Pattern Recognition Letters
journal homepage: www.elsevier.com

Compact Color-Texture Description for Texture Classification

Fahad Shahbaz Khan[1], Rao Muhammad Anwer[2], Joost van de Weijer[3], Michael Felsberg[1], Jorma Laaksonen[2]

[1]*Computer Vision Laboratory, Linköping University, Sweden*
[2]*Department of Information and Computer Science, Aalto University School of Science, Finland*
[3]*Computer Vision Center, CS Dept. Universitat Autonoma de Barcelona, Spain*

ABSTRACT

Describing textures is a challenging problem in computer vision and pattern recognition. The classification problem involves assigning a category label to the texture class it belongs to. Several factors such as variations in scale, illumination and viewpoint make the problem of texture description extremely challenging. A variety of histogram based texture representations exists in literature. However, combining multiple texture descriptors and assessing their complementarity is still an open research problem. In this paper, we first show that combining multiple local texture descriptors significantly improves the recognition performance compared to using a single best method alone. This gain in performance is achieved at the cost of high-dimensional final image representation. To counter this problem, we propose to use an information-theoretic compression technique to obtain a compact texture description without any significant loss in accuracy. In addition, we perform a comprehensive evaluation of pure color descriptors, popular in object recognition, for the problem of texture classification. Experiments are performed on four challenging texture datasets namely, KTH-TIPS-2a, KTH-TIPS-2b, FMD and Texture-10. The experiments clearly demonstrate that our proposed compact multi-texture approach outperforms the single best texture method alone. In all cases, discriminative color names outperforms other color features for texture classification. Finally, we show that combining discriminative color names with compact texture representation outperforms state-of-the-art methods by 7.8%, 4.3% and 5.0% on KTH-TIPS-2a, KTH-TIPS-2b and Texture-10 datasets respectively.

1. Introduction

Classifying textures is a difficult problem in computer vision and pattern recognition. The task is to associate a class label to its respective texture category. In recent years, a variety of texture description approaches have been proposed (Ojala et al., 2002; Guo et al., 2010; Lazebnik et al., 2005; Chen et al., 2010; Guo et al., 2012; Zhao et al., 2012; ul Hussain and Triggs, 2012; Varma and Zisserman, 2010). These approaches can be divided into two categories, namely sparse and dense representations. The sparse representation works by detecting feature points either based on interest point or dense sampling strategy. Feature description is then performed on these sampling points (Lazebnik et al., 2005; Zhang et al., 2007). The second strategy, dense representations, involves extracting local features for each pixel in an image (Ojala et al., 2002; Guo et al., 2010; Chen et al., 2010). In this paper, we investigate the problem of texture classification using dense local texture representations.

A variety of texture description approaches exist in literature (Ojala et al., 2002; Guo et al., 2010; Lazebnik et al., 2005; Chen et al., 2010; Guo et al., 2012; Zhao et al., 2012). One of the most successful approaches is that of Local Binary Patterns (LBP) (Ojala et al., 2002) based image representations. Other than texture classification, LBP have been successfully employed to solve other vision problems as well, such as object detection (Zhang et al., 2011), face recognition (Ahonen et al., 2004) and pedestrian detection (Wang et al., 2009). LBP describes the neighbourhood of a pixel by its binary derivatives which are used to form a short code to describe the pixel neighbourhood. A variety of LBP variants have been proposed (Guo et al., 2010; Ylioinas et al., 2013, 2012). Combining multiple texture features, such as variants of LBP features, is still an open research problem. The work of Guo et al. Guo et al. (2012) proposes a learning framework to combine variants of LBP features for texture classification. Tan and Triggs (Tan and Triggs, 2007) propose to combine Gabor wavelets and LBP features for the problem of face recognition. In this paper, we propose to use a heterogeneous feature set by combining multiple texture description methods.

Combining multiple texture description methods have an inherent problem of high-dimensional final image representations. Recently, Elfiky et al. Elfiky et al. (2012) proposed to use a divisive information theoretic clustering (DITC) method (Dhillon et al., 2003) to counter the problem of high-dimensionality of bag-of-words based spatial pyramid representations. The DITC compression was shown to reduce the dimensionality of image representations without any significant loss in accuracy. Similar to the work of Elfiky et al. (2012), we propose to use the DITC approach to compress the high-dimensional multi-texture representation. However, different to the work of Elfiky et al. (2012), here we investigate compressing a multi-texture histogram to obtain a single heterogeneous texture representation.

Generally, state-of-the-art texture descriptors operate on grey level images thereby ignoring the color information. Color in combination with shape features has been shown to yield excellent results for object recognition (van de Sande et al., 2010; Khan et al., 2012b, 2013c), object detection (Khan et al., 2012a) and action recognition (Khan et al., 2013a). Color description is a challenging problem due to significant variations in color caused by changes in illumination, shadows and highlights. Recent works have shown that an explicit color representation improves the performance for object recognition (Khan et al., 2012b, 2013c), object detection (Khan et al., 2012a) and action recognition (Khan et al., 2013a). In this paper, we perform a comprehensive evaluation of pure color descriptors, popular in object recognition, for the task of texture classification.

Contributions: We first show that combining multiple texture description methods significantly improves the performance compared to using the single best texture method alone. We further propose to use information theoretic compression approach to compress high-dimensional multi-texture features into a compact heterogeneous texture representation. Finally, we provide a comprehensive evaluation of color features, popular in object recognition, for the task of texture classification. This paper extends our earlier work (Khan et al., 2013b) for texture classification that only evaluated the contribution of color for texture recognition. Beyond the work in Khan et al. (2013b), we here investigate the problem of combining multiple local texture descriptors for robust texture description. We perform extensive experiments on four challenging texture datasets namely, KTH-TIPS-2a, KTH-TIPS-2b, FMD and Texture-10.

The results of our experiments clearly demonstrate that combining multi-texture descriptors significantly improves the performance compared to the single best method alone. We further show that multi-texture representations can be compressed efficiently without any significant loss in accuracy. Finally, our comprehensive evaluation of color features suggest that discriminative color names outperforms other color descriptors for texture recognition. By combining the best color descriptor with our compact heterogenous texture representation provides state-of-the-art results on three of the four texture datasets.

The paper is organized as follows. In Section 3 we investigate the problem of combining multiple texture descriptors. A comprehensive evaluation of pure color descriptors for texture description is provided in Section 4. In Section 5 we provide experimental results. Section 6 finishes with concluding remarks.

2. Related Work

A variety of texture description approaches have been proposed in recent years (Ojala et al., 2002; Guo et al., 2010; Lazebnik et al., 2005; Chen et al., 2010; Guo et al., 2012; Zhao et al., 2012; Ylioinas et al., 2013; Leung and Malik, 2001). Varma and Zisserman (2010) propose a statistical approach for texture modeling using the joint probability distribution of filter responses. A multiresolution approach based on local binary patterns (LBP) is proposed by Ojala et al. (2002) for gray-scale and rotation invariant texture classification. The LBP is one of the most successful approaches for texture classification with several variants existing in literature (Guo et al., 2010; Ylioinas et al., 2013, 2012). Chen et al. (2010) propose a method based on Weber's law consisting of two components namely differential excitation and orientation. An image is represented by concatenating the two components in a single representation. ul Hussain and Triggs (2012) introduce an approach that uses lookup-table based vector quantization for texture description. A set of low and mid-level perceptually inspired image features are proposed by Sharan et al. (2013) for texture classification.

Combining multiple texture representations for robust classification (Guo et al., 2012; Li et al., 2004; Tan and Triggs, 2007; Hong et al., 2014) is an interesting problem. The work of Tan and Triggs (2007) combines Gabor wavelets and LBP for the problem of face recognition. Ylioinas et al. (2011) combine contrast information together with local binary patterns for improved gender classification. A combination of HOG, LBP and Gabor features is used by Li et al. (2013) for gender classification. To counter the dimensionality of the proposed image representation, Partial Least Squares (PLS) is used to learn a low-dimensional representation. Hong et al. (2014) propose a numerical variant of LBP which is efficient and rotation invariant. The method is combined with other cues by a covariance matrix. Guo et al. (2012) propose a learning framework to fuse a variety of LBP variants such as conventional LBP, rotation invariant patterns, local patterns with anisotropic structure, completed local binary patterns and local ternary patterns. Similar to Guo et al. (2012), we investigate the problem of combining multiple texture description approaches. However, instead of only combining LBP variants (Guo et al., 2012), we here investigate fusing multiple texture descriptors to obtain a single heterogeneous texture representation.

A variety of color description approaches have been proposed in the field of object and scene recognition (Gevers and Smeulders, 1999; Bosch et al., 2006; van de Weijer and Schmid, 2006; van de Sande et al., 2010; Khan et al., 2012b, 2013c). Bosch et al. (2006) propose to compute SIFT descriptors directly on HSV channels for image classification. A comprehensive evaluation of color descriptors is performed by van de Sande et al. (2010). It has been shown that using an explicit color descriptor significantly improves the performance for object recognition (Khan et al., 2012b, 2013c), object detection (Khan et al., 2012a), texture recognition (Khan et al., 2013b) and action recognition (Khan et al., 2013a). In this work, we perform a comprehensive evaluation of pure color descriptors, popular in object recognition, for the problem of texture classification.

3. Combining Multiple Texture Descriptors

Here we present our framework of combining multiple texture features and obtaining a compact heterogeneous texture representation. We combine five texture descriptors namely, completed local binary patterns (Guo et al., 2010), WLD descriptor (Chen et al., 2010), binary Gabor pattern (Zhang et al., 2012), local phase quantization descriptor (Rahtu et al., 2012) and binarized statistical features (Kannala and Rahtu, 2012). We start by providing a brief overview of the five texture descriptors used in this work.

Completed local binary patterns (Guo et al., 2010): The completed local binary patterns (CLBP) extends the conventional LBP operator by incorporating local difference sign-magnitude transform information (LDSMT)[1]. The LDSMT further consists of two components, namely the difference sign and difference magnitude encoded by a binary code. Likewise the conventional LBP, a region is also represented by its center pixel encoded by a binary code after global thresholding. The final image representation is obtained by concatenating the three binary code maps to form a single histogram.

WLD descriptor (Chen et al., 2010): The WLD descriptor is inspired by Weber's Law and encodes both differential excitations and orientations at locations. The first component, differential excitation, captures the ratio between the intensity difference of a pixel with its neighbors and the intensity of the current pixel. The second component captures the gradient orientation of the current pixel.

Binary Gabor patterns (Zhang et al., 2012): The binary Gabor patterns (BGP) is a rotation invariant texture descriptor. Unlike MR8 filters (Varma and Zisserman, 2010), BGP uses predefined rotation invariant binary patterns and does not require a pre-training phase to learn a texton dictionary. Unlike LBP, where each sign is binary coded from the difference of two single pixels, BGP adopts the difference of regions to counter the noise sensitivity problem.

Local phase quantization (Rahtu et al., 2012): The local phase quantization (LPQ) descriptor works by quantizing the phase information of the Fourier transform and is robust to image blur. To counter the problem of heavy image blur, the approach uses short-term Fourier transform with a uniform function. A data correlation scheme is also incorporated into the descriptor which plays a crucial role in case of a sharp image. The LPQ descriptor is shown to provide excellent results for both texture and face recognition tasks.

Binarized statistical descriptor (Kannala and Rahtu, 2012): The binarized statistical image feature (BSIF) represents each pixel by a binary code. These binary codes are constructed by learning a set of basis vectors from natural images using independent component analysis and an efficient scalar quantization scheme. The number of basis vectors determines the length of the pixel binary codes used to construct the final histogram of an image.

In our approach, each image is represented by the five aforementioned texture description methods. The final representation is obtained by concatenating all five texture representations into a single histogram, $H = [h_{t1}, h_{t2}, h_{t3}, h_{t4}, h_{t5}]$. This multi-texture histogram is then input to the classifier for texture classification.

3.1. Compact Multi-Texture Representation

The multi-texture representation has the disadvantage of being high-dimensional (more than 3k of size) for an image. This is problematic as it significantly increases the computational time and memory usage in the classification stage. Recently, Elfiky et al. (2012) proposed a compression approach using the DITC algorithm (Dhillon et al., 2003) to counter the high-dimensionality issue of the bag-of-words based spatial pyramid representation. In this work, we also use the same underlying approach to compress the high-dimensional multi-texture representation. However, the difference with the work of Elfiky et al. (2012), is that here we investigate the DITC algorithm to solve the problem of compressing multi-texture histogram to obtain a single heterogeneous texture representation.

The DITC algorithm has been shown to obtain excellent results in reducing large histograms to compact ones. The algorithm is designed to find a fixed number of clusters that minimize the loss in mutual information between clusters and the category labels of training images. The DITC algorithm works on the class-conditional distributions over the texture histograms. The class-conditional estimation is measured by the probability distributions $p\left(R|h\right)$, where $R = \{r_1, r_2, ..., r_O\}$ is the set of O classes. The DITC algorithm works by estimating the drop in mutual information I between the histogram H and the class labels R. The transformation from the original texture histogram H to the new representation $H^R = \{H_1, H_2, ..., H_J\}$ (where every H_j represents a group of words in the original uncompressed histogram) is equal to

$$\Delta I = I\left(R; H\right) - I\left(R; H^R\right) \\ = \sum_{j=1}^{J} \sum_{h \in H_j} p\left(h\right) KL(p(R|h), p(R|H_j)), \tag{1}$$

where KL is the Kullback-Leibler divergence between the two distributions defined by

$$KL(p_1, p_2) = \sum_{z \in Z} p_1(z) \log \frac{p_1(z)}{p_2(z)}. \tag{2}$$

The multi-texture histogram bins with similar discriminative power are merged together over the classes. For more details, we refer to Dhillon et al. (2003) on the DITC algorithm.

[1]We experimented with different variants of LBP and found CLBP to provide superior performance.

Fig. 1. Example images from the four texture datasets, KTH-TIPS-2a, KTH-TIPS-2b, FMD and Texture-10, used in our experiments.

4. Combining Color and Texture

There exist two main strategies namely, early and late fusion, to combine color and texture information (Maenpaa and Pietikainen, 2004; Khan et al., 2013b). Early fusion works by computing texture descriptor on the color channels. In this way, a joint color-texture representation is obtained that combines the two cues at the pixel-level. Early fusion based image representation has the advantage of being more discriminative since the two cues are combined at the pixel level. However, early fusion representations suffers from the problem of high dimensionality.

Contrary to early fusion, late fusion combines the two cues at the image level. A separate histogram is constructed for color and texture. The two visual cues are then combined by concatenating the separate histograms into a single representation. The late fusion approach has shown to provide superior results for texture recognition (Maenpaa and Pietikainen, 2004; Khan et al., 2013b), object recognition (Khan et al., 2012b), object detection (Khan et al., 2012a) and action recognition (Khan et al., 2013a). Therefore, in this work, we use late fusion scheme for combining color and texture information. Next, we provide an overview of pure color descriptors.

4.1. Pure Color Descriptors

Here, we provide a brief overview of the pure color descriptors, popular in object recognition, for the problem of texture description.

RGB histogram (van de Sande et al., 2010): We use the standard RGB descriptor as a baseline. The RGB histogram is constructed by combining the three histograms from the R, G and B channels. The descriptor has 45 dimensions.

rg histogram (van de Sande et al., 2010): The rg histogram is based on the normalized RGB color model. The descriptor is 45 dimensional. It is invariant to light intensity changes and shadows.

Opponent-angle histogram (van de Weijer and Schmid, 2006): Unlike other pure color descriptors based on the (transformed) RGB values of the image, the opponent-angle histogram is constructed based on image derivatives. The histogram has 36 dimensions.

HUE histogram (van de Weijer and Schmid, 2006): The HUE descriptor was proposed by van de Weijer and Schmid (2006) and consists of 36 dimensions. In this descriptor, the hue is weighted by the saturation of a pixel in order to counter the instabilities in hue.

Transformed Color Distribution (van de Sande et al., 2010): The transformed color descriptor is derived by normalizing each channel of RGB histogram. The descriptor has 45 dimensions. It is invariant with respect to scale and light intensity.

Color Moments and Invariants (van de Sande et al., 2010): In the work of van de Sande et al. (2010), the color moment histogram is constructed by using all generalized color moments up to the second degree and the first order. The color moment invariants are constructed using generalized color moments. The color moments histogram has 36 dimensions whereas the color moment invariants has 24 dimensions.

Hue-saturation descriptor: The hue-saturation descriptor is invariant to luminance variations. The histogram has 36 dimensions (nine bins for hue times four for saturation).

Color names (van de Weijer et al., 2009): Most of the color descriptors discussed above are designed to achieve photometric invariance. Instead, color names descriptor aims at providing a certain degree of photometric invariance with discriminative power. The color names are used in daily life by humans to communicate color, such as "black", "blue" and "orange". Here, we use the 11 dimensional color names mapping learned from the Google images by van de Weijer et al. (2009).

Discriminative color descriptors (Khan et al., 2013c): The discriminative color descriptors by Khan et al. (2013c) take an information theoretic approach to the problem of color description. The method works by clustering color values together based on their discriminative power with an objective function to minimize the drop of mutual information of the final color representation. In this work, we use the three universal color representations with 11, 25 and 50 dimensions, respectively.

5. Experimental Results

To validate the performance of the proposed framework, we use four challenging datasets, namely KTH-TIPS-2a, KTH-

TIPS-2b, FMD and Texture-10. The KTH-TIPS-2a dataset consists of 11 texture categories with images at 9 different scales, 3 poses and 4 different illumination conditions. We use the standard protocol (Caputo et al., 2005; Sharma et al., 2012; Chen et al., 2010) by reporting the average classification performance over the 4 test runs. In each time, all the images from 1 sample are used for test while the images from the remaining 3 samples are used as a training set. The KTH-TIPS-2b dataset also consists of 11 texture categories. Here, for each test run, all images from 1 sample are used for training while all the images from remaining 3 samples are used for testing. The FMD dataset consists of 10 texture categories with 100 images (Sifre and Mallat, 2013; Sharan et al., 2013) for each class where 50 images are used for training and 50 for testing. The Texture-10 dataset consists of 10 different texture categories (Khan et al., 2013b) where 25 images per class are used for training and 15 for testing. Figure 1 shows example images from the four texture datasets.

Throughout our experiments, we use one-versus-all SVM using the χ^2 kernel (Zhang et al., 2007). Each test instance is assigned the category label of the classifier giving the highest response. The final classification score is obtained by calculating the mean recognition rate per category.

Table 1. Classification accuracy (%) of different texture representations on four texture datasets. In all cases, combining multi-texture representations significantly improves the performance compared to the single best texture method.

Method	Dimension	KTH-TIPS-2a	KTH-TIPS-2b	FMD	Texture-10
CLBP Guo et al. (2010)	1944	76.1 ±5.6	61.5 ±2.3	43.6	76.9
WLD Chen et al. (2010)	512	68.5 ±5.1	56.0 ±2.8	43.8	74.7
BGP Zhang et al. (2012)	216	76.8 ±4.9	63.3 ±3.4	43.2	66.0
LPQ Rahtu et al. (2012)	256	67.7 ±5.6	54.4 ±2.7	41.0	75.3
BSIF Kannala and Rahtu (2012)	256	70.0 ±5.7	54.3 ±2.8	34.4	66.0
CLBP + WLD	2456	78.1 ±4.8	63.7 ±2.8	46.6	77.8
CLBP + WLD + BGP	2672	79.2 ±5.1	65.1 ±2.3	48.1	78.6
CLBP + WLD + BGP + LPQ	2928	79.9 ±4.9	67.6 ±2.6	49.5	78.9
CLBP + WLD + BGP + LPQ + BSIF	3184	**80.8 ±5.3**	**68.9 ±2.9**	**51.6**	**79.3**

Table 2. Classification accuracy (%) obtained when using the original high-dimensional texture and compact texture representations. Note that the compression method reduces the dimensionality with little or no loss in accuracy.

Method	Dimension	KTH-TIPS-2a	KTH-TIPS-2b	FMD	Texture-10
Original Texture Feature	3184	80.8 ±5.3	68.9 ±1.7	**51.6**	**79.3**
Compact Texture (DITC)	**500**	**82.2 ±5.4**	**69.0 ±1.6**	49.0	78.0

5.1. *Experiment 1: Combining Texture Features*

We start by providing results for multi-texture representations. The results are presented in Table 1. For the CLBP descriptor, we use multiple radius values since it was shown to improve the performance compared to using a single radius value. On the KTH-TIPS-2a and Texture-10 datasets, CLBP provides the best performance compared to other single texture features. Among the five texture descriptors, the best results are achieved when using the BGP descriptor on the KTH-TIPS-2b dataset and WLD descriptor on the FMD dataset. In case of FMD and KTH-TIPS-2b datasets, the BSIF descriptor alone provides inferior results compared to other four texture descriptors. However, the performance still improves by 2.1% and 1.3% respectively on these datasets by adding the BSIF descriptor.

Combining the five texture representations in a single representation significantly improves the performance on all datasets. On the KTH-TIPS-2a dataset, a significant gain of 4.0% is obtained by combining multiple features compared to the single best representation. Similarly, gains of 5.6%, 7.8% and 2.4% are obtained by combining multiple texture features on the KTH-TIPS-2b, FMD and Texture-10, respectively. The results clearly suggest that different texture representations possess complementary information and should be combined to obtain a significant performance boost.

5.2. *Experiment 2: Compact Multi-Texture Features*

As discussed above, combining multi-texture representations improve the overall performance. However, this performance improvement comes at the price of high dimensionality. Here, we present the results obtained, using the approach described in Section 3.1, to compress the high dimensional multi-texture representation.We fix the final dimension of our multi-texture representation to 500.

Table 2 shows the results obtained on the four texture datasets. The DITC compression method reduces the dimensions from 3184 to 500 without any significant loss in accuracy. Surprisingly, on the KTH-TIPS-2a and KTH-TIPS-2b datasets, the low-dimensional compact representation improves the performance compared to the original representation. This demonstrates that the DITC method removes the redundancy while increasing the discriminative power in certain cases such as KTH-TIPS-2a and KTH-TIPS-2b datasets.

We also compared our texture compression approach with the discriminative texture feature selection method (Guo et al., 2012) on Texture-10 dataset. The method (Guo et al., 2012) learns a selection of LBP patterns based on robustness, discriminative power and representation of the features. We use the same feature representation (CLBP), having rotation invariance and a pixel neighborhood of 16, for both compression methods. The original representation is reduced to 500 using the two compression methods. The original feature representation with 8k dimensions provides a recognition rate of 71.3%. The feature selection method (Guo et al., 2012) obtains a classification rate of 70.0%. Our DITC based compression method improved the performance by providing an accuracy of 72.6%.

Additionally, we also compare the DITC compression method with conventional approaches for very low-dimensional representations. We compare with standard compression methods namely, PCA, PLS and Diffusion maps. Figure 2 shows results obtained using different compression techniques on the FMD and Texture-10 datasets. The three compression methods, PCA, PLS and Diffusion maps provide inferior performance on both datasets. The DITC based compression method significantly outperforms other compression methods even for very compact texture representations.

5.3. *Experiment 3: Pure Color Descriptors*

Here, we provide results of our comprehensive evaluation of color descriptors for texture recognition. Table 3 shows the results obtained using different color description methods on the four texture datasets. On the KTH-TIPS-2a and KTH-TIPS-2b datasets, RGB descriptor provides a recognition score of 55.5%

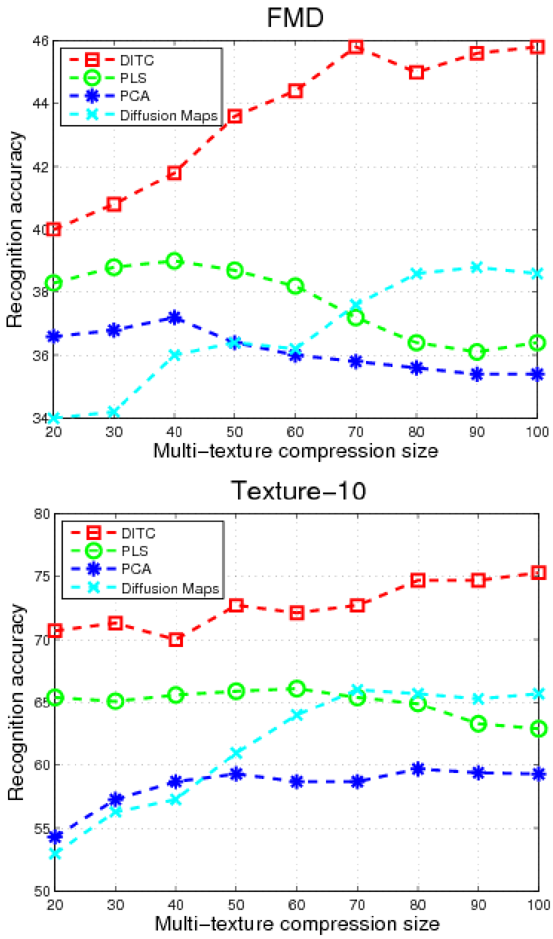

Fig. 2. Classification accuracy (%) obtained by compressing the multi-texture representation using different compression methods. Top row: results on the FMD dataset. Bottom row: results on the Texture-10 dataset. The best results are obtained using the DITC based compression technique.

and 42.1% respectively. The conventional color names provides a classification performance of 56.8% and 44.2% respectively. The best results are obtained using discriminative color descriptors with 50 dimensions. Similarly, on the FMD and Texture-10 datasets, the discriminative color names with 50 dimensions provide the best performance.

The results clearly demonstrate the effectiveness of using a discriminative color description approach that aims at maximizing the discriminative power while maintaining a certain degree of photometric invariance. Therefore, we select the discriminative color descriptors with 50 dimensions as an explicit color representation. In our final experiment, we combine the discriminative color descriptors with our proposed compact texture representation. The texture and color representations are concatenated in a late fusion manner which is then input to the classifier.

5.4. Comparison with State-of-the-art

Table 4 shows a comparison with state-of-the-art approaches on four texture datasets. On the KTH-TIPS-2a dataset, the method of Sharma et al. (2012) based on local-high-order statistics provides a classification accuracy of 73.0%. The approach by Lee et al. (2012) based on local color vector binary patterns

Table 3. Comparison (%) of pure color descriptors on four texture datasets. Note that the best performance is obtained by using discriminative color names with 50 dimensions.

Method	Dimension	KTH-TIPS-2a	KTH-TIPS-2b	FMD	Texture-10
RGB	50	55.5 ±5.8	42.1 ±1.8	20.3	52.3
rg	30	54.3 ±6.2	43.3 ±2.3	22.2	52.7
HUE	36	53.3 ±6.1	43.1 ±2.1	21.6	50.7
Opp-angle	36	50.1 ±6.2	45.4 ±1.7	17.4	34.0
Transformed color	45	52.8 ±5.3	44.8 ±1.8	23.0	40.0
Color moments	30	54.9 ±5.7	45.1 ±1.6	26.0	50.1
Color moments inv	24	50.1 ±5.5	41.0 ±2.4	10.0	44.6
HS	36	53.6 ±5.2	42.9 ±2.9	26.0	44.6
Color names	11	56.8 ±5.8	44.2 ±1.7	25.6	56.0
Discriminative color descriptors	11	55.7 ±5.6	43.9 ±2.1	22.0	50.7
Discriminative color descriptors	25	57.4 ±5.8	46.4 ±2.2	25.6	54.0
Discriminative color descriptors	50	**60.1 ±5.7**	**48.1 ±1.9**	**27.4**	**58.0**

Table 4. Comparison (%) with state-of-the-art approaches on four texture datasets. Our approach provides the best performance on KTH-TIPS-2a, KTH-TIPS-2b and Texture-10 datasets.

Method	KTH-TIPS-2a	KTH-TIPS-2b	FMD	Texture-10
LHS Sharma et al. (2012)	73.0	-	-	-
TFT Timofte and Gool (2012)	-	66.3	55.7	-
PIF Sharan et al. (2013)	-	-	**57.1**	-
CNLBP Khan et al. (2013b)	-	-	-	77.0
PM Khan et al. (2013b)	-	-	-	73.0
WLD Chen et al. (2010)	56.4	-	-	-
MWLD Chen et al. (2010)	64.7	-	-	-
SDIC Sifre and Mallat (2013)	-	-	41.4	-
LQP ul Hussain and Triggs (2012)	64.2	-	-	-
LTP Tan and Triggs (2010)	60.0	-	-	-
CMR Zhang et al. (2013)	69.4	-	-	-
ELBP Liu et al. (2012b)	-	58.1	-	-
SRP Liu et al. (2012a)	-	-	48.2	-
LBP-HF Ahonen et al. (2009)	-	54.6	-	-
VZ-MR8 Varma and Zisserman (2010)	-	46.3	-	-
CMLBP Li and Fritz (2012)	73.1	-	-	-
aLDA Liu et al. (2010)	-	-	44.6	-
ETF Satpathy et al. (2014)	62.6	-	-	-
LBPD Hong et al. (2014)	74.9	-	-	-
LVCBP Lee et al. (2012)	61.7	53.6	38.4	58.7
This paper	**82.7**	**70.6**	54.2	**82.0**

achieves a recognition rate of 61.7%. Our approach, while being compact, outperforms the state-of-the-art methods with a significant gain of 7.8% over the best reported result. On the KTH-TIPS-2b dataset, the extended LBP approach (Liu et al., 2012b) provides an accuracy of 58.1%. A combination of LBP and Fourier features achieves an accuracy of 54.6%. Our approach outperforms existing methods on this dataset by providing a recognition accuracy of 70.6%.

On the FMD dataset, a training-free approach by Timofte and Gool (2012) obtains a recognition accuracy of 55.7%. Our approach, despite its simplicity, achieves an accuracy of 54.2%. The best results on this dataset are obtained using perceptually inspired features (Sharan et al., 2013). It is worthy to mention that our approach neither uses any ground-truth masks nor any perceptually inspired features. Such features are complementary to the approach presented in this paper and can be combined to obtain further boost in performance. Finally, on the Texture-10 dataset, our approach outperforms the color names and LBP fusion methods (Khan et al., 2013b) by achieving a recognition accuracy of 82.0%.

6. Conclusion

In this paper we investigated the problem of texture recognition in images. Firstly, we have shown that fusing different texture representations significantly improves the performance compared to the single best method. To counter the high-dimensionality problem of the image representation, we proposed to use the DITC approach. Additionally, we performed a comprehensive evaluation of pure color descriptors, popular in image classification, for the task of texture recognition.

The results show that our compact texture representation with a dimensionality of only 500 significantly improved the performance over existing texture classification methods. Among the color descriptors, the discriminative color descriptors provide the best results. Finally, we fused the discriminative color descriptors with our compact texture representation and showed that it can achieve state-of-the-art performance.

In this work, we used a simple late fusion technique to combine the color and texture features. Future work includes investigating sophisticated fusion approaches to combine the color and texture descriptions. A further comparison of the DITC approach with other compression approaches (Jiang, 2009; Scholkopf et al., 1998) can provide a further insight on its applicability to other computer vision applications.

Acknowledgements

This work has been supported by SSF through a grant for the project CUAS, by VR through a grant for the project ETT, through the Strategic Area for ICT research ELLIIT, CADICS and The Academy of Finland (Finnish Centre of Excellence in Computational Inference Research COIN, 251170). We also acknowledge the grants 255745 and 251170 of the Academy of Finland, SSP-14183 of EIT ICT Labs, and the D2I SHOK project. The project TIN2013-41751 of Spanish Ministry of Science. The calculations were performed using computer resources within the Aalto University School of Science "Science-IT" project.

References

Ahonen, T., Hadid, A., Pietikainen, M., 2004. Face recognition with local binary patterns, in: ECCV.

Ahonen, T., Matas, J., He, C., Pietikainen, M., 2009. Rotation invariant image description with local binary pattern histogram fourier features, in: SCIA.

Bosch, A., Zisserman, A., Munoz, X., 2006. Scene classification via plsa, in: ECCV.

Caputo, B., Hayman, E., Mallikarjuna, P., 2005. Class-specific material categorisation, in: ICCV.

Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W., 2010. Wld: A robust local image descriptor. PAMI 32, 1705–1720.

Dhillon, I., Mallela, S., Kumar, R., 2003. A divisive information-theoretic feature clustering algorithm for text classification. JMLR 3, 1265–1287.

Elfiky, N., Khan, F.S., van de Weijer, J., Gonzalez, J., 2012. Discriminative compact pyramids for object and scene recognition. PR 45, 1627–1636.

Gevers, T., Smeulders, A.W.M., 1999. Color based object recognition. PR 32, 453–464.

Guo, Y., Zhao, G., Pietikainen, M., 2012. Discriminative features for texture description. PR 45, 3834–3843.

Guo, Z., Zhang, L., Zhang, D., 2010. A completed modeling of local binary pattern operator for texture classification. TIP 19, 1657–1663.

Hong, X., Zhao, G., Pietikainen, M., Chen, X., 2014. Combining lbp difference and feature correlation for texture description. TIP 23, 2557–2568.

ul Hussain, S., Triggs, B., 2012. Visual recognition using local quantized patterns, in: ECCV.

Jiang, X., 2009. Asymmetric principal component and discriminant analyses for pattern classification. PAMI 31, 931–937.

Kannala, J., Rahtu, E., 2012. Bsif: Binarized statistical image features, in: ICPR.

Khan, F.S., Anwer, R.M., van de Weijer, J., Bagdanov, A., Lopez, A., Felsberg, M., 2013a. Coloring action recognition in still images. IJCV 105, 205–221.

Khan, F.S., Anwer, R.M., van de Weijer, J., Bagdanov, A.D., Vanrell, M., Lopez, A.M., 2012a. Color attributes for object detection, in: CVPR.

Khan, F.S., van de Weijer, J., Ali, S., Felsberg, M., 2013b. Evaluating the impact of color on texture recognition, in: CAIP.

Khan, F.S., van de Weijer, J., Vanrell, M., 2012b. Modulating shape features by color attention for object recognition. IJCV 98, 49–64.

Khan, R., van de Weijer, J., Khan, F.S., Muselet, D., Ducottet, C., Barat, C., 2013c. Discriminative color descriptors, in: CVPR.

Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation using local affine regions. PAMI 27, 1265–1278.

Lee, S.H., Choi, J.Y., Ro, Y.M., Plataniotis, K., 2012. Local color vector binary patterns from multichannel face images for face recognition. TIP 21, 2347–2353.

Leung, T., Malik, J., 2001. Representing and recognizing the visual appearance of materials using three-dimensional textons. IJCV 43, 29–44.

Li, M., Bao, S., Dong, W., Wang, Y., Su, Z., 2013. Head-shoulder based gender recognition, in: ICIP.

Li, S.T., Li, Y., Wang, Y.N., 2004. Comparison and fusion of multiresolution features for texture classification, in: ICMLC.

Li, W., Fritz, M., 2012. Recognizing materials from virtual examples, in: ECCV.

Liu, C., Sharan, L., Adelson, E., Rosenholtz, R., 2010. Exploring features in a bayesian framework for material recognition, in: CVPR.

Liu, L., Fieguth, P., Clausi, D., Kuang, G., 2012a. Sorted random projections for robust rotation-invariant texture classification. PR 45, 2405–2418.

Liu, L., Zhao, L., Long, Y., Kuang, G., Fieguth, P., 2012b. Extended local binary patterns for texture classification. IVC 30, 86–99.

Maenpaa, T., Pietikainen, M., 2004. Classification with color and texture: jointly or separately? PR 37, 1629–1640.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. PAMI 24, 971–987.

Rahtu, E., Heikkila, J., Ojansivu, V., Ahonen, T., 2012. Local phase quantization for blur-insensitive image analysis. IVC 30, 501–512.

van de Sande, K.E.A., Gevers, T., Snoek, C.G.M., 2010. Evaluating color descriptors for object and scene recognition. PAMI 32, 1582–1596.

Satpathy, A., Jiang, X., Eng, H.L., 2014. Lbp-based edge-texture features for object recognition. TIP 23, 1953–1964.

Scholkopf, B., Smola, A., Muller, K.R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10, 1299–1319.

Sharan, L., Liu, C., Rosenholtz, R., Adelson, E., 2013. Recognizing materials using perceptually inspired features. IJCV 103, 348–371.

Sharma, G., ul Hussain, S., Jurie, F., 2012. Local higher-order statistics (lhs) for texture categorization and facial analysis, in: ECCV.

Sifre, L., Mallat, S., 2013. Rotation, scaling and deformation invariant scattering for texture discrimination, in: CVPR.

Tan, X., Triggs, B., 2007. Fusing gabor and lbp feature sets for kernel-based face recognition, in: AMFG.

Tan, X., Triggs, B., 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. TIP 19, 1635–1650.

Timofte, R., Gool, L.V., 2012. A training-free classification framework for textures, writers, and materials, in: BMVC.

Varma, M., Zisserman, A., 2010. A statistical approach to texture classification from single images. IJCV 32, 1705–1720.

Wang, X., Han, T., Yan, S., 2009. An hog-lbp human detector with partial occlusion handling, in: ICCV.

van de Weijer, J., Schmid, C., 2006. Coloring local feature extraction, in: ECCV.

van de Weijer, J., Schmid, C., Verbeek, J.J., Larlus, D., 2009. Learning color names for real-world applications. TIP 18, 1512–1524.

Ylioinas, J., Hadid, A., Guo, Y., Pietikainen, M., 2012. Efficient image appearance description using dense sampling based local binary patterns, in: ACCV.

Ylioinas, J., Hadid, A., Pietikainen, M., 2011. Combining contrast information and local binary patterns for gender classification, in: SCIA.

Ylioinas, J., Hong, X., Pietikainen, M., 2013. Constructing local binary pattern statistics by soft voting, in: SCIA.

Zhang, J., Huang, K., Yu, Y., Tan, T., 2011. Boosted local structured hog-lbp for object localization, in: CVPR.

Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C., 2007. Local features and kernels for classification of texture and object catergories: A comprehensive study. IJCV 73, 213–218.

Zhang, J., Zhao, H., Liang, J., 2013. Continuous rotation invariant local descriptors for texton dictionary-based texture classification. CVIU 117, 56–75.

Zhang, L., Zhou, Z., Li, H., 2012. Binary gabor pattern: An efficient and robust descriptor for texture classification, in: ICIP.

Zhao, G., Ahonen, T., Matas, J., Pietikainen, M., 2012. Rotation-invariant image and video description with local binary pattern features. TIP 21, 1465–1477.