

Evaluation of Statistical Features for Medical Image Retrieval

Cecilia Di Ruberto¹ and Giuseppe Fodde¹

University of Cagliari, Department of Mathematics and Computer Science,
Via Ospedale 72, 09124 Cagliari, Italy, dirubert@unica.it, giufodde@libero.it.

Abstract. In this paper we present a complete system allowing the classification of medical images in order to detect possible diseases present in them. The proposed method is developed in two distinct stages: calculation of descriptors and their classification. In the first stage we compute a vector of thirty-three statistical features: seven are related to statistics of the first level order, fifteen to that of second level where thirteen are calculated by means of co-occurrence matrices and two with absolute gradient; the last thirteen finally are calculated using run-length matrices. In the second phase, using the descriptors already calculated, there is the actual image classification. Naive Bayes, RBF, Support Vector Machine, K-Nearest Neighbor, Random Forest and Random Tree classifiers are used. The results obtained from the proposed system show that the analysis carried out both on textured and on medical images lead to have a high accuracy.

Keywords: texture, feature extraction, feature selection, classification, medical image analysis.

1 Introduction

Texture analysis is a process that allows the characterization of different surfaces and objects through the identification of their specific statistical properties. Through rigorous techniques of image capture, you can get a texture concerning a certain surface that uniquely identifies its structure depending on the lighting and the intensity captured during acquisition. From this you can extract characteristics or features that allow to describe a texture image, through an appropriate mathematical formulation. In this work we propose a complete system that allows the classification of medical images in order to detect possible diseases present in them. The proposed method is developed in two distinct stages: extraction of descriptors and classification. In the first stage we extract thirty-three features from the image: seven are related to statistics of the first level order, fifteen to that of the second level (thirteen are calculated using the co-occurrence matrices and two by the absolute gradient) and, finally, the last eleven are calculated using the run-length matrices. In the second phase, instead, using the descriptors already calculated, there is the actual image classification by using different classifiers. The results obtained from the proposed system are

encouraging and show that the analysis carried out both on textured and on medical images lead to have a high accuracy.

2 Related Works

In literature there are several methods for the extraction of features from medical images, each of them based on a different type of texture. In addition, most of them are only concerned with the extraction of the feature of texture. In the field of Image Processing the term texture refers to any and repetitive geometric arrangement of the gray levels of an image. The texture provides important information about the spatial arrangement of the gray levels and the relationship with their surroundings. The human visual system determines and recognizes easily different types of texture characterizing them in a subjective manner and, although for a human observer it is simple and intuitive to associate a surface with a particular texture, to give a rigorous definition for this is very difficult. In fact, there is no general definition of texture and a methodology for measuring the texture acceptable to all.

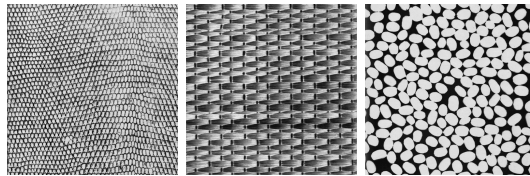


Fig. 1. Examples of textures from Brodatz album [?].

Typically you only use qualitative definitions to describe texture with attributes as coarse, granular, random, ordered, filiform, punctuating, fine-grained, etc (see Fig. ??). It can easily guess that the quantitative analysis of texture passes through statistical and structural relations among the basic elements (the texels) of what we call just texture.

The analysis of texture has three main aspects: classification, segmentation and shape from texture. Classification concerns the search for particular regions of texture among different predefined classes of texture. The classification of textures is carried out using statistical methods that define the descriptors of the texture, i.e. from the images in gray level, a measure of the characteristics of texture is evaluated through co-occurrence matrices, run-length matrix, contrast, homogeneity, entropy, etc.. This type of statistical approach is particularly suitable when the texture is made up of very small and complex elementary primitives, typical of microstructures. Segmentation determines the boundaries between regions with different textures. Segmentation occurs when the previous statistical approaches do not provide accurate measures of the textural characteristics and therefore are insufficient to characterize the texture of a region. In

addition, the segmentation is indispensable when you do not have any knowledge nor on the number of classes of texture features nor on a priori characteristics. Shape from texture, instead, is essential for the reconstruction of the surface of objects starting from the information associated with the macrostructure, such as density, size and orientation.

There are various methods for feature extraction and texture analysis and one very important is the one that makes use of the statistics of various orders useful for our purpose. The statistical approach is particularly suitable for the analysis of microstructures. The elementary primitives of the microstructures are determined by analyzing the characteristics of the texture associated with a few pixels of the image. The approaches of this type comprise three different orders: first, second and higher order. *Statistics of first order* measure the likelihood of observing a gray value in random position in the image. The statistics of the first order can be calculated from the histogram of the image gray levels. This depends only on the single gray level of the pixel and not on the interaction co-occurring with the pixels of the surroundings, thus it does not depend on the absolute position and reciprocal pixels. It is ascertained experimentally that this type of statistics is the most strong, stable and representative of an image. *Statistics of second order* are defined as the likelihood of observing a pair of gray levels measured at the ends of a segment randomly positioned in the image with a random orientation. The scientist Julesz [?] proposed the theory of textons to explain the early mechanisms that occur in the brain in discriminating two pairs of regions with different textures. The textons are visual events (i.e. collinearity, terminations, closures, etc..) whose presence is detected and used to discriminate the texture. Terminations are the heads of a segment or an angle. This category includes all those statistics to determine the mutual correlation between the gray levels. *Statistics of higher order* involve the analysis of higher-order statistics information. This type of statistics contains information regarding the runs of gray levels, which make up the image matrix. This information concerns to the size of these runs, i.e. their lengths. It is possible to define also the sets of consecutive pixels that all have the same level of gray and then go to make up a run.

Image analysis techniques have played an important role in various medical applications. In general, applications involve the automatic features extraction from the images which are then used for classification, such as the distinction of normal or healthy tissue from abnormal or sick one. Depending on the particular classification, the features extracted capture morphological, color or texture properties of the image. The properties of calculated texture are closely related to the application domain to be used. For example, Sutton and Hall [?] discussed the classification of lung diseases by examining texture features. Some diseases, such as interstitial fibrosis affecting the lungs, lead to have change of textures in the radiographic images, on the contrary of the lesions that are clearly delineated. In such applications, the methods of textures analysis are ideal for such images. Sutton and Hall [?] proposed three types of texture features to distinguish the sick pulmonary tissue from the healthy one. These features were calculated on the basis of a measure of isotropic contrast, a measure of direc-

tional contrast and a sampling of Fourier energy domain. In their experiments of classification, the best results were obtained using the directional contrast measure. Harms [?] used the image texture combined with the features of color to diagnose leukemic malignant tumors in samples of blood cells. He extracted particular texture features, such as the "textons" (regions with nearly uniform color), and analyzed the related features, such as the total number of pixels of textons having a specific color, the average radius and the size for each color and various shape features of textons. The features of texture, combined with the color, have increased the percentage of correct classification of type of cell in the blood compared to the classification done using only one color. Landeweerd and Gelsema [?] extracted various statistics of the first order (as the average gray level in a region) and second order (as arrays of co-occurrences of gray levels) to differentiate the different types of white blood cells. Insana [?] used the features of texture of ultrasound images to estimate the parameters of dispersion of the tissue. He made a meaningful use both of knowledge about the physics used in ultrasound images and of characteristics of the tissue to design the model of the texture. Chen [?] used the fractal texture features for classifying the ultrasonic images of the liver and for improving the analysis X-rays of the chest. Lundervold [?] used fractal texture features in combination with other functions to analyze the ultrasound images of the heart (the ultrasound images are temporal sequences of the left ventricle of the heart).

3 The proposed system

Let us now explain the system designed to perform the texture analysis of medical images. The method implemented is divided into two main parts: in the first we have the extraction of texture features while in the second part we have the image classification using the descriptors obtained in the first phase. After choosing to analyze the entire image or only a part of it, we calculate thirty-three features: run-length matrices (eleven features), absolute gradient (two features), first order statistics (seven features) and co-occurrence matrices (thirteen features). This first part is implemented using the Matlab development tool. In the second step, we classify the data previously obtained: through the Weka environment, in fact, we build the dataset containing the calculated descriptors. These data are classified using various types of classifiers. We also use a feature selection, as sometimes the obtained accuracy results are not satisfactory, while in other cases the feature selection do not improve the results already obtained. We can see graphically the system diagram in Fig. ??.

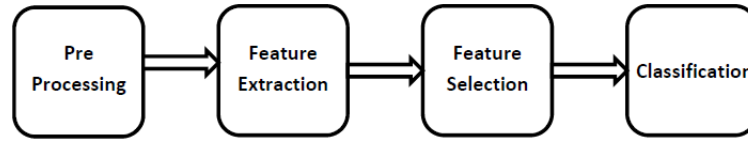


Fig. 2. Proposed system diagram

3.1 Pre-processing

In this first phase, we need sometimes to resize and crop the image manually in order to isolate the regions of interest from medical images and to eliminate portions of them not related to human tissue but to the instruments used during the mammography generation.

3.2 Features extraction

The first features (higher-order statistics) to be analyzed are related to arrays of run lengths (GLRLM), from which we obtain eleven features: Short Run Emphasis, Long Run Emphasis, Gray Level Nonuniformity, Run Length Nonuniformity, Run Percentage, Low Gray-Level Run Emphasis, High Gray Level Run Emphasis, Short Run Low Gray-Level Emphasis, Short Run High Gray-Level Emphasis, Long Run Low Gray-Level Emphasis and Long Run High Gray-Level Emphasis. After the run-lengths matrices, we calculate other two descriptors, still related to higher order statistics, those of the absolute gradient: mean and variance. Other implemented descriptors are the seven statistics of the first order: mean, standard deviation, entropy, kurtosis, uniformity, smoothness and skewness. The latest features (statistics of the second order) are related to the co-occurrence matrices (GLCM) : contrast, angular second moment, correlation, entropy, variance, homogeneity, sum average, sum variance, sum entropy, difference variance, difference entropy, two information measures of correlation, as proposed by Haralick [?].

3.3 Feature Selection

This phase belongs to the second part of our system when, using the descriptors calculated, we have the step of classification of real images. During this step, we select the most predictive attributes based on their correlation with the attribute class (positive correlation) in descending order. The most frequently features removed from feature selection, both for Brodatz's images and for DDSM database, are: contrast, correlation, homogeneity, difference variance, difference entropy, Info measure correlation 1 and 2 from GLCM and Gray Level non Uniformity and Run Percentage from GLRLM. This part is introduced why during the testing phase we have achieved not very satisfactory accuracy results. It was then decided to try to improve these results adopting this technique that, in some cases, has led to an improvement of the initial accuracy up to 10% more.

In few cases, however, the feature selection has not brought improvements to the results already obtained.

3.4 Classification

The phase of classification is the latest one. In fact, after the descriptors extractions and the feature selection, then we have to choose the classifiers that in our approach are six: SVM, RBF, Random Tree, Random Forest, Nave Bayes and k-NN. For this stage we decide to use the technique of ten fold cross-validation where the original dataset is divided into subsets each composed by the same number of samples (in all our cases ten). The data are first classified by analyzing the original dataset and by applying the classifiers listed above, and then by making the feature selection on the original dataset and by proceeding with the application of the various classifiers.

4 Experimental results and comparisons

The implemented system is applied to two different types of images. The first type includes all the images taken from Brodatz album [?]. Texture is fundamental for analyzing medical tissue and methods of texture analysis are widely studied, confirming the utility and the goodness in classifying a human tissue as healthy or sick. Therefore, we also apply our system to a medical database, the DDSM (Digital Database for Screening Mammography) [?]. We test the performance of our system by comparing the accuracy value obtained by our classification and that presented by eight alternative methods. We analyze the same subset of images (from Brodatz's album) proposed in each paper. The accuracy values, that is a measure of how many images are correctly classified after feature selection, are obtained using six different classification models, as described in section 3. In all the cases, the feature selection never decreases the accuracy level obtained without it. Finally, the best result is compared with the accuracy value presented in each considered alternative method in classifying the same set of images. Let's show the comparison results for each of them.

4.1 Brodatz album

In this experiment, we use 37 classes of textures of Brodatz album from which we create non-overlapping sub-images. We compare the performance of our system on the same set of Brodatz textures analyzed in each considered alternative method, each time creating a sub-set of textures ad hoc for each of them. Naive Bayes, RBF, Random Tree, Random Forest, k-Nearest Neighbors and Support Vector Machine classifiers are used.

In the paper [?] for classification of Brodatz textures five feature sets are calculated: Sobel edge detector, discrete cosine transform (DCT), speeded up robust features (SURF), gray level run length matrix and the eigenvalues. The classification is performed using decision trees, random trees and support vector

machines. The first set of images (Set 1) is composed by the following classes: D1, D5, D6, D15, D21, D41, D49, D60, D67, D74, D82, D87, D94, D102, D109. The average accuracy is 68.36% and the best result is 84.44% obtained with the SVM. In our work the best result is 100% obtained with the KNN. Paper [?] presents a new feature selection scheme that automatically determines a reduced subset of methods whose integration produces classification results comparable to those obtained when all the available methods are integrated but with a significantly lower computational cost. The second set of image (Set 2) is composed by: D3, D15, D32, D37, D41, D54, D91, D94. The average classification is 91.13% and the best result is 96.8% obtained with all methods. In our work the best result is 100% obtained with Random Forest and KNN. In the paper [?] a combined statistical and structural approach is used. The spectrum of image is used for unsupervised texture classification. The third set of images (Set 3) is composed by two classes: D84 and D112. An average correct classification of 96% is obtained. In our work the best result is 100% obtained with all classifiers. In the paper [?] features computed as statistics (e.g. histograms) of local filters responses are reported as the most powerful descriptors for texture classification and segmentation. The set of images (Set 4) is composed by: D3, D4, D6, D21, D24, D49, D68, D71, D82, D87. The average classification is 89.86%. In our work the best result is 93.5% obtained with the KNN. In the paper [?] texture images are splitted into 64 non-overlapping sub-images and then these sub-images are decomposed through wavelet transformation to obtain sub-band images. These sub-band images are further used to extract statistical texture features. For images classification neural networks are used. The set of images (Set 5) is composed by: D9, D12, D15, D19, D21, D24, D29, D30, D112. The average classification is 92,58% and the best result is 100% obtained with the MRBF. In our work the best result is 100% obtained with KNN. In the paper [?] the main focus is to do texture segmentation and classification: GLCP method is used to extract features from texture images and GSVM have been proposed to do classification on extracted features. The set of images (Set 6) is composed by: D4, D6, D57, D64. The average classification is 95,13% and the best result obtained is 98,49% with GSVM. In our work the accuracy varies from 74,2% to 100% and the best result is 100% obtained using all classifiers except the Random-Tree. In the paper [?] a set of invariant descriptors of each image is extracted. These descriptors are vector-quantized from key-points. PLSI and NMF to perform unsupervised classification are used. The set of images (Set 7) is composed by: D35, D64, D74, D99. The average classification is 62,92%. In our work the accuracy varies from 95% to 100% where the best result is 100% obtained using all classifiers except the Random-Tree and Random-Forest. The paper [?] describes the usage of wavelet packet neural networks (WPNN) for texture classification problem. The proposed schema is composed of a wavelet packet feature extractor and a multi-layer perceptron classifier. Set 8 is composed by: D9, D12, D15, D16, D19, D28, D54, D68, D94, D112. The overall success rate is about 95%. In our work the success rate varies from 83% to 98,6% obtained with KNN and RBF.

The best accuracy values obtained by our system and those of other methods present in literature are reported in table ???. As it can be seen from the results shown in the table, the our method leads to a high accuracy in the classification of textured images.

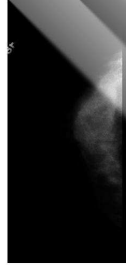
Images	Our System		Other Methods	
	<i>Classifier</i>	<i>Accuracy</i>	<i>Classifier</i>	<i>Accuracy</i>
Set 1	KNN	100%	SVM[?]	84.44%
Set 2	R-Forest	100%	All Methods[?]	96.8%
Set 3	All Class.	100%	AVG[?]	96%
Set 4	KNN	93.5%	WAVELET[?]	96.9%
Set 5	KNN	100%	MRBF[?]	100%
Set 6	SVM	100%	GSVM[?]	98.49%
Set 7	RBF	100%	PLSI[?]	64.46%
Set 8	KNN	98.6%	WPNN[?]	95.7%

Table 1. The best accuracy values on Brodatz textures achieved by our method and others: Set 1,..., Set 8 are the sets of textures analysed by each method.

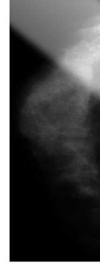
4.2 Digital Database for Screening Mammography (DDSM)

This database [?] contains a series of mammography screenings stored in four different categories: Normal, Cancer, Benign, Benign Without Callback. For each of these categories we use 20 different images, for a total of 80 images to classify. The classifiers are the same used with Brodatz images. Again, we first make a classification with the original data, then we refine it by using a feature selection. The process starts downloading from the archive some images for each case contained in DDSM. More specifically, eighty pictures are downloaded in total, twenty for each case. The feature extraction are made both on the entire mammographic image and on a cropped sub-image after eliminating superfluous background (see Fig. ??). The first images analyzed are those related to the "normal" case, followed by "cancer", "benign" and finally "benign without callback" cases. Once calculated values for the first two cases, the results obtained are used to create a first dataset consisting of only two classes ("normal" and "cancer"), to see the behavior of the system in the diagnosis of possible diseases. The other dataset is composed proceeding with the calculation of descriptors for the last two cases and integrating them to the previous one, going to get a multi-class classification in which there are all the four categories listed above. So, we have two datasets: one composed by forty images and only two classes and the other one composed of eighty images and four classes.

In [?] the classification is conducted by CAD algorithms where the success rate is about 95%. Our best classification rate is 98,5% obtained by Random-Forest. The best accuracy values obtained by our system are reported in table ???.



(a) Original Image



(b) Cropped Image

Fig. 3. A mammography image example from DDSM [?].

Classes	Original Images				Cropped Images			
	Without FS		With FS		Without FS		With FS	
<i>Class</i>	<i>Acc</i>	<i>Class</i>	<i>Acc</i>	<i>Class</i>	<i>Acc</i>	<i>Class</i>	<i>Acc</i>	
2 classes	RBF 92.6%	SVM 92.6%	KNN 93.5%	R-Tree 97.1%				
4 classes	KNN 81.4%	KNN 81.4%	KNN 96.2%	R-Forest 98.5%				

Table 2. The best accuracy values (Acc) on mammography images (without Feature selection (FS) and with Feature Selection) for the relative classifier (Class).

5 Conclusions and Future Works

The purpose of this work is to combine various statistical methods for features descriptors calculation in order to classify images containing different textural information. We combine together the various statistics of first, second and higher level order to see if a joined application of these methods is more discriminating than the application both of a different approach and similar methods. The main aim is to look for a system that is able to find in medical images various possible diseases through a texture analysis quickly. The results obtained in different experimental phases show how the implemented system is able to massively discriminate the textures of the images analyzed going to track pathologies within a diseased tissue. The rates of accuracy are just fine, but we achieve further refinements through feature selection. So, the implemented method is able to identify various diseases quickly and very satisfactorily. This work could be further developed going to implement and test other methods for texture analysis as the wavelet transforms, the Gabor filters and the Fourier spectrum and to see so if the addition of new descriptors allows a further discrimination compared to that already obtained, thus to increase the values of accuracy already achieved. In addition, it is also necessary to increase the size of the dataset to allow a better classification, without maybe having feature selection. Finally, given the encouraging experimental results, of course our future interests and possible developments will include a detailed analysis of a possible dependance analysis

among the chosen features.

Acknowledgments

This work has been funded by Regione Autonoma della Sardegna (R.A.S.) Project CRP-17615 DENIS: Dataspace Enhancing Next Internet in Sardinia.

References

1. Brodatz, P.: Texture: A photographic album for Artists and Designers. Dover Publications, (1966)
2. Julesz, B.: Textons, the Elements of Texture Perception, and Their Interactions. *Nature* 290, 91–97 (1981)
3. Sutton, R., Hall, E. L.: Texture Measures for Automatic Classification of Pulmonary Disease. *IEEE Transactions on Computers* C-21, 667–676 (1972)
4. Harms, H., Gunzer, U., Aus, H. M.: Combined Local Color and Texture Analysis of Stained Cells. *Computer Vision, Graphics, and Image Processing* 33, 364–376 (1986)
5. Landeweerd, G. H., Gelsema, E. S.: The Use of Nuclear Texture Parameters in the Automatic Analysis of Leukocytes. *Pattern Recognition* 10, 57–61 (1978)
6. Insana, M. F., Wagner, R. F., Garra, B. S., Brown, D. G., Shawker, T. H.: Analysis of Ultrasound Image Texture via Generalized Rician Statistics. *Optical Engineering* 25, 743–748 (1986)
7. Chen, C. C., Daponte, J. S., Fox, M. D.: Fractal Feature Analysis and Classification in Medical Imaging. *IEEE Transactions on Medical Imaging* 8, 133–142 (1989)
8. Lundervold, A.: Ultrasonic Tissue Characterization - A Pattern Recognition Approach, Technical Report. Norwegian Computing Center Oslo Norway, (1992)
9. Haralick, R. M.: Statistical and Structural Approaches to Texture. *IEEE Proceedings* 7, 786–804 (1979)
10. Hearsh, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, W. P.: The digital database for Screening Mammography. *Medical Physics Publishing*, 212–218 (2001)
11. Desager, C., Geerts, S., Van der Schueren, F., Ledda A.: Texture Analysis and Classification of Brodatz Textures using ESAC Framework. *THE E-LAB MASTERS THESES 2010-2011*, (2011)
12. Puig, D., Garcia, M.: Automatic texture feature selection for image pixel classification. *Pattern Recognition* 39, 1996–2009 (2006)
13. Umarani, C., Ganesan, L., Radhakrishnan, S.: Combined Statistical and Structural Approach for Unsupervised Texture Classification. *International Journal Of Imaging Scienze And Engineering*, 162–165 (2008)
14. Karoui, I., Fablet, R., Boucher, J. M., Pieczynski, W., Augustin, J. M.: Fusion of textural statistics using a similarity measure: application to texture recognition and segmentation. *Pattern Analysis & Applications* 11, 425–434 (2008)
15. Chuan-Yu, C., Shih-Yu, F.: Image Classification using a Module RBF Neural Network. *Proceedings of the First International Conference on Innovative Computing, Information and Control*, (2006)
16. Hee-Kooi, K., Hong-Choon, O., Ya-Ping, W.: Image Texture Classification using Combined Grey Level Co-occurrence Probabilities and Support Vector Machines. *Proceedings of Fifth International Conference on Computer Graphics, Imaging and Visualization*, 180–184 (2008)

17. Lei, Q., Qingfang, Z., Shuqiang, J., Qingming, H., Wen, G.: Unsupervised texture classification: Automatically discover and classify texture patterns. *Image and Vision Computing* 26, 647–656 (2008)
18. Sengur, A., Turkoglu, I., Ince Cevdet, M.: Wavelet packet neural networks for texture classification. *Expert Systems with Applications* 32, 527–533 (2007)