Supervised Texture Classification Using a Novel Compression-Based Similarity Measure

Mehrdad J. Gangeh[1], Ali Ghodsi[2], and Mohamed S. Kamel[1]

[1] Center for Pattern Analysis and Machine Intelligence,
Department of Electrical and Computer Engineering, University of Waterloo,
Ontario N2L 3G1, Canada,
{mgangeh,mkamel}@pami.uwaterloo.ca
[2] Department of Statistics and Actuarial Science,
University of Waterloo, Ontario N2L 3G1, Canada,
aghodsib@uwaterloo.ca

Abstract. Supervised pixel-based texture classification is usually performed in the feature space. We propose to perform this task in (dis)similarity space by introducing a new compression-based (dis)similarity measure. The proposed measure utilizes two dimensional MPEG-1 encoder, which takes into consideration the spatial locality and connectivity of pixels in the images. The proposed formulation has been carefully designed based on MPEG encoder functionality. To this end, by design, it solely uses P-frame coding to find the (dis)similarity among patches/images. We show that the proposed measure works properly on both small and large patch sizes. Experimental results show that the proposed approach significantly improves the performance of supervised pixel-based texture classification on Brodatz and outdoor images compared to other compression-based dissimilarity measures as well as approaches performed in feature space. It also improves the computation speed by about 40% compared to its rivals.

1 Introduction

Texture images can be divided to two broad types: stationary that contains only one texture type per image and nonstationary that consists of more than one texture type per image [1]. The main application domain on stationary texture images is supervised classification of each texture image into one class; whereas on nonstationary texture images, there are two main application domains [1, 2]. First, unsupervised texture segmentation that partitions the texture image into disjoint regions of uniform texture. Second, pixel-based texture classification, which is similar to texture segmentation in the sense that the given texture image is segmented to uniform texture regions. The difference, however, is that in pixel classification, the segmentation is performed using supervised techniques [2]. In this paper, our focus is on supervised pixel classification and hence, we deal with nonstationary texture types.

Common trend in literature on pixel-based texture classification is the computation of some texture features for every pixel using its neighboring pixels and a particular texture method [2–4]. However, as texture is a complicated phenomenon, there is no definition that is agreed upon by the researchers in the field [5, 6]. This is one of the reasons that there are various feature-based techniques in the literature, each of which tries to model one or several properties of texture depending on the application in hand. The performance of each of these features depends on the texture type and there is no single feature method that performs well on all different texture types [2, 3]. To avoid this problem, textures can be represented in (dis)similarity space. In this approach, pairs of texture patches are compared by a (dis)similarity measure reflecting their mutual resemblance.

Among similarity measures in the literature, the metric based on the notion of Kolmogorov complexity, i.e., so called normalized information distance (NID) [7] has attracted the attention of many researchers. However, due to non-computability of Kolmogorov complexity, it has been mainly approximated using real-world compressors [8] introducing normalized compression distance (NCD). NCD has attractive characteristics, e.g., it is parameter-free, i.e., does not use any feature or background knowledge about the data; and it is quasi-universal, (NID is universal, i.e., it minorizes all other distances, but NCD inherits this from NID to some extent [8]).

NCD was originally defined on binary strings with the explanation that all data types can be converted to binary strings. Many initial applications on which NCD was applied successfully were based on 1D data such as in bioinformatics or plagiarism detection. The extension of NCD application to 2D data such as images, however, does not seem to be straightforward. While some researchers linearize 2D data to represent them using 1D strings [9, 10], this causes the loss of the spatial locality and connectivity of neighboring pixels. The effect of linearization on the overall performance of NCD-based system has been empirically investigated in [9] with this important conclusion: "images may not be fully expressible as a string, at least using current compression algorithms". Using 2D compressors such as JPEG and JPEG2000 for NCD on images led to contradictory results in the literature: while [11] shows that using JPEG2000 on 2D satellite images yields better results than converting images to 1D and using string compressors, [9] and [12] show that JPEG and JPEG2000 does not work well as compressors for computing NCD-based similarity measure on images.

An alternative approach is using MPEG encoders as 2D compressors in NCD. The main advantage of MPEG compared to JPEG encoder is that while JPEG is designed for compressing one image, MPEG encodes frames of images and hence, by concatenating two images as two frames, they can be compressed in reference to each other which is desired in NCD. In this paper, we propose a novel formulation based on MPEG encoder for measuring (dis)similarity between images/patches. We will show that this new measure works well on both small and large patch sizes. Introducing this new measure in this paper, we will also show that the results of pixel-based texture classification can be significantly improved compared to other NCD-based approaches in the literature.

2 Compression-Based Dissimilarity Measure

In this section, we first briefly review the concept of NID and NCD and then provide the formulation for our proposed approach. Some illustrative results are then presented to show the effectiveness of the proposed approach on both small and large patch sizes.

2.1 Normalized Compression Distance

The normalized compression distance (NCD) [8] is an approximation for normalized information distance (NID) [7], a universal parameter-free similarity measure based on Kolmogorov complexity that minorizes all other distance measures [7].

To understand the definition of the NID, we need to define two notations: $K(x)$ and $K(x|y)$. The former is the Kolmogorov complexity of string x, which is defined as the length of the shortest binary program to compute x on a universal computer such as universal Turing machine, whereas the latter is the conditional Kolmogorov complexity, which is defined as the length of a shortest program to compute x if y is provided as an auxiliary input for the reference [7]. The NID is defined as

$$NID(x,y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}}. \tag{1}$$

Since Kolmogorov complexity is a noncomputable measure, the NID defined in (1) is computed by approximating Kolmogorov complexity using a compressor denoted by C as follows [8]

$$NCD(x,y) = \frac{\min\{C(xy), C(yx)\} - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}, \tag{2}$$

where xy means that the strings x and y are concatenated. To have more insight into (2), we consider the case that $C(y) \geq C(x)$[1] and the compressor is symmetric such that $C(xy) = C(yx)$. In this case, we can rewrite (2) as $NCD(x,y) = \frac{C(xy)-C(x)}{C(y)}$, which means that the NCD distance between x and y is improvement on compressing y using x (the numerator, which is also denoted as $C(y|x)$) over compressing y by its own (the denominator) [8]. This interpretation will help to explain our proposed measure later in next subsection.

2.2 Proposed Distance Measure

Since we are using MPEG-1 encoder in our proposed (dis)similarity measure, we first provide some description on how this encoder works. MPEG-1 is a 2D encoder and thus, it takes into account the spatial locality and connectivity of the neighboring pixels in images for compression. MPEG-1 was originally designed

[1] The opposite condition can be interpreted similarly as NCD distance defined in (2) is symmetric.

for compressing movies based on three different coding schemes, i.e., intra-frame (I-frame) coding, predictive frame (P-frame) coding (also called inter-frame coding), and bidirectional frame (B-frame) coding [13]. I-frame coding is performed on individual frames without reference to other frames using discrete cosine transform (DCT). P-frame codes a frame in reference to the previous one by using a block matching algorithm for motion estimation and using DCT on the residual. Finally, B-frame coding compresses a frame with reference to its next and previous frames. To utilize MPEG-1 as compressor in compression-based similarity measures, patches/images are considered as two successive frames and compressed using MPEG-1 encoder. This avoids the need to linearize the images that causes the loss of spatial locality as explained in Section 1. Since there are only two frames (two images whose similarity are to be computed), B-frame coding is not utilized.

Now, if we want to use MPEG-1 as compressor for (dis)similarity measure, we need to use proper formulation based on how MPEG-1 works. To this end, based on the description provided above on MPEG-1 encoder and also the explanation provided on (2) at the end of Subsection 2.1, we would like to propose our new dissimilarity measure considering these two points: First, we utilize MPEG-1 for the computation of $C(x|y)$ (the conditional compression of x given y) using only P-frame coding and bypass I-frame coding as it does not provide any information on the similarity of x and y and we denote it using $C_p(x|y)$. Since the P-frame coding indicates the differences between two frames, which is essential in finding the (dis)similarity between them, we encode it with maximum resolution, i.e., minimum quantization scale, which is one in MPEG-1 (quantization scale for I-frame does not have any effect as I-frame coding is bypassed). Second, we notice that because the second image/patch is compressed in reference to the first one, $C_p(x|y)$ (also $C(x|y)$) is not symmetric. However, if both x and y are from the same distribution (class), we expect $C_p(x|y)$ to be close to $C_p(y|x)$ (because x and y are from the same class and it does not make very much difference whether we compress x in respect to y or y in respect to x), while if x and y are from different distributions (classes), $C_p(x|y)$ and $C_p(y|x)$ should be largely different. Hence, we propose our new measure as follows

$$d_N(x,y) = \frac{|C_p(x|y) - C_p(y|x)|}{C(x|x) + C(y|y)}, \tag{3}$$

where the absolute of the difference is taken in the numerator to ensure positive distances. $C(x|x) + C(y|y)$ is used as the normalizing factor. In $C(x|x)$ and $C(y|y)$, since both frames are the same, P-frame coding generates zero (the difference between two frames is zero). Thus, $C(x|x)$ is equivalent to $C(x)$ in (3). However, since in MPEG-1 encoder, there are at least two frames, we use $C(x|x)$ notation instead of $C(x)$. I-frame quantization scale can be maximized in this case. The proposed distance is symmetric and nonnegative.

Although MPEG-1 has been also used in [14] for dissimilarity measure, our proposed measure is different in following aspects. Firstly, Our proposed formulation is different from what they have proposed. Their distance measure is defined as follows

$$d_{CK}(x, y) = \frac{C(x|y) + C(y|x)}{C(x|x) + C(y|y)} - 1, \tag{4}$$

where $C(x|y)$ is computed based on both I- and P-frames coding, while in our approach, it is computed solely based on P-frame coding (denoted by $C_p(.|.)$). Secondly, in (4), the compression is maximized by using large quantization scales for both I- and P-frames coding through MPEG-1 external parameters to prefer compressibility over image quality [14]. In our approach, since P-frame is essential in finding the (dis)similarity between two frames, we encode it with maximum resolution. Thirdly, our proposed measure performs properly on both small and large patches while $d_{CK}(x, y)$ cannot represent dissimilarity between small patches properly. This is explained more in next subsection.

2.3 Illustrative Results

To better realize how $d_{CK}(x, y)$ works, we have computed the distances among patches of 17×17, 33×33, 65×65, and 129×129 extracted from two texture images of Brodatz, i.e., D4 (Fig. 1a) and D5 (Fig. 1f) as shown in Fig. 1b-1e. As can be seen, the distances computed (300 patches per class) among patches are normalized to the interval of $[0, 1]$ to ease the comparison and displayed using color code. We expect to see smaller distances among patches extracted from the same class, i.e., in $c_i - c_i, i = 1, 2$ areas and larger distances among the patches extracted from two different classes, i.e., in $c_i - c_j, i, j = 1, 2$ & $i \neq j$ areas (see Fig. 1b as reference). However, except for large patch size of 129×129, this behavior cannot be observed in Fig. 1b-1e. This problem can be also seen for any other texture pair and the main reason is explained next.

The major problem with $d_{CK}(x, y)$ defined in (4) is that it compresses the concatenated patches based on both I- and P-frames. This is while only P-frame coding is based on the (dis)similarity of patches and I-frame coding is performed using DCT solely based on the frequency contents of a patch/image. This causes that for small patch sizes, where the compression based on P-frame is still limited (due to small search region) comparing to I-frame coding, the distances mainly be dominated by I-frame coding, i.e., frequency contents and distributions of the first frame. Hence, the patches from the texture class that have low frequency contents show lower distances (in this case D5; one can investigate this by taking the Fourier transform of both textures and looking at their spectrum). This is while in this example, due to more homogeneity of D4, we expect lower distances among the patches extracted from D4, i.e., in region $c_1 - c_1$.

Fig. 1g-1j shows the distances among the same patches used for d_{CK} to illustrate the effectiveness of the proposed distance on finding the (dis)similarities among texture pairs. It can be seen that the distances are consistently small among the patches of the same class for all patch sizes and also the distances among the patches extracted from D4, which is a more homogeneous texture than D5, are smaller. This behavior is consistent on other texture pairs as our experiments indicate (not shown here due to space limit).

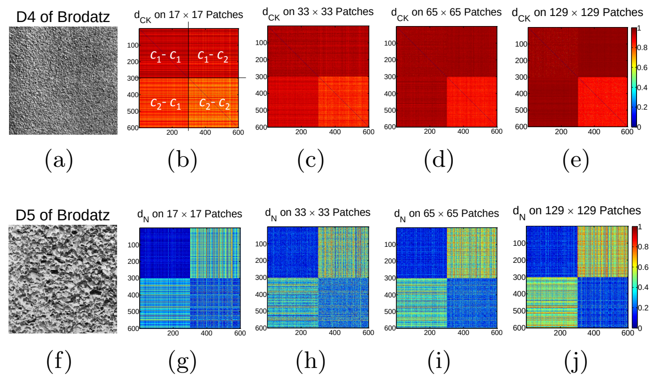

Fig. 1: The distances computed on patches extracted from (a) D4 and (f) D5 of Brodatz album. (b) to (e) distances computed on various patch sizes as indicated in the figures using d_{CK} and (g) to (j) using proposed measure (d_N).

3 Experimental Setup and Results

The effectiveness of the proposed similarity measure is shown in the application of supervised pixel-based classification on nonstationary texture images, i.e., texture images consisting of several different texture areas. In this application, there is a trade-off between the patch sizes at smooth areas and on the borders. While large patch size at the uniform texture areas improves the performance of classification (as more information is included to identify the textures correctly), small patch sizes are more desired on the borders to prevent mixing textures from two different classes.

Here, the distances are first computed on 200 patches per class with the size of 17×17 extracted from the training images. These are used to train a support vector machine (SVM) with linear kernel $k_{\text{tr}} = d_{\text{tr}}.d'_{\text{tr}}$ (d_{tr} is the distance matrix computed on the patches extracted from the training set). This kernel is p.s.d. as it is obtained using an inner product. The optimal cost function (C^*) of the SVM is tuned in a 5-fold cross-validation on the training set. Then the patches of the same size are extracted from the test image and the distances among these patches and the training patches are computed using the proposed approach. A linear kernel is computed subsequently using $k_{\text{ts}} = d_{\text{ts}}.d'_{\text{tr}}$ (d_{ts} is the computed distances from the test to training patches), which is used in the trained SVM.

Data used is the same as what is used in [2]. It is consisting of some texture composites from Brodatz and some outdoor images. The test images are shown on the first column of Fig. 2. The results are compared to two other distance measures using d_{CK} and NCD approach and also to two feature-based approaches published in [2] that yield the best results on these texture images, i.e., local binary pattern ($LBP^{riu2}_{8,1}$) and MeasTex (Gabor, 5NN) (refer to Table 3 of [2]). To get rid of the speckle-noise type in final classification, the same as in [2], a median filter with the same size as the patch sizes (17×17 in this case) is applied

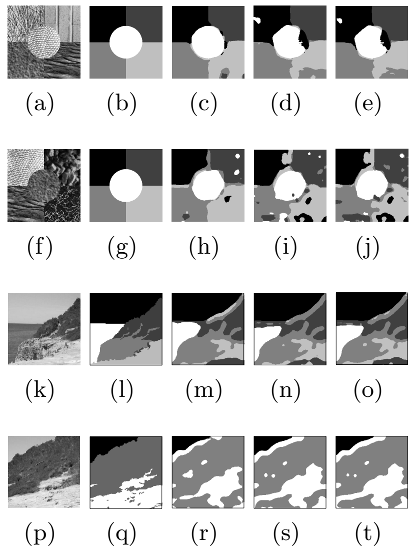

Fig. 2: The results of supervised pixel-based texture classification on Brodatz and outdoor images. (a, f, k, and p) test images; (b, g, l, and q) ground truth; (c, h, m, and r) proposed method; (d, i, n, and s) d_{CK}; (e, j, o, and t) NCD.

Table 1: The classification rate (%) compared among the proposed method and other distance- or feature-based approaches. The results on LBP (local binary pattern) and MeasTex (Gabor, 5NN) methods are based on what is reported in [2] for the same images.

Approach	Test Images			
	Fig. 2a	Fig. 2f	Fig. 2k	Fig. 2p
Proposed	**89.5**	**83.2**	**75.8**	**72.3**
d_{CK}	82.1	74.0	75.0	71.6
d_{NCD}	83.3	73.3	75.3	71.4
$LBP_{8,1}^{riu2}$ [2]	85.4	77.5	69.4	37.9
MeasTex (Gabor, 5NN) [2]	83.7	70.5	68.5	55.1

to the final classified pixels. The results are shown quantitatively in Table 1 and qualitatively in Fig. 2. As can be seen, our results are significantly better than other distance-based approaches and also from what is reported in [2].

4 Discussion and Conclusion

In this paper, we have proposed a new compression-based distance measure using MPEG-1 encoder that takes into account the spatial locality and connectivity of pixels in images. The proposed measure computes distances based on P-frame coding and can properly find the distances on both small and large patch sizes, unlike d_{CK} which works only on large patches. By bypassing the I-frame coding, which is not necessary in the computation of distances anymore (except for the case that the patches are the same), our method improves the performance in terms of speed by 40% compared to the d_{CK}. The effectiveness of the proposed measure was shown on supervised pixel-based texture classification for both texture images from Brodatz and outdoor images yielding a significantly improved performance.

References

1. Petrou, M., Sevilla, P.G.: Image Processing Dealing with Texture. John Wiley & Sons, West Sussex (2006)
2. Garcia, M., Puig, D.: Supervised texture classification by integration of multiple texture methods and evaluation windows. Image and Vision Computing **25**(7) (July 2007) 1091–1106
3. Randen, T., Husøy, J.: Filtering for texture classification: A comparative study. IEEE Trans. Pattern Analysis and Machine Intelligence **21**(4) (April 1999) 291–310
4. Melendez, J., Puig, D., Garcia, M.: Multi-level pixel-based texture classification through efficient prototype selection via normalized cut. Pattern Recognition **43**(12) (2010) 4113 – 4123
5. Mirmehdi, M., Xie, X., J. Suri, E.: Handbook of Texture Analysis. Imperial Collage Press, London (2008)
6. Ahonen, T., Pietikainen, M.: Image description using joint distribution of filter bank responses. Pattern Recognition Letters **30**(4) (March 2009) 368–376
7. Li, M., Chen, X., Li, X., Ma, B., Vitányi, P.: The similarity metric. IEEE Trans. Information Theory **50**(12) (2004) 3250 – 3264
8. Cilibrasi, R., Vitányi, P.: Clustering by compression. IEEE Trans. Information Theory **51**(4) (2005) 1523 – 1545
9. Mortensen, J., Wu, J., Furst, J., Rogers, J., Raicu, D.: Effect of image linearization on normalized compression distance. In Slezak, D., Pal, S.K., Kang, B.H., Gu, J., Kuroda, H., Kim, T.H., eds.: Signal Processing, Image Processing and Pattern Recognition. Volume 61 of Communications in Computer and Information Science. (2009) 106–116
10. Macedonas, A., Besiris, D., Economou, G., Fotopoulos, S.: Dictionary based color image retrieval. Journal of Visual Communication and Image Representation **19**(7) (2008) 464 – 470
11. Cerra, D., Mallet, A., Gueguen, L., Datcu, M.: Algorithmic information theory-based analysis of earth observation images: An assessment. IEEE Geoscience and Remote Sensing Letters **7**(1) (2010) 8 – 12
12. Vázquez, P., Marco, J.: Using normalized compression distance for image similarity measurement: an experimental study. The Visual Computer (2011)
13. Ghanbari, M.: Standard Codecs: Image Compression to Advanced Video Coding. The Institution of Electrical Engineers, London, UK (2003)
14. Campana, B., Keogh, E.: A compression-based distance measure for texture. Statistical Analysis and Data Mining **3**(6) (2010) 381–398