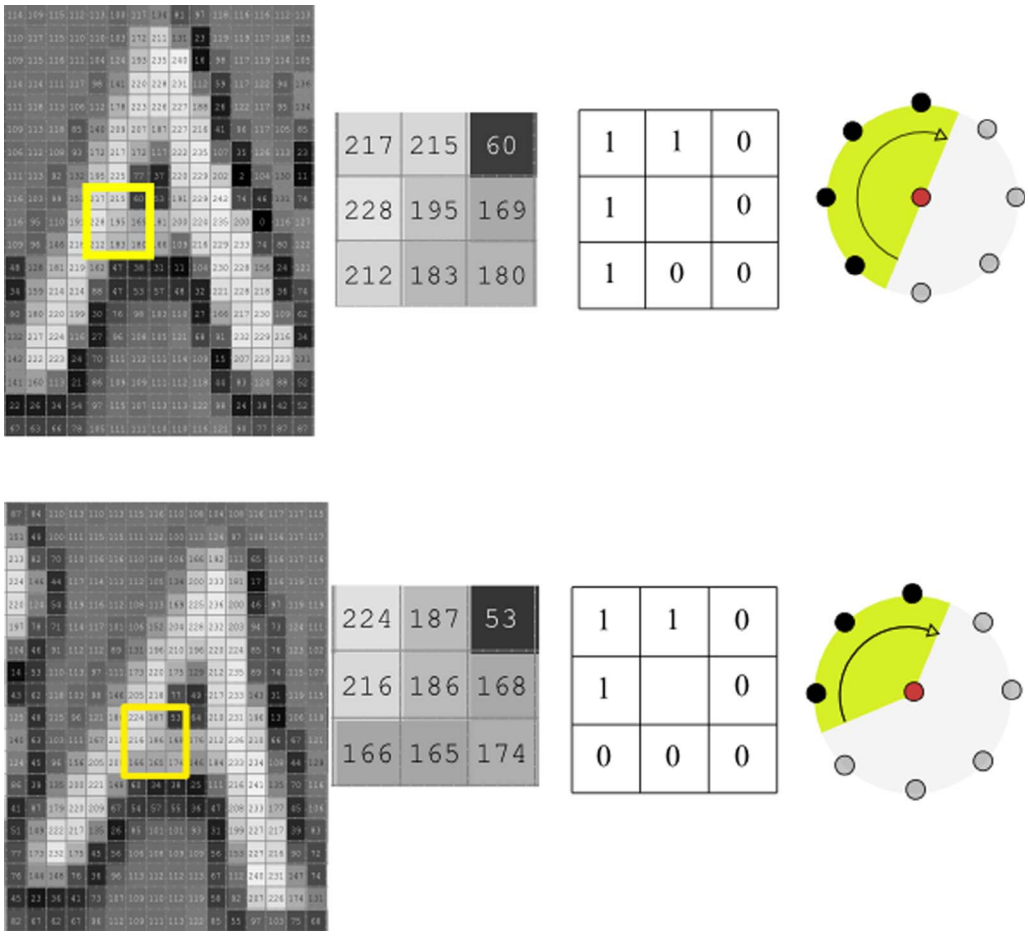


Fig. 3. Example of problem related to LBP. When a pattern is transformed by a small linear transformation such as a small rotation, the LBP values can be severely affected, here the LBP decimal value changes from 195 to 131 (a decrease of 32,8%) but the impact would be different if noise had appeared in another location. In contrast, the proposed angular representation used to build the ELBCM descriptor is more stable ($3\pi/8$ instead of $4\pi/8$ i.e. a decrease of 25%) and the impact does not depend on the location of the pixel in the neighborood.

corresponding nriLBP is noted pLBPd.

Fig. 4c shows three examples of different LBP computed using the three ways to rank colors. On Fig. 4c(a) the luminance is used. On Fig. 4c(b), colors are ranked according to their distance to the origin (as made in ELBPc). To finish on Fig. 4c(c), the signed distances are projected on the centroid color vector (as made in ELBPd).

The results are all different. The first pattern (Fig. 4c(a)) corresponds to a uniform LBP with $n = 3$ and $r = 6$ (using the notations of Fig. 2), the second one (Fig. 4c(b)) is also a uniform LBP with $n = 5$, $r = 5$, but the third pattern (Fig. 4c(c)) is not. To finish, the LBP values computed marginally on R, G and B are given on Figures from Fig. 4c (d)–(f).

3.4. Summary

The three main categories of features, namely space, color and texture, are respectively noted **S**, **C** and **T** in the following. The features and notations used in the paper are collected in Table 1. The spatial location is expressed either by the coordinates *xy* (in pixels) or by the distance to the center of the region (x_0,y_0) to be described so as to provide rotation invariance. Concerning color, the following spaces are tested: luminance *I*, RGB, *rgI* (*r* and *g* are the normalized channels and *I* is the luminance), the Gaussian color model, HSV, RGB with histogram equalization (*HE*), opponent space *Opp* and its normalized version *nOpp*. To finish, *L1* [39] is defined at each pixel as $max(r,g,b)$ which corresponds to the value associated to the most saturated color. Note that *L1*,*rg*,*Opp*,*nOpp* have the problem to be ill-defined for low saturated colors. Concerning texture, we investigate the luminance spatial first-order derivatives *Ix Iy*, the norm *n* and orientation *a* of the gradient, and its second derivatives *Ixx*,*Iyy*. To avoid increasing unnecessarily the number of features used to construct the covariance matrix, gradients are generally not computed for each color channel individually. However, this can be made when the separability power has to be improved, for person re-identification for example [15]. They are noted *Cx*,*Cy* in the table. The color derivatives can also be computed in a vectorial way by the DiZenzo method [38]. *Gabuv* denotes the Gabor responses [16] and *GabLBP* corresponds to its association with LBP, denoted GLRCD in [22]. To finish, the ELBP descriptors proposed in Section 3.3 are compared to existing versions of LBP: the variance and the classical (denoted LBCM [21]) or uniform LBP with or without rotation invariance.

4. Experiments

After introducing the datasets used in the experiments in Section 4.1, the different LBP descriptors are evaluated under illumination changes in Section 4.2. Texture classification, object matching and person re-identification are then successively addressed in Sections 4.3,4.5,4.3.1,4.4,4.5.

4.1. Image database

Three datasets are used to address texture classification, object retrieval and person re-identification.

KTH_TIPS [40]. This dataset is composed by ten different texture classes represented by 81 samples of size 200×200 pixels. Some of them are displayed on Fig. 5. This dataset is challenging because each texture class is represented by images taken at different scales, illuminations and poses.

ALOI [41]. The *Amsterdam Library of Object Images*[1] comprises 1000

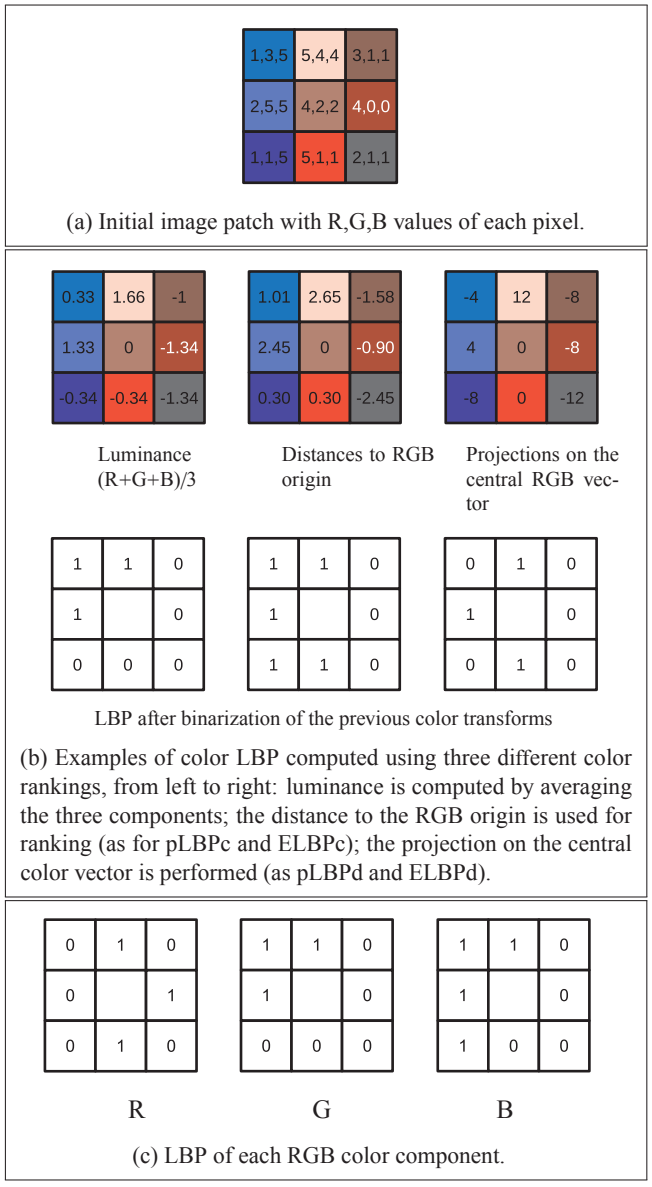


Fig. 4. Comparison between 3 color vectorial versions of LBP and the marginal LBP (a LBP for each component). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

various objects which are more or less colorful, structured and textured. Each object is pictured under different conditions: 12 different illumination colors (this dataset is called *ALOIcolor*) and 24 lighting directions (dataset called *ALOIangle*). The first row of Fig. 6 shows the examples of 10 first objects and the second row shows the first object under different illumination conditions (images from *ALOIangle*). This library is used in Section 4.2 to evaluate the invariance and disctinctness power of the proposed LBP features and in Section 4.4 to address object retrieval.

CAVIAR4REID [42]. This dataset has been extracted from the CAVIAR dataset to specifically evaluate person re-identification algorithms. It comprises 1220 images of 72 different pedestrians manually selected. 50 of them have been acquired by two cameras placed at two different locations. For each pedestrian, the images have been selected so as to show very different image resolutions (from 17×39 to 72×144), lighting conditions, poses and occlusions. Fig. 7 shows a few images of two pedestrians, which surely allows understanding the difficulty of the problem.

Using these datasets, the criterion used for evaluation is the recognition rate, *i.e.* the number of images that are well-labeled on the total number of images.

Table 1
Features used in the covariance descriptors.

	Space **S**
xy	$[x,y]$
d	$\sqrt{(x-x_0)^2+(y-y_0)^2}$
	Color **C**
I	Luminance *I*
RGB	$[R,G,B]$
HSV	[H, S, V]
Gaussian	$[E,E_\lambda,E_{\lambda\lambda}]$
L1	$max(r,g,b)$
HE	RGB + histo. equ.
rgI	normalized RG + I
Opp	Opponent color space
nOpp	normalized Opponent $[c_1,c_2,c_3,]$
	Texture **T**
Ix, Iy	gradient components $[I_x,I_y]$
n, a	gradient norm and angle $\left[\nabla^I=\sqrt{I_x^2+I_y^2},\phi^I=artan\frac{I_y}{I_x}\right]$
Ix, Iy, Ixx, Iyy	1^{st} and 2^{nd} derivatives $[I_x,I_y,I_{xx},I_{yy}]$
Di Zenzo	vectorial color gradient of [38]
Cx, Cy	marginal color gradients $[C1_x,C1_y,C2_x,C2_y,C3_x,C3_y]$
Gab uv	Gabor responses of the first u scales and v orientations [16] $[g_{00}(x,y)\cdots,g_{(u-1)(v-1)}]$
LBP	LBP decimal value ∈ [0,255] for $LBP_{8,1}$
Gab LBP	[Gab14 + LBP] as in [22]
uLBP	uniform LBP value ∈ [0,9] for $LBP_{8,1}$
nriLBP	non-rotation invariant uniform LBP ∈ [0,58] for $LBP_{8,1}$
ror LBP	rotation invariant LBP ∈ [0,255] for $LBP_{8,1}$
V	LBP variance of Eq. (4) as in [14]
pLBPb	color LBP using the mean color of the region to be described as the reference vector
pLBPc	color LBP using a reduced ordering using the origin of the colorspace as the reference point
pLBPd	color LBP defined by color ranking after projection on the central color vector
ELBPa	marginal ELBP: $\alpha_i(nriLBP_i)\ \forall$ channel$i$ = 1. .k
ELBPb	ELBP defined on pLBPb: $\alpha(pLBPb)$
ELBPc	ELBP defined on pLBPc: $\alpha(pLBPc)$
ELBPd	ELBP defined on pLBPd: $\alpha(pLBPd)$

4.2. LBP descriptors

The databases *ALOIcolor* and *ALOIangle* are useful to appraise the invariance of the descriptors under illumination variations, and therefore assess their descriptive or discriminative power. For each of the 20 first objects of the database, the intraclass and interclass correlations, respectively denoted $\mathscr{C}_{nn}$ and $\mathscr{C}_{nm}$, are computed as the average correlations on R, G and B channels. Then, the ratios of these two values are computed to indicate the trade-off between invariance and distinctness. The results are collected in Table 21. Obviously, the proposed LBP features get higher ratios than existing LBP. Generally speaking, this gain is obtained by a higher separability power (lower interclass correlation $\mathscr{C}_{nm}$) while the invariance inside a class is similar. Note that on *ALOIangle*, the intraclass correlation is lower than existing LBP, which tends to indicate that our descriptors are less invariant to rotation changes than to lighting variations. On the other hand, the distinctness remains very good. In the next section, the performance of the proposed features will be analyzed when included in a covariance region descriptor for texture classification.

[1] ALOI is available here: http://aloi.science.uva.nl/.

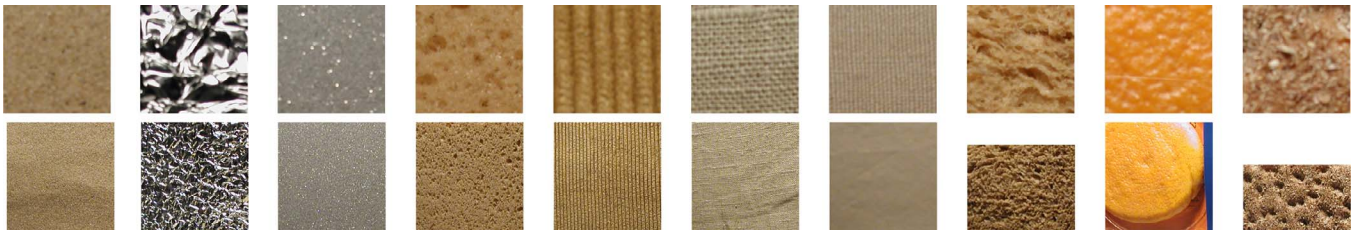


Fig. 5. A few samples from the KTH_TIPS_Color dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. A few samples from the ALOI datasets. First row: different objects categories. Second row: different illuminants and lighting angles of the first object.

4.3. Texture classification

For each class of the database *KTH_TIPS*, 40 images are selected randomly to constitute the training database (the same samples for each experiment), while the 41 remaining ones are kept for testing. A k-nearest-neighbor classification is performed using $k = 5$. In the experiments, each texture image is described by one single covariance matrix. The texture features are first computed on the luminance I in Section 4.3.1 and on color in Section 4.3.2.

4.3.1. Evaluation of gray texture descriptors

Every combination of **S**, **C** and **T** from Table 1 is tested, except that the texture features **T** are extracted using luminance information. For sake of concision, the recognition rate is averaged for each separate feature and displayed on Fig. 8; for instance, the result denoted *xy* corresponds to the average of all results envolving *xy*, for all color **C** and texture **T** features. Obviously, the Enhanced LBP produces the best results in terms of texture (yellow bars). Concerning color, the best averaged results are obtained when **C** = *rgI*. Note however that several colorspaces get very good results as well: RGB with histogram equalization (*HE*), the Gaussian colorspace, RGB, the Opponent space and the association of $L1$ with the luminance. HSV, nOpp and $L1$ gets bad results probably because colors are not saturated enough on this dataset. To finish, as shown by the poor performance obtained when luminance is used alone, color clearly has a positive impact on the classification results, at the cost of course of a larger feature vector.

4.3.2. Evaluation of color texture descriptors

Now the texture features are computed using color information. The Gabor features and the LBP variance *V* have not been included due to their limited performance in the previous tests (see Fig. 8). The marginal second order derivatives are not included neither because they produce worse results with similar size as ELBP (four components). On the contrary, LBP descriptors are kept due top their relatively good results w.r.t their compactness. The feature *d* is chosen as spatial cue **S** and two colorspaces are tested, RGB and *rgI* which have proven their good performance in the previous experiment. Fig. 9 shows the classification rates obtained for this dataset using most of the feature combinations enlisted in Table 1. Each of the four graphs is related to a different combination of colorspaces used for **C** and **T**. The yellow bars correspond to the existing LBP features, the orange ones are the pLBP, and the green ones correspond to the proposed ELBP features. All

Fig. 7. A few samples from the CAVIAR4REID database. First row: pedestrian 3. Second row: pedestrian 55.

Table 2
Intra and inter correlation of LBP features in RGB ($\times 10^{-2}$) on the ALOI database, and their ratio (the higher is the best).

	TEXTURE FEATURES											
	LBP	uLBP	rorLBP	nriLBP	LCVBP	pLBPb	pLBPc	pLBPd	ELBPa	ELBPb	ELBPc	ELBP
	ALOI COLOR DATABASE (20 FIRST OBJECTS)											
$\mathscr{C}_{nn}$	73.60	65.24	73.02	66.64	27.02	72.15	71.79	70.73	64.46	71.37	71.40	71.70
$\mathscr{C}_{nm}$	1.22	1.42	2.48	1.37	0.4	0.75	0.61	0.26	0.85	0.61	0.74	0.60
$\mathscr{C}_{nn}/\mathscr{C}_{nm}$	60.33	45.94	29.44	48.64	67.65	96.2	117.69	272.03	75.84	117	96.5	119.5
	ALOI ANGLE DATABASE (20 FIRST OBJECTS)											
$\mathscr{C}_{nn}$	62.43	62.08	88.93	61.94	21.50	46.93	44.17	35.62	51.89	49.36	48.46	51.43
$\mathscr{C}_{nm}$	1.02	1.44	2.27	1.42	0.36	0.82	0.71	0.38	0.70	0.47	0.45	0.58
$\mathscr{C}_{nn}/\mathscr{C}_{nm}$	61.21	43.11	39.18	43.62	59.72	57.23	62.21	93.74	74.13	105.02	107.69	88.67

texture features are computed using RGB.

Roughly speeking, the results are better when *rgI* is included (compare Fig. 9 (a) with (c) and, Fig. 9(b) with (d)). A clear conclusion is more difficult to reach for the color used to compute texture features since the results vary depending on the type of feature. ELBPa provides the best results (Fig. 9 (a)), and this remains true for most color combinations, except when RGB is used exclusively (on Fig. 9 (d)). The other color ELBP versions also figure prominently in the results, the recognition rate being always greater than 93%.

4.4. Object recognition

Using the ALOI datasets, experiments are performed using the first image of each class as the reference for matching. Therefore each test object is assigned to the class for which the Riemanian distance of the covariance matrices is the minimum.

Concerning *ALOIcolor* dataset, the tested features are **S** = *xy* and **C** = *rgI* or RGB. **T** = RGB, *rgI* or even *I* when considered relevant. The results are visible on Fig. 10(a), using the same color keys as on Fig. 9. Most of the displayed results correspond to **C** = *rgI* because RGB produces worse results. The version ELBPa (ELBP independently computed on R, G and B) collects the best performance for this dataset. Generally speaking, despite the challenging conditions, *i.e.* the large number of objects and only one reference object, the results are quite satisfactory (⩾82 %).

The recognition rates for *ALOIangle* are displayed on Fig. 10(b). For sake of concision, it only collects the results obtained for **S** = *xy*, **C** = *rgI* and **T** features computed on RGB. Here also ELBPa performs very well.

As for the previous experiments in Section 4.3, the vectorial versions ELBPb, ELBPc and ELBPd get quite similar performance, which are all better than the existing vectorial color version LCVBP [35].

4.5. Person re-identification

The problem here consists in retrieving a person, based on one or several observations made from different viewpoints, typically from two different cameras, or simply at different locations and/or

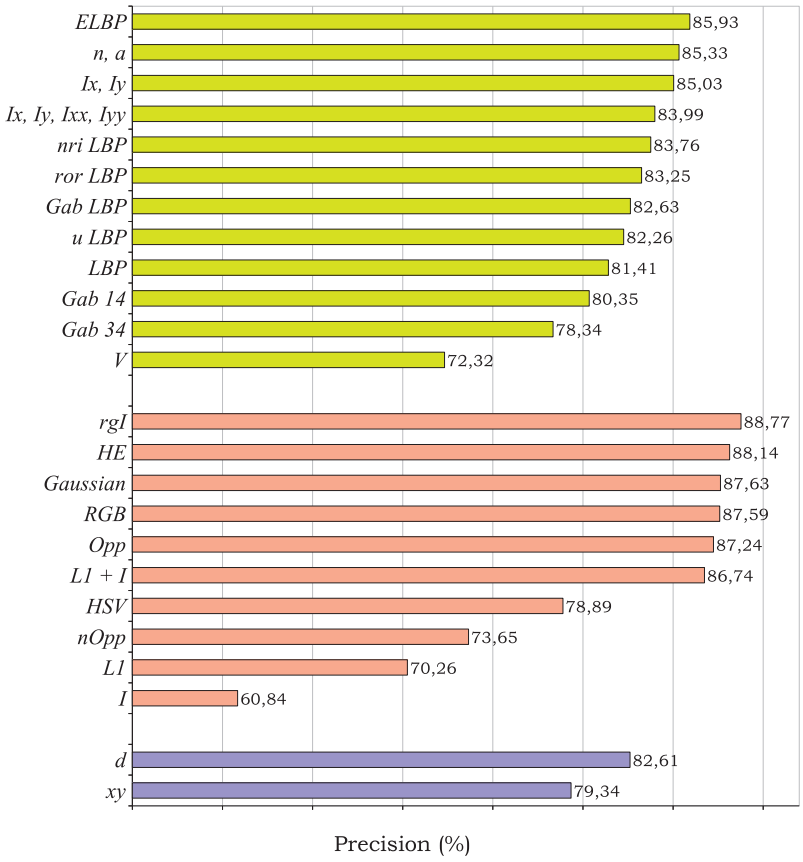
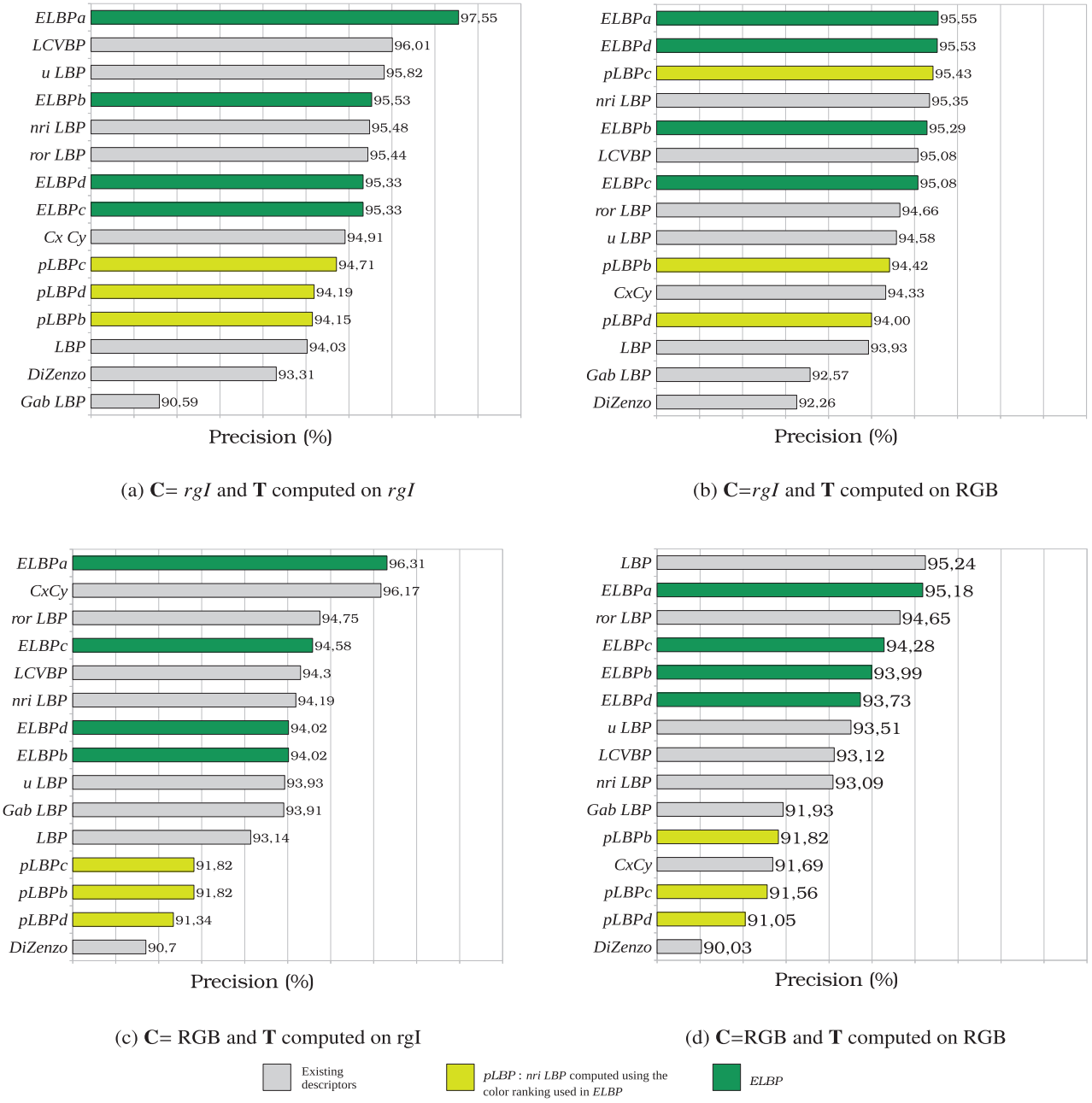


Fig. 8. Average recognition rate (%) for each feature on *KTH_TIPS* dataset (texture **T** in yellow, color **C** in orange and space **S** in blue). The texture features **T** are computed on luminance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) **C**= *rgI* and **T** computed on *rgI*

(b) **C**=*rgI* and **T** computed on RGB

(c) **C**= RGB and **T** computed on rgI

(d) **C**=RGB and **T** computed on RGB

Fig. 9. Classification rates for *KTH_TIPS* using color (RGB or *rgI*) to compute the texture features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

timestamps, after occlusion for instance. Because the distortions caused on the appearance can be very huge (see Fig. 7), it is far more complicated than in previous datasets to find good descriptors, *i.e* which well discriminate objects while being invariant to the acquisition conditions. The experiments are performed on the *CAVIAR4REID* dataset. Two covariance matrices are used for description, instead of one. Indeed, it is assumed that the trunk, the upper part of the body and the head convey the most relevant information to describe the person while legs, and arms are more prone to variations. Therefore, as illustrated on Fig. 11(*a*), a first Gaussian kernel is centered on the object at $(H/2,W/2)$ with parameters $\sigma_x = H/4$ and $\sigma_y = W/4$ where H and W are the dimensions of the region to be described. The second window has the same characteristics but is located at $(H/4,W/2)$ (upper part). Concerning features, $\mathbf{S} = xy$ because orientation is meaningful for person identification; **C** = RGB and rgI; the texture features **T** are computed using either RGB or I. All the LBP-based features are tested.

As shown by the results of Fig. 11(b), the performance are poor compared to the previous experiments (around 20 %) due to the difficulty of the task. Here the best results are obtained for **C** = RGB, because images are not extremely saturated. Here also, the use of ELBP features allows to improve the classification results compared to the existing features.

5. Discussions

The computation of the enhanced LBP is very fast because look-up tables can be used to make a direct correspondence between the uLBP values and the trigonometric quantities. However, four texture features

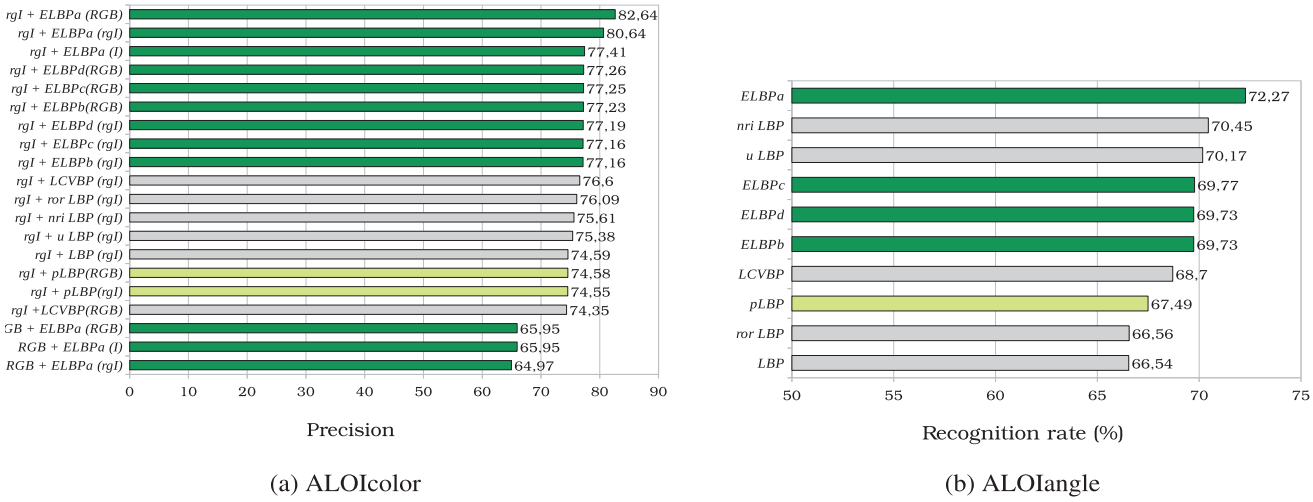


Fig. 10. Classification rates (%) for the ALOI datasets: (a) *ALOIcolor*, (b) *ALOIangle*.

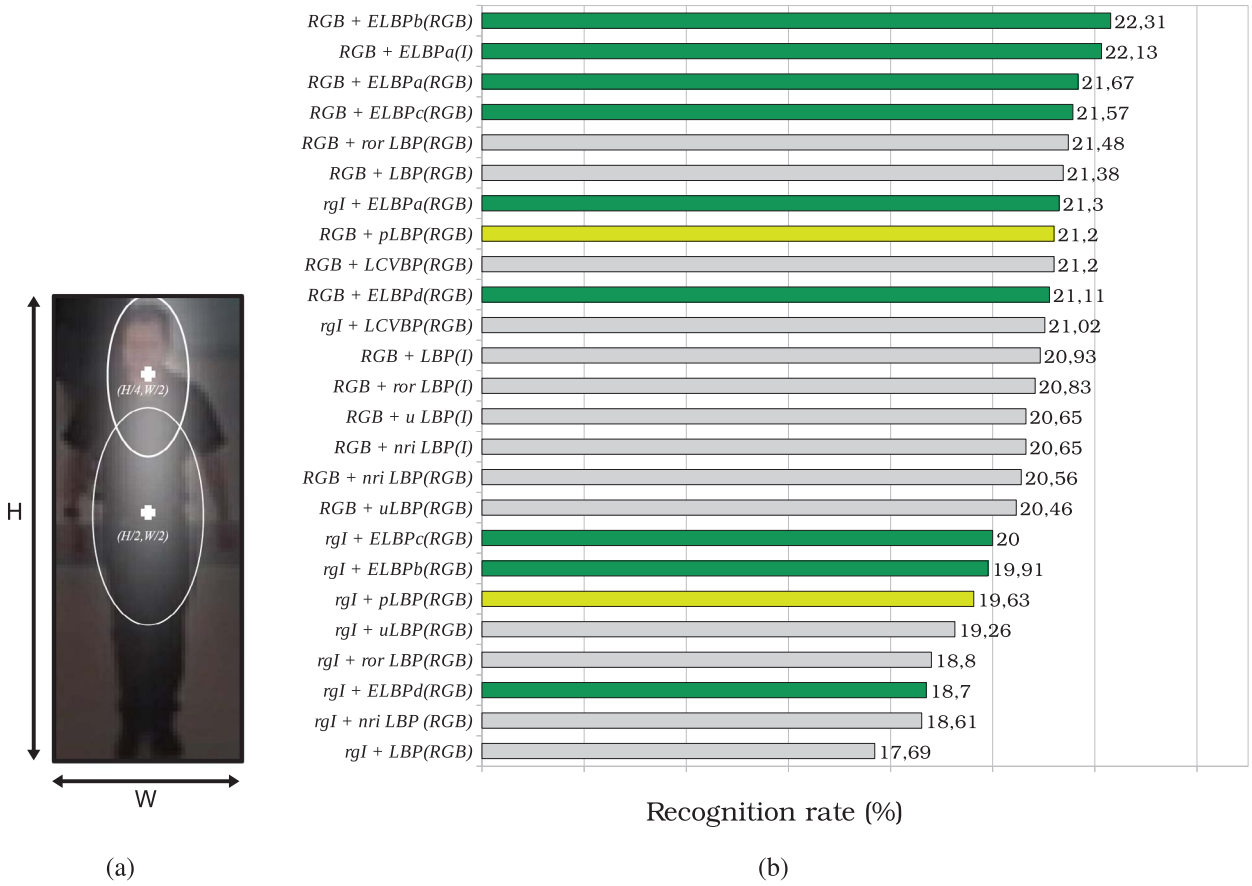


Fig. 11. Experiment on CAVIAR4REID. (a) Two covariance matrices are used to describe the person and, the features are weighted using Gaussian kernels. (b) Recognition rate for *CAVIAR4REID* database.

are used instead of one so the resulting feature vector is necessarily larger than for classical LBP descriptors. Consequently, the computation of the covariance matrix can take 50 % more complexity than when including uLBP, and around 25 % compared to LCVBP [35]. The performances are however improved on each dataset, with a gain that goes from 0.5% to 11 % with respect to LCVBP, depending on the dataset. Consequently, the choice made on the texture feature should depend on the application, whether the priority is put on the accuracy then ELBPa should be selected, or on the execution times or the memory resources then a nriLBP computed on luminance provides good results in most experiments. Note that the computation time of covariance matrices can significantly be reduced by using integral images, by choosing the adapted data structures and employing specific instructions, yielding real-time execution on small embedded processors [6] even for large covariance matrices such as ELBCM.

The Table 3 makes a sum-up of the results for the four datasets KTH_TIPS, ALOI color, ALOI angle and CAVIAR4REID. For 9 LBP versions, the table collects the maximum recognition rate (among the

Table 3
For each dataset, the table collects the maximum recognition rate obtained by using each of the 9 best LBP features, and the corresponding rank. For example, regarding sequence KTH_TIPS, the use of the classical LBP leads to a maximum recognition rate of 95.24 % (obtained for **C** = RGB and **T** computed on RGB, as seen on Fig. 9(d)). This is the 8th best results.

		TEXTURE FEATURES								
Dataset		LBP	uLBP	rorLBP	nriLBP	LCVBP	ELBPa	ELBPb	ELBPc	ELBPd
KTH_TIPS	Rate (%)	95.24	95.82	95.44	95.48	96.01	97.55	95.53	95.33	95.53
	Rank	8	3	7	6	2	**1**	4	9	4
ALOI Color	Rate (%)	74.59	75.38	76.09	75.61	76.6	82.64	77.23	77.25	77.26
	Rank	9	8	6	7	5	**1**	4	3	2
ALOI Angle	Rate (%)	66.54	70.17	66.56	70.45	68.7	72.27	69.73	69.77	69.73
	Rank	9	3	8	2	7	**1**	5	4	5
CAVIAR4REID	Rate (%)	21.38	20.65	21.48	20.65	21.2	22.13	19.91	21.57	21.11
	Rank	4	7	3	7	5	**1**	9	2	6

results of Figs. 9, 10 or 11) and their rank. The use of ELBPa leads to the higher recognition rate. However, it is difficult to predict which of the color versions is the most performant for a given dataset. The use of color for computing texture features improves significantly their discriminant power. Note the experiments have also shown that the choice on the colorspace used in the covariance matrix as well as the one used for LBP computation have a quite significant impact on the results. RGB provides very good results, but *rgI* improves them when images are saturated enough. To address these problems, the features can be automatically selected depending on the application, on the basis of a training dataset. This can be done in a similar way as in the COSMATI method [15] or specifically for color features [43]. When used for visual tracking in videos, the number of features can be reduced since temporal coherency helps reducing ambiguities. The features could also be selected automatically during time depending on the context [6].

6. Conclusions

This paper has proposed a new way to embed LBP in covariance region descriptors. From the nriLBP (non-rotation invariant uniform *LBP*), the angular portion corresponding to the '1' is described by their trigonometric values. Therefore the resulting *LBP* feature offers the same separability power as *nriLBP* with an increased resilience to small rotations or noise. In addition, the transform is fast because the cosine and sine values can be precomputed once for all and stored into a look-up table. Then, the extension to color is evaluated. While most existing works have proposed the concatenation of the three *LBP* features or the conversion to invariant colorspaces, one only work has been published up to now concerning vectorial extensions. The question is how to rank colors which are tridimensional values. Here, four different extensions are proposed, which differ on the color ranking, and several colorspaces are tested. A preliminary experiment has shown that the proposed features show a good trade-off between interclass separability and intraclass invariance.

The descriptors have been evaluated using three different datasets built for texture classification, object recognition and person re-identification. Several color LBP covariance descriptors are compared, as well as their gradient and wavelets-based versions. In most experiments, the highest recognition rates are provided by one of the ELBCM, which confirms the relevance of the proposals.

References

[1] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
[2] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[3] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded up robust features, Computer Vision–ECCV 2006, Springer, 2006, pp. 404–417.
[4] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, IEEE Comput. Vis. Pattern Recogn. Vol. 1 2005, pp. 886–893.
[5] O. Tuzel, F. Porikli, P. Meer, Region covariance: A fast descriptor for detection and classification, Computer Vision–ECCV 2006, 2006, pp. 589–600.
[6] F. Laguzet, A. Romero, M. Gouiffès, L. Lacassagne, D. Etiemble, Color tracking with contextual switching: real-time implementation on CPU, J. Real-Time Image Process. 10 (2) (2013) 403–422.
[7] J.Y. Tou, Y.H. Tay, P.Y. Lau, Gabor filters as feature images for covariance matrix on texture classification problem, Advances in Neuro-Information Processing, Springer, 2009, pp. 745–751.
[8] P. Li, Q. Wang, Local log-Euclidean covariance matrix (L2ECM) for image representation and its applications, Computer Vision, ECCV 2012, Springer, 2012, pp. 469–482.
[9] M.T. Harandi, C. Sanderson, A. Wiliem, B.C. Lovell, Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures, 2012 IEEE Workshop on Applications of Computer Vision (WACV), IEEE, 2012, pp. 433–439.
[10] S. Said, L. Bombrun, Y. Berthoumieu, Texture classification using Rao's distance on the space of covariance matrices, Geometric Science of Information, Springer, 2015, pp. 371–378.
[11] O. Tuzel, F. Porikli, P. Meer, Pedestrian detection via classification on Riemannian manifolds, IEEE Trans. Pattern Anal. Mach. Intell. 30 (10) (2008) 1713–1727.
[12] J. Yao, J.-M. Odobez, Fast human detection from joint appearance and foreground feature subset covariances, Comput. Vision Image Understand. 115 (10) (2011) 1414–1426.
[13] F. Porikli, O. Tuzel, P. Meer, Covariance tracking using model update based on lie algebra, IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 1 IEE, 2006, pp. 728–735.
[14] A. Romero, M. Gouiffès, L. Lacassagne, Covariance descriptor for multiple object tracking and re-identification with colorspace evaluation, in: Asian Conference on Computer Vision, 2012. ACCV 2012, 2012.
[15] S. Bak, F. Brémond, et al., Re-identification by covariance descriptors, Person Re-Identification (2013) 71.
[16] Y. Pang, Y. Yuan, X. Li, Gabor-based region covariance matrices for face recognition, IEEE Trans. Circ. Syst. Video Technol. 18 (7) (2008) 989–993.
[17] K. Guo, P. Ishwar, J. Konrad, Action recognition using sparse representation on covariance manifolds of optical flow, 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, 2010, pp. 188–195.
[18] Y.H. Habiboğlu, O. Günay, A.E. Çetin, Covariance matrix-based fire and flame detection method in video, Mach. Vis. Appl. 23 (6) (2012) 1103–1113.
[19] W. Förstner, B. Moonen, A metric for covariance matrices, Quo vadis geodesia (1999) 113–128.
[20] C. Zhu, C.-E. Bichot, L. Chen, Multi-scale color local binary patterns for visual object classes recognition, 2010 20th International Conference on Pattern Recognition (ICPR), 2010, pp. 3065–3068.
[21] S. Guo, Q. Ruan, Facial expression recognition using local binary covariance matrices, 4th IET International Conference on Wireless, Mobile & Multimedia Networks (ICWMMN 2011), 2011, pp. 237–242.
[22] Y. Zhang, S. Li, Gabor-LBP based region covariance descriptor for person re-identification, 2011 Sixth International Conference on Image and Graphics (ICIG), 2011, pp. 368–371.
[23] A. Romero, M. Gouiffès, L. Lacassagne, Enhanced Local Binary Covariance Matrices ELBCM for texture analysis and object tracking, IRAGE 2013, Berlin, Germany, ACM International Conference Proceedings Series, Association for Computing Machinery, 2013.
[24] F. Nielsen, R. Bhatia, Matrix Information Geometry, Springer, 2012.
[25] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2) (2004) 137–154.
[26] A. Sanin, C. Sanderson, M.T. Harandi, B.C. Lovell, Spatio-temporal covariance descriptors for action and gesture recognition, 2013 IEEE Workshop on Applications of Computer Vision (WACV), IEEE, 2013, pp. 103–110.
[27] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[28] B. Ma, Y. Su, F. Jurie, et al., Bicov: a novel image representation for person re-identification and face verification, British Machine Vision Conference, 2012.

[29] M. Hirzer, C. Beleznai, P.M. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification, Image Analysis, Springer, 2011, pp. 91–102.
[30] C. Zhu, C.-E. Bichot, L. Chen, Image region description using orthogonal combination of local binary patterns enhanced with color information, Pattern Recogn. 46 (7) (2013) 1949–1963.
[31] S. Bak, E. Corvee, F. Bremond, M. Thonnat, Boosted human re-identification using Riemannian manifolds, Image Vis. Comput. 30 (6) (2012) 443–452.
[32] M. Pietikainen, Computer Vision Using Local Binary Patterns, Computational Imaging and Vision, Springer, London, 2011.
[33] M. Gouiffès, C. Collewet, C. Fernandez-Maloigne, A. Trémeau, A study on local photometric models and their application to robust tracking, Comput. Vis. Image Understand. 116 (8) (2012) 896–907.
[34] J.Y. Choi, K. Plataniotis, Y.M. Ro, Using colour local binary pattern features for face recognition, 2010 17th IEEE International Conference on Image Processing (ICIP), 2010, pp. 4541–4544.
[35] S.H. Lee, J.Y. Choi, Y.M. Ro, K. Plataniotis, Local color vector binary patterns from multichannel face images for face recognition, IEEE Trans. Image Process. 21 (4) (2012) 2347–2353.
[36] Z. Lu, X. Jiang, A. Kot, A novel lbp-based color descriptor for face recognition, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 1857–1861.
[37] V. Barnett, The ordering of multivariate data, J. Roy. Stat. Soc. Ser. A (General) (1976) 318–355.
[38] S.D. Zenzo, A note on the gradient of multi-image, Comput. Vis., Graph., Image Process. 33 (1986) 116–125.
[39] A. Romero, M. Gouiffès, L. Lacassagne, Feature points tracking adaptive to saturation, 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2011, pp. 277–282.
[40] E. Hayman, B. Caputo, M. Fritz, J.-O. Eklundh, On the significance of real-world conditions for material classification, Computer Vision-ECCV 2004 (2004) 253–266.
[41] J. Geusebroek, G. Burghouts, A. Smeulders, The Amsterdam library of object images, Int. J. Comput. Vis. 61 (1) (2005) 103–112.
[42] D.S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification, British Machine Vision Conference (BMVC), 2011.
[43] R.R. Varior, G. Wang, J. Lu, T. Liu, Learning Invariant Color Features for Person Reidentification, IEEE Trans. Image Process. 25 (7) (2016) 3395–3410.