



ISSN: 0067-2904 GIF: 0.851

Texture Features Analysis using Gray Level Co-occurrence Matrix for Abnormality Detection in Chest CT Images

Faleh H. Mahmood^{1*}, Wafaa A. Abbas²

¹ Remote sensing Research Unit, College of Science, University of Baghdad, Baghdad, Iraq ²College of Pharmacy, University of Baghdad, Baghdad, Iraq

Abstract

Texture is an important characteristic for the analysis of many types of images because it provides a rich source of information about the image. Also it provides a key to understand basic mechanisms that underlie human visual perception. In this paper four statistical feature of texture (Contrast, Correlation, Homogeneity and Energy) was calculated from gray level Co-occurrence matrix (GLCM) of equal blocks (30×30) from both tumor tissue and normal tissue of three samples of CT-scan image of patients with lung cancer. It was found that the contrast feature is the best to differentiate between textures, while the correlation is not suitable for comparison, the energy and homogeneity features for tumor tissue always greater than its values for normal tissue.

Keywords: Texture feature, Co-occurrence matrix, CT-scan, statistical feature.

تحليل خصائص النسيج بأستخدام مصفوفة الحدوث (GLCM) لكشف الشذوذ في الصور المقطعية للحديث

فالح حسن محمود 1*، وفاء عبدالامير عباس²

¹ وحدة الاستشعار عن بعد ، كلية العلوم ، جامعة بغداد، بغداد ، العراق ² فرع العلوم المختبرية السريرية ، كلية الصيدلة، جامعة بغداد، بغداد ، العراق

الخلاصة

النسيج هو من الخصائص المهمة لتحليل أنواع متعددة من الصور لأنها توفر مصدرا غنيا من المعلومات حول الصورة. كما أنه يوفر مفتاح لفهم الآليات الأساسية التي تكمن وراء الإدراك البصري البشري. في هذا البحث، تم حساب الخصائص الإحصائية للنسيج (التباين، الارتباط، ، التجانس والطاقة) من مصفوفة مصفوفة الحدوث (GLCM) لكتل متساوية (30×30) لكل من نسيج الورم والأنسجة الطبيعية وباستخدام ثلاث عينات لصور اشعة مقطعية لمرضى مصابين بسرطان الرئة . وقد تبين أن ميزة التباين هو أفضل للتمييز بين القوام، في حين أن الارتباط غير مناسب للمقارنة، وميزات الطاقة والتجانس لنسيج الورم دائما أكبر من قيمها للأنسجة الطبيعية.

Introduction

Image analysis techniques have played an important role in several medical applications. In general, the applications involve automatic extraction of features from the image which is then used for a variety of classification tasks, such as distinguishing normal tissue from abnormal tissue [1]. Texture is an important characteristic for the analysis of many types of images [2-4] because it

*Email:faleh_sine@yahoo.com

provides a rich source of information about the image [4], among the traditional application areas of texture analysis are industrial inspection, biomedical image analysis, etc. A texture area in an image can be characterize by non-uniform or varying spatial distribution of intensity , the intensity variation reflects some change in the scene being imaged [5].

Texture feature are complex visual patterns composed of entities or sub patterns that have specific characteristics (brightness, slop, and size) [6]. Texture is attractive not only because it is an important component in image analysis for solving a wide range of applied recognition, segmentation, and synthesis problems, but also it provides a key to understand basic mechanisms that underlie human visual perception [7]. Texture features refer to visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity. These features contain important information about the structural arrangement of surfaces and their relationship to the surrounding environment [8]. Simplest texture features can achieve by calculating statistical properties, like mean and variance from the gray level histogram of the image. However, the performance of these kinds of first-order statistics is usually poor. Haralick et al. (1973) [9] calculated second-order gray scale statistics using gray-level co-occurrence matrices (GLCM) and defined fourteen statistical measures for texture. Co-occurrence matrices give information about patterning of the texture, and it co-occurrence used to calculate textural properties from them. These features are sensitive to illumination variations, but have been very popular in different texture analysis applications [10].

Method

In this paper, an image classification system proposed based on Haralick texture features extracted from a slice of DICOM Lung CT images, to differentiate between lung cancer tissue and normal lung tissue. The system process is shown in Figure-1, start by Choice of two sections of equal size from different tissues in the original input image, and then create multiple GLCM from the two sections by specifying an array of offsets. These offsets define in pixel relationships of varying direction and distance in four directions and 30 distances as follow:

Offsets = [01; 02; 03; 04;; 030]	(0°, Horizontal direction(H)).
[-1 1; -2 2; -3 3; -4 4;; -30 30]	(45°, Right-down direction (RD)).
[-1 0; -2 0; -3 0; -4 0;; -30 0]	(90°, Vertical direction (V)).
[-1 -1; -2 -2; -3 -3; -4 -4;; -30 -30]	(135°, Left-down direction (LD)).

Several statistics were derived from created GLCMs. These statistics provide information about the texture of an image. The statistical features calculated in this work are: Contrast, Correlation, Energy and Homogeneity of multiple GLCMs for each normal and tumor section. The following sections will describe in detail the above stages. The system implemented with MATLAB software version R2014a.



Figure 1-Block diagram of implemented system

Texture Features Analysis

An important approach to region description is to quantify is texture content. Although no formal definition to texture exists, intuitively this descriptor provides measures of properties such as smoothness, coarseness and regularity. The three principal approaches used in image processing to

describe the texture of the region are statistical, structural, and spectral. Statistical approaches yield characterization of textures as smooth, coarse, grainy, and so on [11].

Texture analysis methods are divided into two main categories: *stochastic* and *structural*. The approaches differ in their principles of describing and treating textures. In a stochastic model, textures are considered to be formed by random processes. These kinds of textures are analyzed studying the statistical properties of the intensity values of pixels. In the structural approach, texture is considered to consist of textural primitives, often called textons that are located on the texture with certain placement rules. [10]

Gray Level Co-Occurrence Matrix (GLCM)

Several texture metrics that contain spatial information are based on the co-occurrence matrix, they also known as the spatial gray-level dependence matrix. Forming the co-occurrence matrices is an initial step that compiles spatial as well as statistical information for computing the texture metrics described later. The spatial information considered is the relative position of pairs of pixels, defined with distance *d* and orientation θ that describe the location of the second pixel with respect to the first. A co-occurrence matrix is formed for each such position. In this manner, each co-occurrence matrix prepares the data to emphasize primarily structure or streaks in a given direction and a grain size that is at least as large as the selected distance. Typically, four values of θ , namely 0°, 45°, 90°, and 135°, cover the orientations, and the most common choice of distance is d = 1 when θ is 0° or 90°, and

$d = \sqrt{2}$ when θ is 45° or 135 [12].

Mathematically, for a given image *I* of size $K \times K$, the elements of a gray-level Co-occurrence matrix M_{CO} for a displacement vector d (= dx, dy) is defined as:

$$M_{co} = \sum_{x=1}^{K} \sum_{y=1}^{K} \begin{cases} 1, & \text{if } I(x, y) = i \text{ and } I(x + d_x, y + d_y) = j \\ 0, & \text{otherwise} \end{cases}$$
.....1

A single GLCM might not be enough to describe the textural features of the input image. For example, a single horizontal offset might not be sensitive to texture with a vertical orientation. For this reason, create multiple GLCMs for a single input image. To create multiple GLCMs, specify an array of offsets. These offsets define pixel relationships of varying direction and distance. For example, an array of offsets that specify four directions (horizontal, vertical, and two diagonals) and four distances are defined. In this case, the input image is represented by 16 GLCMs. The calculated statistics features from these GLCMs, we can take the average of each feature in four directions for the same distance and plot versus offset.

Such as illustrated in the following example which creates an offset that specifies four directions and 4 distances for each direction.

Offsets = [0 1; 0 2; 0 3; 0 4; -1 1; -2 2; -3 3; -4 4; -1 0; -2 0; -3 0; -4 0; -1 -1; -2 -2; -3 -3; -4 -4;]

Figure-2 illustrates the spatial relationships of pixels that are defined by this array of offsets, where D represents the distance from the pixel of interest [13, 14].



Figure 2-Co-occurrence matrix offset for extracting texture features in 4 directions and 4 distances [15].

Haralick Features

In 1973, Haralick [9] introduced 14 statistical features. These features are generated by calculating the features for each one of the co-occurrence matrices obtained by using the directions 0° , 45° , 90° , and 135° , then averaging these four values. The distance parameter can be selected as one or higher. A vector of these 14 statistical features is used for characterizing the co-occurrence matrix contents [16], only four of them are defined here:

• *The Contrast*: Measures the local contrast of an image. The Contrast is expected to be low if the gray levels of each pixel pair are similar, given by:

$$Contrast = \sum_{i}^{M} \sum_{j}^{N} (i - j)^{2} P[i, j]$$
(2)

• *The Correlation*: Provides a correlation between the two pixels in the pixel pair. The Correlation is expected to be high if the gray levels of the pixel pairs are highly correlated.

$$Correlation = \sum_{i}^{M} \sum_{j}^{N} \frac{(i-\mu)(j-\mu)P[i,j]}{\sigma^{2}}$$
(3)

• *The Homogeneity*: Measures the local homogeneity of a pixel pair. The Homogeneity is expected to be large if the gray levels of each pixel pair are similar, given by:

$$Homogeneity = \sum_{i}^{M} \sum_{j}^{N} \frac{P[i, j]}{1 + |i - j|}$$

$$\tag{4}$$

The Energy: Measures the number of repeated pairs. The Energy is expected to be high if the occurrence of repeated pixel pairs is high, given by:

$$Energy(Angular Second Moment) = \sum_{i}^{M} \sum_{j}^{N} P^{2}[i, j]$$
(5)

Where M, N are the dimension of the image P[i,j] for $i=1,2,3,\ldots,M$, $j=1,2,\ldots,N$. μ and σ are the mean and standard deviation of image respectively. [9, 16-18]. **Results and Discussion**

The introduced proposed system in this paper have been tested on several digitized lung CT scan images from three patients with lung cancer, first input the images ,then selected two blocks from input image of each sample one of them from tumor tissue and the other from normal tissue of size (30×30) as shown in Figures-3, 4 and 5.

Multiple GLCMs produced from the two blocks by specifying an array of offsets to the program. These offsets define pixel relationships of varying directions (horizontal, vertical, and two diagonals) and 30 distances for each direction. In this case, the selected subimage (Block) is represented by 120 GLCMs, then four statistical features (Contrast, Correlation, Energy and Homogeneity) calculated from these GLCMs, and take the average for each feature in four direction for the same distance see Tables-1, 2 and 3, To facilitate the comparison between features of tumor tissue and normal tissue each feature plotted versus offset.

As Figure-6 show that there are a clear difference in each feature for tumor tissue and normal tissue with the same offset and for different samples, where; in the contrast, there is a big difference between them within the offset 5 - 20, and the contrast of normal tissue is greater than the contrast of tumor tissue, while high correlation appears in small distance between pairs of pixels, then decrease as the distance increase, this is true for tumor and normal tissues for all samples. Energy and homogeneous are with the same behavior, where the value of these features of tumor tissue is greater than normal tissue in particular within the offset 0-25 for the energy and 5-20 for the homogeneous. These behaviors frequent in all tested samples.



Figure 3-Two selected sections of tumor and normal tissue from sample1 and the sections histogram.



Figure 4- Two selected sections of tumor and normal tissue from sample 2 and the sections histogram.



Figure 5-Two selected sections of tumor and normal tissue from sample 3 and the sections histogram.

offset	(*,,,,,	Tum		a set of 30 distances for sample 1. <i>Normal region</i>				
No.	Contrast	Corr.	Energy	Homogeneity	Contrast	Corr.	Energy	Homogeneity
1	0.003	0.520	0.992	0.998	0.127	0.708	0.552	0.936
2	0.004	0.000	0.992	0.998	0.242	0.464	0.479	0.886
3	0.004	0.001	0.992	0.998	0.327	0.305	0.433	0.854
4	0.004	0.001	0.991	0.998	0.389	0.206	0.401	0.829
5	0.005	0.000	0.991	0.998	0.441	0.126	0.383	0.813
6	0.005	0.001	0.990	0.998	0.492	0.042	0.371	0.800
7	0.005	0.000	0.990	0.997	0.514	0.011	0.363	0.792
8	0.006	-0.250	0.990	0.997	0.517	0.005	0.368	0.793
9	0.007	-0.250	0.989	0.997	0.495	0.023	0.386	0.801
10	0.005	0.000	0.989	0.997	0.484	0.011	0.406	0.808
11	0.006	0.000	0.989	0.997	0.478	-0.006	0.420	0.811
12	0.006	nan*	0.988	0.997	0.474	-0.027	0.429	0.814
13	0.007	0.000	0.987	0.997	0.462	-0.016	0.435	0.818
14	0.007	0.000	0.986	0.996	0.457	-0.015	0.443	0.822
15	0.008	0.000	0.984	0.996	0.454	-0.017	0.450	0.820
16	0.009	0.000	0.983	0.996	0.460	-0.039	0.453	0.820
17	0.010	0.000	0.981	0.995	0.457	-0.055	0.459	0.821
18	0.011	nan*	0.979	0.995	0.430	-0.048	0.475	0.828
19	0.012	0.000	0.976	0.994	0.386	-0.027	0.497	0.837
20	0.014	nan	0.972	0.993	0.344	0.013	0.512	0.844
21	0.016	0.000	0.968	0.992	0.305	0.064	0.541	0.854
22	0.016	0.000	0.970	0.992	0.247	0.117	0.596	0.878
23	0.019	nan*	0.964	0.991	0.206	0.068	0.651	0.897
24	0.023	0.000	0.955	0.988	0.181	0.033	0.695	0.910
25	0.030	0.000	0.943	0.985	0.183	-0.154	0.754	0.927
26	0.041	nan*	0.925	0.979	0.159	-0.098	0.773	0.934
27	0.061	nan*	0.896	0.970	0.113	0.007	0.804	0.943
28	0.102	0.000	0.853	0.949	0.153	0.014	0.764	0.923
29	0.216	0.000	0.853	0.892	0.139	0.013	0.771	0.930
30	0.290	0.000	0.925	0.855	0.097	0.017	0.836	0.952

 Table 1- The average of statistical features (Contrast, Correlation, Energy and Homogeneity) in four directions

 (0°, 45°, 90°, 135°) for tumor and normal tissue in a set of 30 distances for sample 1.

*represent there is no correlation

offset	offset Tumor region				Normal region				
No.	Contras t	Corr.	Energy	Homogeneity	Contrast	Corr.	Energy	Homogeneity	
1	0.156	0.685	0.377	0.922	0.192	0.859	0.251	0.904	
2	0.294	0.407	0.302	0.853	0.407	0.708	0.189	0.824	
3	0.396	0.205	0.273	0.802	0.667	0.537	0.153	0.758	
4	0.474	0.049	0.265	0.763	0.921	0.378	0.136	0.714	
5	0.519	-0.039	0.265	0.740	1.150	0.242	0.127	0.681	
6	0.537	-0.075	0.267	0.731	1.325	0.133	0.122	0.658	
7	0.533	-0.064	0.267	0.734	1.468	0.039	0.122	0.641	
8	0.517	-0.031	0.265	0.742	1.588	-0.040	0.124	0.631	
9	0.502	0.001	0.263	0.749	1.697	-0.114	0.128	0.618	
10	0.492	0.025	0.262	0.754	1.741	-0.152	0.131	0.610	
11	0.474	0.062	0.267	0.763	1.756	-0.171	0.134	0.600	
12	0.474	0.061	0.272	0.763	1.721	-0.158	0.138	0.601	
13	0.466	0.077	0.277	0.767	1.660	-0.145	0.149	0.608	
14	0.479	0.055	0.275	0.761	1.602	-0.138	0.163	0.617	
15	0.514	-0.014	0.278	0.743	1.566	-0.176	0.173	0.629	
16	0.542	-0.069	0.288	0.729	1.526	-0.238	0.188	0.642	
17	0.558	-0.099	0.295	0.721	1.437	-0.300	0.205	0.652	
18	0.586	-0.158	0.299	0.707	1.308	-0.337	0.225	0.659	
19	0.601	-0.194	0.308	0.700	1.172	-0.341	0.248	0.666	
20	0.599	-0.190	0.321	0.700	1.078	-0.256	0.261	0.671	
21	0.582	-0.171	0.330	0.709	1.023	-0.185	0.272	0.670	
22	0.570	-0.163	0.336	0.715	0.960	-0.145	0.291	0.667	
23	0.545	-0.155	0.357	0.727	0.920	-0.162	0.315	0.660	
24	0.485	-0.069	0.414	0.758	0.894	-0.238	0.334	0.660	
25	0.450	-0.091	0.521	0.775	0.822	-0.394	0.349	0.671	
26	0.433	-0.227	0.586	0.783	0.723	-0.283	0.369	0.692	
27	0.442	-0.268	0.596	0.779	0.652	-0.407	0.400	0.712	
28	0.442	-0.247	0.604	0.779	0.573	-0.356	0.481	0.735	
29	0.492	-0.196	0.683	0.754	0.625	-0.332	0.638	0.709	
30	0.500	-0.205	0.674	0.750	0.597	-0.322	0.642	0.712	

 Table 2- The average of statistical features (Contrast, Correlation, Energy and Homogeneity) in four directions (0°, 45°, 90°, 135°) for tumor and normal tissue in a set of 30 distances for sample 2

offset		Tum		et of 30 distances for sample 3 Normal region				
No.	Contrast	Corr.	Energy	Homogeneity	Contrast	Corr.	Energy	Homogeneity
1	0.209	0.473	0.450	0.896	0.459	0.780	0.280	0.827
2	0.341	0.144	0.388	0.831	1.192	0.445	0.238	0.753
3	0.399	0.011	0.370	0.804	1.885	0.147	0.223	0.712
4	0.408	-0.009	0.368	0.801	2.327	-0.026	0.205	0.689
5	0.401	0.009	0.368	0.804	2.465	-0.054	0.192	0.673
6	0.400	0.016	0.365	0.802	2.415	-0.005	0.185	0.668
7	0.398	0.015	0.369	0.803	2.326	0.058	0.165	0.648
8	0.380	0.031	0.385	0.812	2.278	0.090	0.162	0.645
9	0.362	0.054	0.403	0.822	2.289	0.092	0.165	0.650
10	0.369	0.028	0.404	0.819	2.445	0.055	0.156	0.631
11	0.385	0.001	0.397	0.810	2.553	0.015	0.160	0.629
12	0.389	-0.001	0.398	0.809	2.513	-0.017	0.165	0.633
13	0.373	0.050	0.404	0.817	2.235	-0.034	0.173	0.638
14	0.367	0.075	0.403	0.818	2.047	-0.016	0.180	0.647
15	0.377	0.078	0.393	0.813	2.038	0.038	0.193	0.657
16	0.406	0.036	0.377	0.799	2.102	0.056	0.190	0.647
17	0.437	-0.013	0.365	0.787	2.151	0.029	0.204	0.656
18	0.469	-0.077	0.358	0.776	1.923	0.038	0.218	0.667
19	0.488	-0.118	0.358	0.772	1.441	0.074	0.248	0.691
20	0.491	-0.104	0.362	0.773	1.161	0.095	0.280	0.715
21	0.453	0.016	0.349	0.787	1.049	0.100	0.302	0.738
22	0.416	0.096	0.355	0.799	1.057	0.064	0.316	0.736
23	0.371	0.187	0.395	0.816	1.005	0.109	0.344	0.751
24	0.428	0.009	0.406	0.787	0.928	0.051	0.365	0.765
25	0.502	-0.194	0.424	0.769	0.836	-0.068	0.365	0.766
26	0.556	-0.246	0.450	0.768	0.714	0.072	0.391	0.787
27	0.496	-0.133	0.457	0.783	0.667	0.249	0.428	0.792
28	0.380	0.055	0.491	0.824	0.716	0.267	0.403	0.756
29	0.310	0.135	0.586	0.856	0.784	0.164	0.549	0.724
30	0.484	0.006	0.677	0.769	0.718	0.151	0.665	0.759

Table 3-The average of statistical features (Contrast, Correlation, Energy and Homogeneity) in four directions(0°, 45°, 90°, 135°) for tumor and normal tissue in a set of 30 distances for sample 3



Figure 6- The average of statistical features (Contrast, Correlation, Energy and Homogeneity) in four directions as a function of offset for sections from tumor tissue and normal tissue of sample 1, 2 and 3.

Conclusion

There are many texture features that can be extracted from the gray-level co-occurrence matrix. In our work, an evaluation of the Haralick texture features is done in order to identify the most significant features that can be used differentiate abnormalities within the lungs for cancer versus normal. The four commonly used features-contrasts, correlation, energy and homogeneity-are studied to differentiate between tumor tissue and normal tissue. Our results indicate that the best feature can be used to differentiate between them is the contrast, followed by each of the Energy and the homogeneous, while the correlation is inappropriate to distinguish between the two types of tissue. There is still further work that can be done in the detecting of the abnormality within the lungs to detect the type of that abnormality whether it will be a lung cancer or not.

References

- 1. Tuceryan, M. and Jain A. 1998. Texture analysis, in *The Handbook of Pattern Recognition and Computer Vision*, 2nd Edition, World Scientific Publishing Co., pp:207-248.
- **2.** Rajalakshmi M. and Subashini P. **2014**. Texture Based Image Segmentation of Chili Pepper X-Ray Images Using Gabor Filter, *International Journal of Advanced Studies in Computer Science* & *Engineering*(IJASCSE), 3(3):44-51.
- **3.** Saad Al-Momen, Loay E. George and Raid K. Naji.**2015**. Texture classification using spline, wavelet decomposition and fractal dimension, *Applied and Computational Mathematics*, science publishing group,4(1):5-10.
- 4. Hatam. H. 2009. Breast Cancer Tissues Recognition Using Fractal Dimension, Fuzzy Logic and Neural Network, M.Sc. Thesis, Department of Computer Science, College of Science, Al-Nahrain University, Iraq.
- 5. Bhiwani, R.J., Khan M.A., and Agrawal, S.M.2010.Texture Based Pattern Classification, *International Journal of Computer Applications*. 1(1):54-56.
- 6. Dalia N. Abdul Wadood. 2014. Detection and Localization of Non-Melanoma Skin Cancer Using Texture Analysis, M.Sc. Thesis, Department of Computer Science, College of Science, Baghdad University, Iraq.
- 7. Zhou. D. 2006. Texture Analysis and Synthesis using a Generic Markov-Gibbs Image Model, Ph.D Thesis, Computer Science, Auckland University.
- 8. Suraksha Ransingh and Manjusha Singh.2015. A Survey Paper on Image Retrieval Based on Colour Feature, *International Journal on Recent and Innovation Trends in Computing and Communication*, 3(2):96-98.
- 9. Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein. 1973. Textural Features for Image Classification, *IEEE Transactions On Systems, Man and Cybernetics*, SMC-3(6):610-21.
- **10.** Markus Turtinen. **2007**.*Learning and Recognizing Texture Characteristics Using Local Binary Patterns*, Academic dissertation, University of Oulu, Finland, C 278, p.21.
- **11.** Rafael C. Gonzalez and Richard E. Woods.**2002**. *Digital Image Processing*, Third Edition, Pearson International Edition,pp:849.
- **12.** Isaac N. Bankman.**2009**.*HandBook of medical image processing and analysis*, Elsevier Inc.,pp:270.
- 13. Ozdemir, I., Norton, D., Ozkan, U. Y., Mert, A. and Senturk, O. 2008. Estimation of tree size diversity using object-oriented texture analysis and ASTER imagery, sensors, 8:4709-4724.
- 14. Benazir . K.K and Vijayakumar.2012.Fingerprint Matching by Extracting GLCM Features, International Conference & Workshop on Recent Trends in Technology, (TCET)2012, *International Journal of Computer Applications* (IJCA),pp:30-34.
- **15.** Girisha A B, M C Chandrashekhar and M Z Kurian.**2013**.FPGA Implementation of GLCM", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*,2(6):2618-2621.
- **16.** Eleyan A. and Dem'irel, H. **2011**. Co-occurrence matrix and its statistical features as a new approach for face recognition, *Turk J Elec Eng & Comp Sci*, 19(1):97-107.
- 17. Gipp, M., Marcus, G., Harder, N., Suratanee, A., Rohr, K., König R., and Männer, R. 2009. Haralick's Texture Features Computed by GPUs for Biological Applications, *International Journal of Computer Science*, 36(1).
- **18.** Zayed N. and Elnemr, H. A. **2015**. Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities, *International Journal of Biomedical Imaging*, 2015, Article ID 267807, pp:1-7.