Computer Communication & Collaboration (2013)
Submitted on 23/April/2013
DOIC: 2292-1036-2013-01-001-18

A Comparison of Multi-scale Approaches for Extracting Image Descriptors from the Co-occurrence Matrix

Loris Nanni, Stefano Ghidoni, Emanuele Menegatti
Department of Information Engineering, University of Padova, Italy

E-mail: nanni@dei.unipd.it

Abstract: One of the first methods for analyzing the texture of an image was proposed in 1979 by Haralick, who introduced the co-occurrence matrix for calculating a set of image statistics. In this paper we focus on novel texture descriptors extracted from the co-occurrence matrix. It is well known that scale is important information in texture analysis, since the same texture can be perceived as different patterns at distinct scales. In this work we present, compare and combine different strategies for extending the texture descriptors extracted from the co-occurrence matrix at multiple scales. The texture descriptors are used to train a support vector machine and some different fusion techniques are compared. Our results are validated using seven image classification problems (mainly medical image classification problems). Our results shown that we improve the performance of the standard approaches. The code for the approaches tested in this paper is available at: http://www.dei.unipd.it/wdyn/?IDsezione=3314&IDgruppo_pass=124&preview=.

Keywords: Texture descriptors, Co-occurrence matrix, Multiple scales strategy, Support vector machine

1 Introduction

The increase in computational power of commonly available computers enabled research centers and companies to face the world of computer vision at the beginning of this century. Image acquisition, storage and fast processing was possible without the need for dedicated hardware and expensive equipment. This is at the origin of the big boost that computer vision had on a high number of applications [1-3], ranging from video surveillance for indoor and outdoor environments to visual inspection of industrial products. Medical imaging is another field that is getting strong interest by computer vision researchers: over the years, medical search engines and specialized databases, such as HUGO, Rfam, and Cancer Cell Map were created. Machine vision technology applied to these databases of images has the potential of rapidly accelerating scientific knowledge. The development of novel and improved tools for automatic analysis and classification has already shown its positive effects in clinical practice and in medical and biological research. In [3], for example, the linear discriminant analysis of wavelet features is demonstrated to be effectively for the detection of tumors in endoscopic images the linear discriminant analysis of wavelet features is demonstrated to be effectively for the detection of tumors in endoscopic images. This can be used in applications that range from traffic accident detection [4] to face identification and verification [5]. In [6], image texture information is used to successfully discriminate polyps in colonoscopy images.

The Local Binary Pattern (LBP) operator [7] is a tool that offers great performance thanks to its simplicity, effectiveness, and robustness, and was successfully used for detecting a variety of tumors and masses. In [8], for example, LBP was used to assign a Marsh-like score to endoscopical

images of pediatric celiac diseases, thus providing concrete help for pathologists. In [9] a Support Vector Machine (SVM) was coupled with the LBP operator to distinguish real masses from normal parenchyma in mammographic images, thus reducing the incidence of false positive samples. LBP can also be combined with other features and descriptors that are useful for data mining purposes. In [10], for instance, LBP was used to explore brain magnetic resonance data, and in [11] the authors demonstrated how a combination of LBP with other texture descriptors is effective in classifying different cell phenotypes using SVM.

Image analysis and classification is often based on the analysis of texture. Texture is a concept which is difficult to define, measure and compare, but provides strong information content: a number of methods for managing texture have therefore been developed, based on very different approaches. The most performing methods presented in the literature include the scale-invariant feature transform (SIFT) [28], speeded up robust feature (SURF) [29], histogram of oriented gradients (HOG) [30], gradient location and orientation histogram (GLOH) [31], region covariance matrix (RCM) [32], edgelet [33], gray level co-occurrence matrix (GLCM) [34], local binary patterns (LBP) [35], non-binary codings [19], color correlogram (CCG) [36], color coherence vectors (CCV) [37], color indexing [38], steerable filters [39] and Gabor filters [40].

In this work we focus on GLCM, which is one of the most general approaches to texture analysis, and was originally developed in 1979 for analyzing satellite images. This approach is based on a set of features (descriptors) that are evaluated starting from a histogram, and has been exploited by several research groups: to increase the discriminability of the descriptors, Gelzinis et al. [41] consider simultaneously different values of the distance parameter that influences the GLCM. Walker et al. [42] have proposed co-occurrence matrix-based features by weighted summation of GLCM elements from areas presenting high discrimination. Furthermore, addition of color information has also been considered for co-occurrence matrices [43]. Multi-scale analysis has also been performed using the GLCM. Hu [44] and Pacifici et al. [45] consider multiple scales by changing the window size from which the GLCM descriptors are extracted. Rakwatin et al. [46] propose to rescale the image to different sizes, extracting co-occurrence descriptors from each size. Nguyen-Duc et al. [47] have obtained improved results on content-based image retrieval employing a combination of contourlet transform and GLCM. First, the contourlet transform is performed for four subbands of the image, then the GLCM features are extracted from each one.

One of the main difficulties when analyzing texture is that results have a strong dependence on resolution and scale, an effect that is much stronger with respect to e.g. edge-based approaches. The aim of this work is to assess the performance improvement obtained using a multi-scale approach. In the literature several multi-scales approaches are coupled with local binary pattern descriptors [49,50]. Here our aim is to show that it is possible to improve the performance of the descriptors extracted from the co-occurrence matrix coupling them with a multi-scale approach.

We extend [17] using very different image classification problems, and combine two different ways for extracting features from the co-occurrence matrix (CM). Moreover we use a better performing classifier: the support vector machine (SVM), instead of the weak nearest neighbor used in [17].

We couple the multi-scale approach both with the standard method based on Haralick's features (HAR) for describing the CM and with a very recent method called SHAPE [16,18,27] where features are extracted considering the co-occurrence matrix as a 3D shape. Two multi-scale approaches are applied: Gaussian scale-space representation and an image pyramid.

Moreover, we report that the fusion between SHAPE-based features and HAR further boosts the performance. Our experiments were validated across seven databases of very different

classification problems (see section 3), our results are also compared with some state-of-the-art descriptors.

The paper is organized as follows. In section 2 we describe our proposed approaches. In section 3 we describe the benchmark databases used to validate our approach. In section 4, we present our experimental results, and in section 5 we conclude our paper with a few suggestions for further research.

2 Proposed Approaches

The focus of the paper is on the descriptors based on the co-occurrence matrix (CM). Our goal is to enhance the performance of the Haralick descriptor and of a set of recently proposed features based on the analysis of the 3D shape of the co-occurrence matrix. Improvements are achieved by means of a multi-scale approach that overcomes the main weakness of texture-based features, which is the dependency on the scale, as previously discussed. Moreover different fusion approaches are compared for improving the performance.

In the following sub-section 2.1 the texture descriptors are explained, in sub-section 2.2 the different multi-scales approaches are detailed and in section 2.3 the support vector machine is briefly explained.

2.1 Descriptors extracted from the co-occurrence matrix

The Gray Level Dependency Matrix (GLDM) [12,34] is a particular type of co-occurrence matrix (CM) obtained as the histogram on a two-dimensional domain of dimension $N_{GL} \times N_{GL}$, where N_{GL} is the number of gray levels in the image, usually 256. Such CM records the gray level transitions between pixels. Each location on the histogram identifies a specific transition between two values. The algorithm to evaluate the GLDM considers pixel couples in the source image: for each one, the bin of the histogram whose coordinates are equal to the values of gray levels of the two pixels is incremented. The way couples are determined depends on two parameters, d and θ, that represent the distance of the two pixels and the direction on which they are aligned, respectively.

In this work, four directions are considered: the horizontal (H), the vertical (V), the diagonal top left-bottom right, or right-down (RD), and the top right-bottom left, or left-down (LD). For example, values of d=1 and θ=0 cause pixels to be grouped in couples that are adjacent to each other on the same row. The original image appears on the left-hand side, with the corresponding GLDM on the right-hand side. It should be noted that the locations placed along the principal diagonal of the GLMD represent transitions between two pixels of the same color because they have coordinates with the same value for x and y. Moreover, pixels in the neighborhood of the diagonal represent transitions between similar gray levels. Since pixel couples with similar color are much more frequent than strong transitions, the GLDM has a large component on the diagonal and in the neighboring region.

Since a GLDM records pixel transitions, it contains information about image texture. This information is however in a raw form, and it is very complex to handle. As a result, the GLDM is never used directly, but rather, a set of features (or indicators) are exploited to process the content of the GLDM. The use of indicators was originally introduced together with the GLDM itself in [43]. The selection of a good set of features that provides optimal information from the matrix for a particular problem is crucial, because a wrong choice at this stage could lead to information loss or to redundancy. It is then worth to investigate about novel indicators to describe GLDMs as they can improve the performance of texture analysis.

The first approach proposed in this paper, for measuring image texture, exploits some indicators that were introduced in the original paper by Haralick [12]. This approach is referred to as HAR. The following 13 features introduced by Haralick are exploited:

1. Energy
2. Correlation
3. Inertia
4. Entropy
5. Inverse difference moment
6. Sum average
7. Sum variance
8. Sum entropy
9. Difference average
10. Difference variance
11. Difference entropy
12. Information measure of correlation 1
13. Information measure of correlation 2

The second approach for measuring image texture is called SHAPE because it explores the shape of the CM that is considered as a 3D function as illustrated in Figure 1. This approach was originally proposed for crowd detection in [16]. The main idea is to intersect the GLDM with a set of horizontal planes at given heights (see Figure 1-a), and develop a set of features based on the contours obtained in this way. The intersection (see Figure 1-b) defines a complex shape that can consist of one or more extracted blobs. The main blob, that is, the one with the largest area, is then selected for extracting features. The main blob is fitted to an ellipse (drawn in red in Figure 1-c) in order to simplify the analysis. This approximation of the main blob shape using an ellipse results in some information loss, but it offers the advantage of making the comparison among curves much easier.

Level curves are considered towards the base of the GLDM, starting at height 1 and going until height 19, with a distance of 2 between two consecutive planes. They are all at a rather low height because that region is very stable: it does not change significantly between two similar frames, while the peaks of the GLDM are much more unstable and can fluctuate because of the image noise. For this reason the GLDM is not even normalized, since a normalization to the highest bin would introduce instabilities. Another possible normalization could be performed with respect to the total volume of the GLDM, which depends mainly on the size of the original image.

For each level, a set of descriptors regarding the ellipses derived from the GLDM are evaluated. The features describing all levels are then jointly analyzed for deriving another set of features that describe the evolution of the level curves.

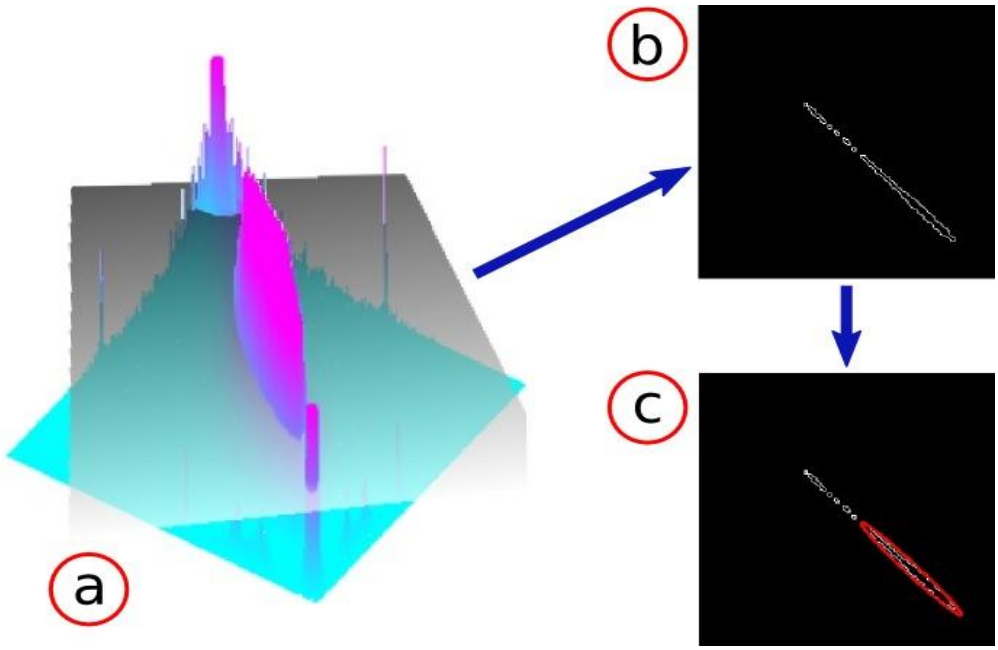

Figure 1: GLDM processing: the matrix is intersected with a set of planes (a). The intersection (b) provides multiple contours: the one with the largest area is then selected and described by an ellipse (c).

A number of features can be obtained from a joint analysis of the ellipses. First, features can describe how evenly (EV) the minor axis of the ellipses decreases as the height of the level curve is increased:

$$EV = \sqrt{\frac{1}{N}\frac{\sum_{i=1}^{N-1}[(w_i - w_{i+1}) - \overline{\Delta w}]^2}{i}}$$

where w_i represents the minor axis (width) of the i-th layer, N the total number of layers, and $\overline{\Delta w}$ is the average difference between ellipses at two consecutive layers. The sum is divided by *i* since in the general case the distance between two consecutive layers cannot be assumed to be constant.

A second descriptor evaluates the Minor Axis Spread (MAS), which provides information about the GLDM steepness. MAS is evaluated as the difference between the maximum and minimum observed for the length of the ellipse minor axis:

$$MAS = w_{max} - w_{min}$$

A third descriptor is the minimum value (MV) for the minor axis. This provides information about the upper part of the GLDM: the larger MV, the smoother the upper part of the function:

$$MV = \min_i w_i$$

A fourth descriptor is the mean of the height/width ratio (WHR) of all ellipses. This measures the relationship between the contrast of the image (the higher the contrast, the longer the major axis) and the presence of strong variations between pixels (which causes a long minor axis). This mean is obtained as follows:

$$WHR = \frac{1}{N}\sum_{i=1}^{N}\frac{w_i}{h_i}$$

A fifth descriptor is the total volume (TV) under the analyzed level curves, which can be calculated as follows:

$$TV = w_1 h_1 l_1 + \sum_{i=2}^{N} w_i h_i (l_i - l_{i-1})$$

where w_i and h_i are the minor and major axes, respectively, of the ellipse found on the i-th layer at height l_i.

Finally, descriptors can be based on the area of the smallest ellipse, or minimum area (MA), as well as on the ratio between the smallest and the largest ellipses, or the maximum-minimum area Ratio (MMAR):

$$MA = \min_i w_i h_i = w_N h_N$$

$$MMAR = \frac{w_1 h_1}{w_N h_N}$$

It should be noted that, by construction, the N-th layer is the one with the smallest ellipse, while the largest one is placed on the first layer considered.

All the descriptors mentioned above measure the characteristics of the ellipses approximating the main components, or blobs, of the level curves. Two other features are exploited to measure other characteristics of the co-occurrence matrix: the first is based on the volume of the peak (PV), or the portion of the GLDM that is above the highest level curve considered:

$$PV = \sum_{i=1}^{M}(H_i - l_N)$$

where l_N is the height of the highest layer, and M is the number of bins with height $H_i > l_N$.

The second feature that does not rely on the elliptic approximation is the number of blank locations (BL), i.e., the number of bins with zero occurrences:

$$BL = \sum_{i=1}^{N_{GL}}\sum_{j=1}^{N_{GL}} \begin{cases} 1 \; iff \; GLDM(i,j) = 0 \\ 0 \; else \end{cases}$$

where GLDM(i,j) is the value of the GLDM at location (i,j).

It should be noted that all the above features indirectly depend on the parameters used to evaluate the GLDM. In our case, to achieve a complete characterization of the input image, features are evaluated at all orientations, i.e. θ={0 °, 45 °, 90 °, 135 °}.

The features described so far are used to provide a characterization of the input image, and can be directly used as input for a classifier: this is the principle exploited in the HAR approach. In the case of the SHAPE approach, features are evaluated not only on the whole co-occurrence matrix (as in [16]), but on the following 13 windows[1] (recall that the GLDM has a fixed dimension of

[1] We have tested also smaller sub-windows without obtaining any performance improvement.

256×256 that depends on the number of gray levels of the original image rather than on its resolution):

1. whole co-occurrence matrix;
2. sub-window of the GLDM, from the coordinates (0, 0) to (127, 127);
3. sub-window of the GLDM, from (128, 128) to (255, 255);
4. sub-window of the GLDM, from (0, 0) to (191, 191);
5. sub-window of the GLDM, from (64, 64) to (255, 255);
6. sub-window of the GLDM, from (0, 0) to (95, 95);
7. sub-window of the GLDM, from (31, 31) to (95, 95);
8. sub-window of the GLDM, from (63, 63) to (127, 127);
9. sub-window of the GLDM, from (95, 95) to (159, 159);
10. sub-window of the GLDM, from (127, 127) to (191, 191);
11. sub-window of the GLDM, from (159, 159) to (223, 223);
12. sub-window of the GLDM, from (191, 191) to (255, 255);
13. sub-window of the GLDM, from (63, 63) to (191, 191).

All windows are placed along the principal diagonal of the GLDM because it is the region where the main component of the histogram is located; also, windows are different in size to be able to get a detailed description of given portion of the GLDM. By analyzing small regions of the GLDM it is possible to better analyze its symmetry.

For each window a different feature vector (called a descriptor) is extracted; the 13 descriptors are then used to train 13 different SVMs, which are combined by weighted sum rule: a weight of 1 is assigned to the first 5 descriptors; a weight of 0.5 is assigned to the others. Each set of 13 descriptors comes from co-occurrence matrices evaluated at θ={0°, 45°, 90°, 135°}, and at a given value of d. When multiple values are used for the distance (e.g. d=1 and d=3), the descriptor is obtained concatenating the features extracted for each value of the distance.

2.2 Multi-scale approaches

We test two multi-scale approaches: image pyramid and Gaussian scale-space representation. The pyramid multiscale[2] representation of an image is a hierarchy of $l=(0,...,L)$ different versions of the same image, obtained as follows. Let us define the original image as $G_0(x,y)$. Each element of the hierarchy at the *l-th* level is obtained using a low-pass-filter and by resizing the image of the (*l-1)-th* level of the pyramid:

$G_l(x,y)$ is obtained from $G_{l-1}(x,y)$ as follows:

$$G_l(x,y) = \sum_{m=1}^{T}\sum_{n=1}^{T} w(m,n)\, G_{l-1}(R_x x + m, R_y y + n)$$

[2] http://www.mathworks.it/it/help/vision/ref/vision.pyramidclass.html

Where: $w(x,y)$ is a $T \times T$ kernel (here T=5) of a low-pass filter; R_x and R_y (here $R_x = R_y = 2$) are the downsampling ratios in x and y directions. An example of the result of the multi-scale pyramid approach applied to a given image is shown in figure 2.

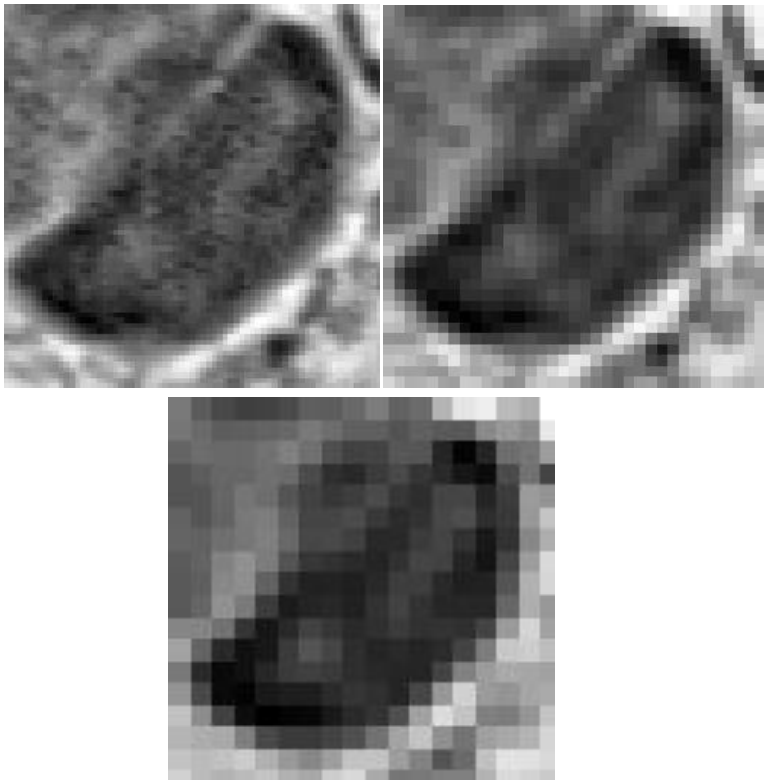

Figure 2. *left*: original image; *center*: first pyramid level image; *right*: second pyramid level image.

The second approach presented in this paper is the Gaussian scale-space representation. In this case, different images are generated using a two-dimensional symmetric Gaussian lowpass filter of size k (here we use k=3 and k=5) with standard deviation 1. The original image is filtered to obtain a set of smoothed versions of the original image.

An example of the result of the multi-scale Gaussian approach applied to a given image is shown in figure 3.

The Support Vector Machine (SVM) [13-15] is used as classifier. In this study, both linear and radial basis function kernels are tested. For each dataset, the best kernel and the best set of parameters are chosen by 5-fold cross validation using the training data.

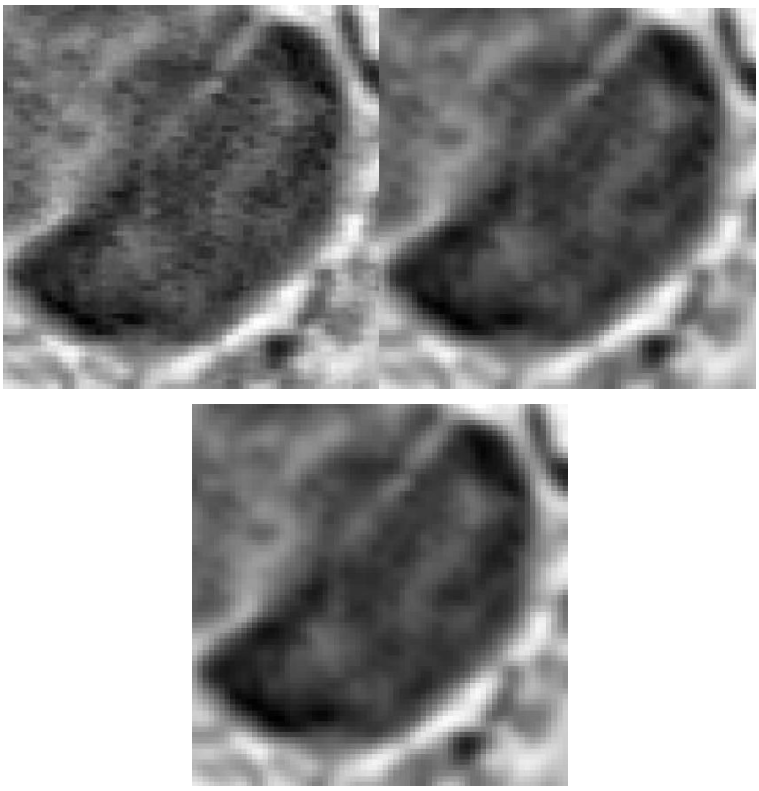

Figure 3. *left*: original image; *center*: image filtered by a lowpass filter of size 3; *right*: image filtered by a lowpass filter of size 5.

3 Datasets

The approaches proposed in this paper have been tested on several different datasets to assess the generality of the developed solution. Such datasets belong to different fields, but the main part is taken from medical imaging, because texture analysis techniques are particularly suited for this kind of images, as previously pointed out. Among the medical datasets, however, very different kinds of image can be found, because e.g. cells, viruses and smoke look very differently, as it can be seen in figure 4.

Datasets used for training and testing the classifier are the following ones:

- PAP, the Pap Smear dataset [20] containing images[4] representing cells to be automatically examined to diagnose cervical cancer.
- VIR, the virus type classification dataset [23] contains images of virus extracted using negative stain transmission electron microscopy. The authors proposed a system that obtains a mean accuracy of ~73.8% using the "object scale dataset" used in this work (the median accuracy among the classes is of 79.0%, but here we report the average mean accuracy). We used the 10-fold validation division shared by the authors but we do not exploit the mask, shared by the authors, for background subtraction, as we used the whole image for extracting the features (since in this way we obtained better results).
- SM, an interesting application of the video surveillance systems is the video-based smoke detection. We used the same subdivision into training/test set proposed in [22] (in which an accuracy of 96.6% was achieved), instead of the 10-fold testing protocol.

- HI, the Histopatology dataset [21] is composed of images from different organs that are representative of the four fundamental tissues.
- BC is a dataset [24] that contains 273 malignant and 311 benign breast cancer images; authors reported an average area under the ROC curve of 0.89.
- PR, the dataset built in [25] containing 118 DNA-binding Proteins and 231 Non-DNA-binding proteins. Texture descriptors are extracted from the 2-D distance matrix that represents each protein, which is obtained from the 3-D tertiary structure of a given protein (considering only atoms that belong to the protein backbone, see [25] for details).
- CHO, a dataset[3] [26] that contains 327 fluorescent microscopy images, belonging to 5 different classes, and is taken from Chinese Hamster Ovary cells. Images are 16 bit grayscale of size 512×382 pixels.

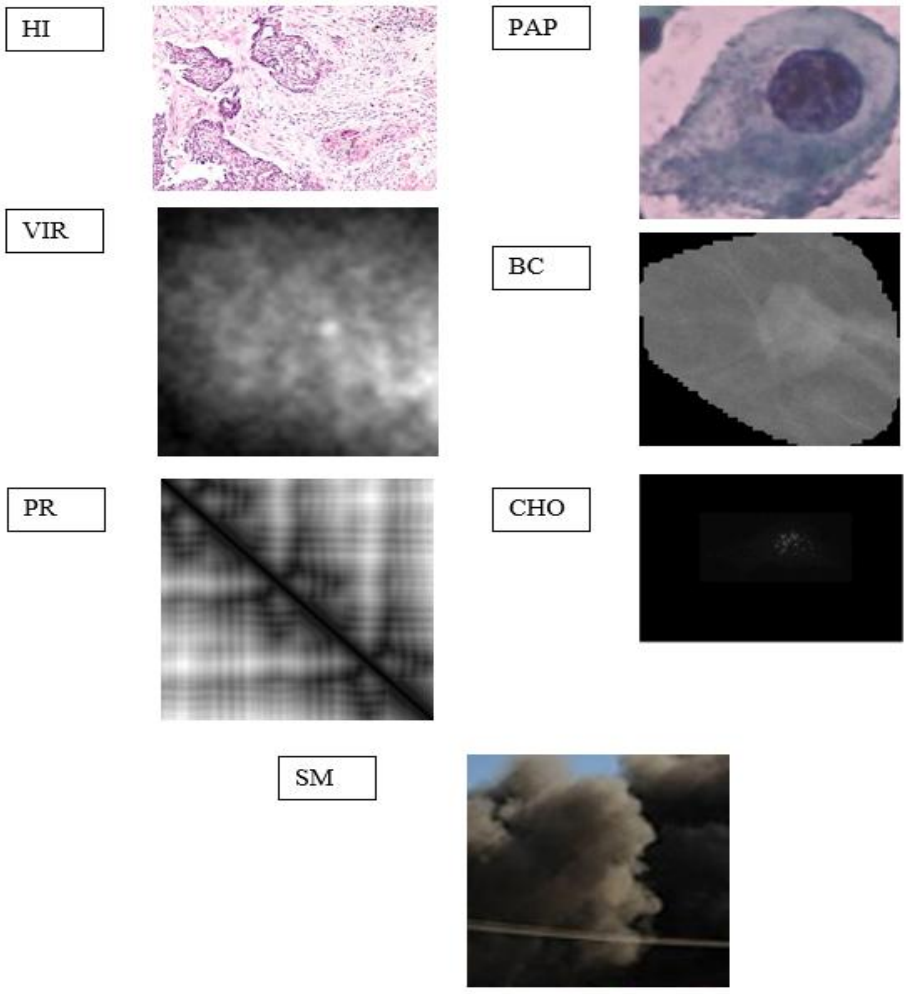

Figure 4. Images of the used datasets.

[3] Download from http://ome.grc.nia.nih.gov/iicbu2008/hela/index.html#cho

A brief description of each dataset is reported in table 1.

Name	Short Name	#Classes	#Samples	Sample size
Histopatology[4]	HI	4	2828	various[5]
Pap smear	PAP	2	917	various[5]
Smoke recognition	SM	2	2868	100×100
Virus types classification	VIR	15	1500	41×41
Breast cancer	BC	2	584	various
Protein classification	PR	2	349	various
Chinese Hamster Ovary	CHO	5	327	512×382

Table 1. Brief description of the datasets

4 Experimental Results

We chose a 10-fold cross-validation as an evaluation protocol for testing each texture descriptor, with the exception of the VIR and SM dataset for which the original testing protocol is used. Two performance indicators were exploited: accuracy (*Acc*)[6] and area under the ROC curve (*AUC*)[7], which provides a better overview of the classification results. AUC is a scalar measure that can be interpreted as the probability that the classifier will assign a higher score to a randomly picked positive sample than to a randomly picked negative sample [48]. In multi-class problem, the AUC is calculated using the one-versus-all approach (a given class is considered as "positive" and all the other classes are considered as "negative") and the average AUC is reported.

The first experiment was aimed at establishing which is the best parameter set for HAR and SHAPE. Results in terms of *Acc* and *AUC* for the different approaches[8] presented in this paper are reported in tables 2 and 3, respectively. Performance has been measured both on single databases and in average, which is the last row of the table, *Av*. Experiments were run considering:

- *HAR*(1), Haralick method using d=1 for extracting the features;
- *HAR*(1,3), the input of SVM is the concatenation of the features obtained with d={1,3};
- *HAR*(1:5), the input of SVM is the concatenation of the features obtained with d={1,2,3,4,5};
- *HAR*(1:3), the input of SVM is the concatenation of the features obtained with d={1,2,3};
- *HAR2*, two SVMs were trained (one with the HAR features extracted with d=1 and the other with d=3) and combined by sum rule;
- *SH*, the shape-based features extracted from the whole co-occurrence matrix and concatenating the features extracted for d={1,3};

[4] the RGB images are converted in gray level images

[5] The images were of different size, we have resized them for obtaining images of at least 100 pixels for both row/column dimensions

[6] i.e. the proportion of "true positives + true negatives" in the population

[7] EUC is implemented as in dd_tools 0.95 davidt@ph.tn.tudelft.nl

[8] We performed several experiments; only the most significant ones are reported here.

- *SH5*, the shape-based features extracted only from the first five windows (the whole co-occurrence matrix and four sub-windows) reported in section 2.1; each of the five descriptors is obtained concatenating the features extracted with d={1,3} and used to train a different SVM. These five SVMs are combined by sum rule;
- *SH*(1), the SHAPE method detailed in section 2.1 where we used d=1 in each of the 13 descriptors;
- *SH*(1,3), the SHAPE method detailed in section 2.1 where we used d={1,3} in each of the 13 descriptors;
- *SH*(1:5), the SHAPE method detailed in section 2.1 where we used d={1,2,3,4,5} in each of the 13 descriptors;

The following conclusion could be drawn from the results:

- The best results are obtained by *HAR*(1,3) and *SH*(1,3);
- Using more sub-windows of the GLDM improves the performance, since *SH*(1,3) outperforms *SH5* and *SH5* outperforms *SH*.
- *HAR*(1,3) outperforms *HAR2*, i.e. better results are obtained when a single SVM is trained with the whole set of features.

Acc	HAR(1)	HAR(1,3)	HAR(1:5)	HAR(1:3)	HAR2	SH	SH5	SH(1)	SH(1,3)	SH(1:5)
PAP	86.67	86.45	85.25	86.78	85.90	84.37	87.43	86.12	87.32	85.46
VIR	66.67	77.27	77.87	77.07	69.33	43.27	55.13	44.73	56.47	54.87
SM	97.01	96.28	97.07	96.21	96.48	95.42	95.55	94.88	95.02	95.68
HI	73.06	77.31	77.17	77.27	74.05	66.02	70.90	69.81	72.64	72.50
BC	85.52	88.45	87.76	88.28	84.31	80.86	82.59	81.72	82.76	80.69
PR	87.76	85.97	85.07	85.07	87.46	80.00	78.81	81.19	80.60	80.90
CHO	95.69	96.31	94.46	96.31	97.23	95.08	95.08	95.69	95.38	92.92
Av	84.62	86.86	86.38	86.71	84.96	77.86	80.78	79.16	81.45	80.43

Table 2. Methods compared using the accuracy as performance indicator.

AUC	HAR(1)	HAR(1,3)	HAR(1:5)	HAR(1:3)	HAR2	SH	SH5	SH(1)	SH(1,3)	SH(1:5)
PAP	88.66	89.23	89.42	**91.33**	88.58	82.46	86.14	**87.17**	86.63	86.01
VIR	94.28	**96.36**	96.10	96.18	94.88	84.59	89.23	85.56	**89.88**	89.76
SM	99.50	99.43	**99.59**	99.40	99.41	99.08	99.25	98.91	99.36	**99.41**
HI	87.65	89.70	89.11	**89.71**	88.13	83.23	86.74	86.52	87.59	**88.37**
BC	93.90	94.05	**94.06**	93.84	90.71	88.76	91.00	91.33	**91.85**	90.98
PR	91.41	**90.10**	87.86	89.02	90.88	84.85	84.70	**85.44**	84.71	84.18
CHO	99.46	**99.71**	99.43	99.55	99.72	99.48	99.57	**99.70**	99.57	99.04
Av	93.55	94.08	93.65	**94.15**	93.18	88.92	90.94	90.66	**91.37**	91.10

Table 3. Methods compared using the error under the ROC curve as performance indicator.

Since the best results are obtained by *HAR*(1:3) and *SH*(1,3) we tested the multi-scale approaches using these two methods as a benchmark. In tables 4 and 5 the following approaches are compared:

- HARg, fusion by mean rule among the *HAR*(1,3) descriptors extracted using the two filtered images (k=3 and k=5);
- HARp, fusion by mean rule among the descriptors extracted using the two images obtained by the pyramidal approach (using l=1 and l=2);
- HARf, fusion by weighted sum rule between *HAR*(1:3) and HARg, the weight of HAR(1:3) is 2 while the weight of HARg is 1;
- SHg, fusion by mean rule among the *SH*(1,3) descriptors extracted using the two filtered images (k=3 and k=5);
- SHp, fusion by mean rule among the *SH*(1,3) descriptors extracted using the two images obtained by the pyramidal approach (with l=1 and l=2);
- SHf, fusion by sum rule between *SH*(1,3) and *SHg*;
- F[9], fusion by weighted sum rule between *HAR*(1:3) and *SH*(1,3), the weight of *HAR*(1:3) is 2 while the weight of *SH*(1,3) is 1;
- *FUS*[9], fusion by weighted sum rule between *HARf* and *SHf*, the weight of *HARf* is 2 while the weight of *SHf* is 1;

Accuracy	**HAR(1:3)**	**HARg**	**HARp**	**HARf**	**SH(1,3)**	**SHg**	**SHp**	**SHf**	**F**	**FUS**
PAP	86.78	85.14	83.17	88.42	87.32	84.37	85.25	86.34	89.18	**89.62**
VIR	77.07	64.93	58.60	78.07	56.47	57.67	30.40	61.20	77.93	**78.33**
SM	96.21	96.15	96.48	96.68	95.02	94.95	92.36	95.35	96.68	**97.08**
HI	77.27	74.97	72.32	**78.02**	72.64	72.21	64.11	74.41	**78.02**	77.98
BC	88.28	83.45	77.24	88.62	82.76	78.79	72.24	83.10	89.33	**89.83**
PR	85.07	86.57	82.09	**86.27**	80.60	84.48	78.21	83.58	84.78	85.67
CHO	96.31	95.08	95.38	96.00	95.38	97.85	95.38	97.23	98.15	**99.08**
Average	86.71	83.75	80.75	87.44	81.45	81.47	73.99	83.03	87.72	**88.22**

Table 4. Fusion approaches, accuracy

AUC	**HAR(1:3)**	**HARg**	**HARp**	**HARf**	**SH(1,3)**	**SHg**	**SHp**	**SHf**	**F**	**FUS**
PAP	91.33	88.15	83.51	91.72	86.63	82.86	84.86	86.55	92.26	**92.49**
VIR	96.18	92.30	91.12	96.23	89.88	90.94	75.35	92.15	96.34	**96.41**
SM	99.40	99.29	99.40	99.50	99.36	99.01	98.33	99.34	99.54	**99.58**
HI	89.71	88.48	87.18	90.10	87.59	88.29	83.74	89.60	90.20	**90.68**
BC	**93.84**	89.39	83.98	93.77	91.85	86.99	80.14	92.35	**95.25**	94.93
PR	**89.02**	86.39	85.45	88.25	84.71	89.34	83.27	87.56	88.62	**89.10**
CHO	99.55	99.28	99.47	99.62	99.57	99.76	99.59	99.81	**99.87**	**99.87**
Average	94.15	91.89	90.01	94.17	91.37	91.02	86.46	92.48	94.58	**94.72**

Table 5. Fusion approaches, area under the ROC curve

It is interesting to point out that even though the HAR approach offers better performance than its HARg and HARp variants, the fusion of the 3 provides higher performance both in average and

[9] Before the fusion the scores of the two methods are normalized to mean 0 and standard deviation 1

in all single datasets but one. This indicates that HARg and HARp variants provide information that cannot be found in HAR.

As it can be seen from the tables, the multi-scale approach outperforms HAR on accuracy (table 2), but does not improve the area under the ROC curve (table 3). Instead the shape-based method has better performance on both indicators when the multi-scale approach is adopted (see *SHf* vs *SH*(1,3)). Moreover, the novel approach *FUS*, based on weighted sum rule, outperforms the fusion *F* based on SHAPE and HAR extracted only from the original image.

The Gaussian approach outperforms the pyramidal approach. It should be noted that pyramidal decomposition was not included in the Gaussian approach because it did not provide any performance improvement.

The proposed approach provides strong performance with respect to others described in the literature. This is the case, for example, of the virus type classification dataset: authors proposed a system that obtains an accuracy of ~73.8%. In the video-based smoke detection authors proposed a system that achieves an accuracy of 96.6%; in breast cancer dataset authors reported an average area under the ROC curve of 89%.

Finally, we report the comparison, see Table 6, between our best method named FUS with one of the best performing variants of local binary pattern: the multi-threshold local quinary coding (MT) [59]. Our results show that the method proposed in this paper outperforms also MT.

Accuracy	**FUS**	**MT**
PAP	**89.62**	86.89
VIR	**78.33**	70.00
SM	**97.08**	96.53
HI	77.98	**80.92**
BC	**89.83**	85.00
PR	**85.67**	81.19
CHO	**99.08**	**99.08**
Average	**88.22**	85.65

Table 6. Comparison between FUS and MT

The main drawback of methods that fuse together several approaches is the required computational time, which makes them unsuitable for real-time applications. The computational time needed to extract HAR(1:3) and SH(1,3) in the HeLa dataset was measured. The computation is quite heavy, because five images are needed (original, the two images obtained by the Gaussian based approach, and the two images obtained by the pyramidal approach). The computational time needed to process 10 images was:

- 19.98 s (i.e. ~2 sec/image) on an Intel i5-2500 processor (4 cores) using Matlab parallel toolbox to exploit the multi-core architecture.
- 15.44 s on an Intel i5-3470 processor (4 cores) using Matlab parallel toolbox to exploit the multi-core architecture.

5 Discussion and Conclusion

In this work we extended the work presented in previous papers on texture analysis techniques based on the co-occurrence matrix and introduced a study on multi-scale approach for Haralick-based features. The two multi-scale approaches applied are a) Gaussian scale-space representation and b) image pyramid decomposition. The best results were obtained using the Gaussian scale-space representation.

We showed that our approach improves the performance of SHAPE by coupling it with a multi-scale approach, moreover also the fusion with standard Haralick-based features has been improved thanks to the introduction of a weighted average over the contributions of the different features. For assessing the results we tested our systems over seven different classification problems, which demonstrated the generality of the approach described in this paper. For all the experiments a SVM classifier was employed. Our results were also compared with some state-of-the-art descriptors.

In our opinion the most valuable result of this paper is that we have shown that it is possible to extract more information from the co-occurrence matrix. As future works we would test different methods for processing the image before extracting the co-occurrence matrix, e.g. Gabor filters or wavelet decomposition. Moreover, we want to test different trained fusion rule [51] as stacking, discriminative accumulation scheme (DAS); Multiclass Multi Kernel Learning (MK).

The code of the methods tested in this paper will be available at
http://www.dei.unipd.it/wdyn/?IDsezione=3314&IDgruppo_pass=124&preview=.

References

[1] Hamilton, N. A., Wang, J. T. H., Kerr, M. C., and Teasdale, R. D., “Statistical and visual differentiation of subcellular imaging,” BMC Bioinformatics, vol. 10, no. 94, 2009.

[2] Murphy, R. F., “Putting proteins on the map,” Nature Biotechnology, vol. 24, no. 10, pp. 1223-1224, 2006.

[3] Karkanis, S. A., Iakovidis, D. K., Maroulis, D. E., Karras, D. A., and Tzivras, M., “Computer-aided tumor detection in endoscopic video using color wavelet features,” IEEE Transaction on Information Technology in Biomedicine, vol. 7, no. 3, pp. 141-52, 2003.

[4] A. Samant,H. Adeli “Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis,” Computer-Aided Civil and Infrastructure Engineering, vol. 15, no. 4, pp. 241-250, 200.

[5] Shen, L., Bai, L., and Fairhurst, M., “Gabor wavelets and general discriminant analysis for face identification and verification,” Image and Vision Computing, vol. 25, pp. 553–563, 2007.

[6] Ameling, S., Wirth, S., Paulus, D., Lacey, G., and Vilarino, F., "Texture-based polyp detection in colonoscopy," Bildverarbeitung für die Medizin, Hans-Peter Meinzer, Thomas Martin Deserno, Heinz Handels et al., eds., pp. 346-50, Berlin and Heidelberg, Germany: Springer, 2009.

[7] Ojala, T., Pietikainen, M., and Maeenpaa, T., “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” Ieee transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[8] Vécsei, A., Amann, G., Hegenbart, S., Liedlgruber, M., and Uhl, A., "Automated marsh-like classification of celiac disease in children using local texture operators," Computers in Biology and Medicine, vol. 41, no. 6, pp. 313-25, 2011.

[9] Oliver, A., Lladó, X., Freixenet, J., and Martí, J., "False positive reduction in mammographic mass detection using local binary patterns," Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Computer Science 4791, pp. 286-293, Brisbane, Australia: Springer, 2007.

[10] Unay, D., and Ekin, A., "Intensity versus texture for medical image search and retrieval " in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008, pp. 241-244.

[11] Nanni, L., and Lumini, A., "A reliable method for cell phenotype image classification," Artificial Intelligence in Medicine, vol. 43, no. 2, pp. 87-97, 2008.

[12] Haralick, R. M., "Statistical and structural approaches to texture," Proceedings of the IEEE, vol. 67, no. 5, pp. 786-804, 1979.

[13] Vapnik, V., and Lerner, A., "Pattern recognition using generalized portrait method," Automation and Remote Control, vol. 24, pp. 774-80, 1963.

[14] Vapnik, V., and Chervonenkis, A., "A note on one class of perceptrons," Automation and Remote Control, vol. 25, 1964.

[15] Kuncheva, L. I., and Whitaker, C. J., "Measures of Diversity in Classifier Ensembles and their Relationship with the ensemble accuracy," Machine Learning, vol. 51, no. 2, pp. 181-207, 2003.

[16] Ghidoni, Stefano and Cielniak, Grzegorz and Menegatti, Emanuele (2012) Texture-based crowd detection and localisation. In: The 12th International Conference on Intelligent Autonomous Systems, June 26-29, 2012, Jeju Island, Korea.

[17] Fernando Roberti de Siqueira, William Robson Schwartz, Helio Pedrini, Multi-scale gray level co-occurrence matrices for texture description, Neurocomputing, http://dx.doi.org/10.1016/j.neucom.2012.09.042

[18] Nanni, L., Ghidoni, S, Brahnam, S., Menegatti, E (2013) A comparison of methods for extracting information from the co-occurrence matrix for subcellular classification. submitted Expert system with applications 2013

[19] Paci, Michelangelo; Nanni, Loris; Lahti, Anna; Aalto-Setala, Katriina; Hyttinen, Jari; Severi, Stefano, Non-Binary Coding for Texture Descriptors in Sub-Cellular and Stem Cell Image Classification, Volume 8, Number 2, April 2013 , pp. 208-219(12)

[20] Jantzen, J., et al., Pap-smear benchmark data for pattern classification, in Nature inspired Smart Information Systems (NiSIS)2005: Albufeira, Portugal. p. 1–9.

[21] Angel Cruz-Roa, Juan C. Caicedo, Fabio A. González, Visual pattern mining in histology image collections using bag of features, Artificial Intelligence in Medicine 52 (2011) 91– 106

[22] Feiniu Yuan. (2011) Video-based smoke detection with histogram sequence of LBP and LBPV pyramids. Fire Safety Journal, 46(3), pp. 132–139.

[23] Gustaf Kylberg, Mats Uppström, Ida-Maria Sintorn, (2011) Virus Texture Analysis Using Local Binary Patterns and Radial Density Profiles. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 16th Iberoamerican Congress, CIARP 2011, Pucón, Chile, November 15-18, 2011.

[24] Junior G.B., Cardoso de Paiva A., Silva A.C., Muniz de Oliveira A.C., (2009) Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM, Computers in Biology and Medicine, Volume 39 , Issue 12, 1063-1072.

[25] Nanni L, Shi JY, Brahnam S, Lumini A. (2010) Protein classification using texture descriptors extracted from the protein backbone image. Journal of Theoretical Biology 2010 Jun 7;264(3):1024-32.

[26] Boland, M. V., Markey, M. K., and Murphy, R. F., "Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images," *Cytopathology*, vol. 33, no. 3, pp. 366-75, 1998.

[27] Loris Nanni, Michelangelo Paci, Sheryl Brahnam, Stefano Ghidoni, Emanuele Menegatti, (2013) Virus image classification using different texture descriptors, submitted IPCV 2013.

[28] Lowe, D.G.. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 2004;60:91–110.

[29] Bay, H., Tuytelaars, T., Van Gool, L.. SURF: Speeded-Up Robust Features. In: European Conference on Computer Vision. 2006, p. 346–359.

[30] Dalai, N., Triggs, B., Rhone-Alps, I., Montbonnot, F.. Histograms of Oriented Gradients for Human Detection. In: IEEE Conference on Computer Vision and Pattern Recognition. 2005, p. 886–893.

[31] Mikolajczyk, K., Schmid, C.. A Performance Evaluation of Local Descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005;27(10):1615 –1630.

[32] Tuzel, O., Porikli, F., Meer, P.. Region Covariance: A Fast Descriptor for Detection And Classification. In: In Proc. 9th European Conf. on Computer Vision. 2006, p. 589–600.

[33] Wu, B., Nevatia, R.. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. In: IEEE International Conference on Computer Vision. 2005, p. 90–97.

[34] Haralick, R.M., Shanmugam, K., Dinstein, I.. Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics 1973;3(6):610–621.

[35] Ojala, T., Pietikinen, M., Harwood, D.. A Comparative Study of Texture Measures with Classification Based on Featured Distributions. Pattern Recognition 1996;29(1):51–59.

[36] Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.. Image Indexing Using Color Correlograms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1997, p. 762–.

[37] Pass, G., Zabih, R., Miller, J.. Comparing Images Using Color Coherence Vectors. In: Proceedings of the ACM international conference on Multimedia. 1996, p. 65–73.

[38] Swain, M.J., Ballard, D.H.. Color Indexing. International Journal of Computer Vision 1991;7:11–32.

[39] Freeman, W.T., Adelson, E.H.. The Design and Use of Steerable Filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 1991;13:891–906.

[40] Jain, A.K., Farrokhnia, F.. Unsupervised Texture Segmentation Using Gabor Filters. Pattern Recognition 1991; 24:1167–1186.

[41] Gelzinis, A., Verikas, A., Bacauskiene, M.. Increasing the Discrimination Power of the Co-occurrence Matrix-based Features. Pattern Recognition 2007;40(9):2367–2372.

[42] Walker, R., Jackway, P., Longstaff, D.. Genetic Algorithm Optimization of Adaptive Multi-Scale GLCM Features. International Journal of Pattern Recognition and Artificial Intelligence 2003;17(1):17–39.

[43] Benco, M., Hudec, R.. Novel Method for Color Textures Features Extraction based on GLCM. Radioengineering 2007;4(16):64–67.

[44] Hu, Y.. Unsupervised Texture Classification by Combining Multi-Scale Features and K-Means Classifier. In: Chinese Conference on Pattern Recognition. 2009, p. 1–5.

[45] Pacifici, F., Chini, M.. Urban Land-Use Multi-Scale Textural Analysis. In: IEEE International Geoscience and Remote Sensing Symposium. 2008, p. 342–345.

[46] Rakwatin, P. and Longepe, N. and Isoguchi, O. and Shimada, M. and Uryu, Y.. Mapping Tropical Forest using ALOS PALSAR 50m Resolution Data with Multiscale GLCM Analysis. In: IEEE International Geoscience and Remote Sensing Symposium. 2010, p. 1234–1237.

[47] Nguyen-Duc, H., Do-Hong, T., Le-Tien, T., Bui-Thu, C.. A New Descriptor for Image retrieval using Contourlet Co-Occurrence. In: Communications and Electronics (ICCE), 2010 Third International Conference on. 2010, p. 169–174.

[48] Fawcett T (2004). ROC Graphs: Notes and Practical Considerations for Researchers, Technical report, Palo Alto, USA: HP Laboratories

[49] Xiao-Ming Ren, Xiao-Feng Wang, Yang Zhao (2012) An Efficient Multi-scale Overlapped Block LBP Approach for Leaf Image Recognition, Intelligent Computing Theories and Applications, Lecture Notes in Computer Science Volume 7390, 2012, pp 237-243

[50] Xueming Qian, Xian-Sheng Hua, Ping Chen, Liangjun Ke, (2011) PLBP: An effective local binary patterns texture descriptor with pyramid representation, Pattern Recognition, Volume 44, Issues 10–11, October–November 2011, Pages 2502-2515

[51] F. Orabona, L. Jie, B. Caputo. (2010) Online-Batch Strongly Convex Multi Kernel Learning. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.