Feature Extraction of Images Texture Based on Co-occurrence Matrix

Hadia S. Abd allah¹, Wijdan Yassen², khalil I. Alsaiif³

¹Department of Computer Science, College of Computer Sciences & Mathematics, University of Mosul, Iraq.
²Department of Computer Science, College of Education for pure Science, University of Basrah, Iraq.

ABSTRACT:

There are many techniques to extracted object properties in an image. In this research a co-occurrence matrix has been adopted for feature extraction of English letters. English letters of size 14 and font time new roman have been stored as image, then preproced by apply truncation to take off all blank area, then filtered to make it noise free. Energy, contrast, correlation and homogeneity of the co-occurrence matrix properties for the stored character images were calculated. Another character models with different size and fonts were adopted to make the database able to cover a wide range of character images for character recognition and classification. Applied technique shows that a companion properties can be extracted as new properties for letters images and give good results.

The experimental results of the proposed algorithm have proved that both energy and homogeneity features have given high recognition compared with the remaining other properties.

KEY WORDS: GLCM, Feature Extracted, Texture.

INTRODUCTION:

Texture can be adopted for identifying regions of interest in an image due to the important characteristic which can be extracted from it. That’s information plays an important approach in image analysis ( B. Sebastian, et al.). Although several descriptors can be characteristic which can be extracted from it. That’s information plays an important approach in image analysis ( B. Sebastian, et al.). proposed to extract and analyze texture, the development of automatic systems for image interpretation and object recognition is a difficult task due to the complex aspects of texture. Scale is an important information in texture analysis, since a same texture can be perceived as different texture patterns at distinct scales (F.Roberti, et al). Texture is another feature that can help to segment images into regions of interest and to classify those regions, because in some images it is critical in obtaining a correct analysis and can't define characteristic of regions. The image of (Fig. 1) has three distinct regions of textures: the texture of the tiger, the texture of the jungle, and the texture of the water. These textures can be quantized and used to
identify the object classes they represent (Computer Vision, mar2000).

Figure (1): An image with different regions, each have of distinct texture

2. Related work

Feature extracting is widely started when the digital instrument came up and so many researchers pay attention to this field. In 2015 (Neelima B., et al., 2015) gave his paper in feature extraction using texture and shape for content based image retrieval, and has adopt the comparative study on feature extraction. Also (Xin Z. et al., 2017) with his research a study for texture feature extraction of high resolution satellite images based on a direction measure and gray level co occurrence matrix fusion algorithm provide an idea about feature extraction which reach high performance to recognize the solution of problems for high resolution satellite images on direction measure and GLCM. In addition feature extraction is used to support the medical direction for extracting features about some important parts, in MRI image in his research Maayad K. et al., use modern technique to segment some peace using contourlet transformation (Maayad K. et al., 2010).

In this research adopt GLCM (Grey-Level Co-occurrence Matrix) for feature extraction to classify images of English letters and other images.

3. Feature Extraction

Feature Extraction will be involved with texture calculation of a data set by using Gray Level Co-occurrence Matrix (GLCM) algorithm. Classifier is an important field in the image processing which mainly depend on feature extraction such as Energy, entropy, contrast, correlation and homogeneity which are essential properties commonly used to extracted the feature values (S. Yamuna, et al., 2015).

Features computed from GLCM are based on the assumption that the texture information in an image is contained in the overall spatial relationship of grey levels between neighboring pixels. GLCM hold wide information about the image, so features can be derived from GLCM, and usually these features are importance (A. Gebejes, et al., 2013).

4. Co-occurrence Matrices:

A co-occurrence matrix is a matrix that is defined over an image to be the distribution of co-occurring value at a given offset. Suppose that you want to record how often certain transitions occur as you go from one pixel to another. Define a spatial relationship such as “to the left of”, “above”, etc. The co-occurrence matrix for this relationship count the number of times that a pixel with value i occurs with relationship with a pixel with value j. Co-occurrence matrices are mainly used to describe region texture, but they can also be used on image maps to measure how often pixels with certain labels occur with certain relationships to other labels (Clausi, et al., 2002). Gray level co-occurrence texture features assume
that the texture information in an image is contained in the overall spatial relationships among the pixels in the image. This is done by first determining the Gray level Co-occurrence Matrix. This is an estimate of the second order probability density function of the pixels in the image. Then features are obtained from the GLCM based on statistical approaches.

The estimated GLCM matrix can be defined in Equation (1), where GLCM \((n,m)\) represent the number of occurrences of pixels with gray levels \(n\) and \(m\) respectively with a separation of \((dr, dc)\) pixels as shown in (Fig. 2). The number of pixels over which this estimate is obtained is given by Equation (2). If the GLCM is normalized with respect to \(R\), its entries then represent the probability of GLCM occurrence of pixel pairs with gray levels \(n\) and \(m\) with separation \((dr,dc)\) where \(dc = 0\) and \(dr\) between 1 and 10 (Rose F. walker, et al., 1995, Howard D., et al., 2013). A clear basic example to generate GLCM matrix can be seen in (Fig. 3) Howard D., et al., 2013).

\[
glc(n, m) = \sum_{(i,j),(i+dr,j+dc) \in ROI} I(img(i, j) = n, img(i + dr, j + dc) = m) \quad \ldots(1)
\]

\[
R_{glcm} = \sum_{(i,j),(i+dr,j+dc) \in ROI} I \quad \ldots(2)
\]

Where

\(I=1\) satisfy relation, \(I=0\) not satisfy

![Figure (2): generation of GLCM (n,m)](image)

![Figure (3): Process Used to Create the GLCM](image)
5. Description of Two-dimensional Co-occurrence Matrices:

Two-dimensional co-occurrence (gray-level dependence) matrices, proposed by Haralick in 1973, are generally used in texture analysis because they are able to capture the spatial dependence of gray-level values within an image (Haralick R. M., et al., 1973). A 2D co-occurrence matrix, P, is an n x n matrix, where n is the number of gray-levels within an image. For reasons of computational efficiency, the number of gray level scan be reduced if one chooses to bin them, thus reducing the size of the co-occurrence matrix. The matrix acts as an accumulator so that P[i, j] counts the number of pixel pairs having the intensities i and j.

Pixel pairs are defined by a distance and direction which can be represented by a displacement vector d = (dx, dy), where dx represents the number of pixels moved along the x-axis, and dy represents the number of pixels moved along the y-axis of an image slice.

In order to quantify this spatial dependence of gray-level values, calculating various textural features proposed by Haralick (Haralick R. M., et al., 1973, Haralick R. M., et al., 1992), including Entropy, Energy, Contrast, Homogeneity, and Correlation. For the formulas and the intuitive interpretations of these features with respect to the texture characterization, refer to the table (1).

Table 1: Some features of co-occurrence

| Feature        | Formula                                      | What is measured                                           |
|----------------|----------------------------------------------|-----------------------------------------------------------|
| Entropy        | \(-\sum_{i}^{M} \sum_{j}^{N} P[i, j] \log P[i, j]\) | Measures the randomness of gray-level distribution. It is expected to be high if the gray levels are distributed randomly throughout the image. |
| Energy         | \(\sum_{i}^{M} \sum_{j}^{N} P[i, j]^2\)       | Measures the number of repeated pairs. It is expected to be high if the occurrence of repeated pixel pairs is high. |
| Contrast       | \(\sum_{i}^{M} \sum_{j}^{N} (i - j)^2 P[i, j]\) | Measures the local contrast of an image. It is expected to be low if the gray levels of each pixel pair is similar. |
| Homogeneity    | \(\sum_{i}^{M} \sum_{j}^{N} \frac{P[i, j]}{1 + |i - j|}\) | Measures the local homogeneity of the pixel pair. It is expected to be large if the gray levels of each pixel pair are similar. |
| Correlation    | \(\sum_{i}^{M} \sum_{j}^{N} \frac{(i - \mu)(j - \mu)P[i, j]}{\sigma^2}\) | Provides a correlation between the two pixels in the pixel pair. It is expected to be high if the gray levels of the pixel pairs are highly correlated. |
6. Proposed algorithm:

The proposed algorithm goes in two phases, the **first phase** will achieved to find the properties for English letter images in addition to some images, by adopting GLCM matrix, then try to classify the letters and images depending on the four properties got from the co-occurrence matrix by applying the following:

- Images acquisition for English letters.
- Apply preprocess on these images (truncation and filtering).
- Evaluate properties (energy, contrast, correlation, homogeneity) of the co-occurrence matrix for these images.

- Store the acquisition images and their properties in a database.

While in the **second phase** a character will be acquisition then a preprocess will apply in the same way of the first phase then the four properties of the co-occurrence matrix will be evaluated to look for the closest character inside database within a threshold value, in case of no match found, the acquisition character will added to the database with its properties as shown in (Fig. 4).

**Figure (4): The proposed system**

7. Results Discussion:

The characteristics extraction is an important stages in the pattern recognition system which takes the qualities that achieve high accuracy in distinguishing and processing speed. The extraction characteristics must achieve the objectives of correct at the option of the qualities learned and reducing the input data to the discrimination system gives accurate recognition or classification. Applying the proposed algorithm on English letters type Times New Roman, font 14, then obtain results as shown in Table(2) to create database to be used later in recognition phase as shown in (Fig 5).

| Letters | Contrast | Correlation | Energy | Homogeneity |
|---------|----------|-------------|--------|-------------|
| A       | 568.5507 | 0.2124      | 0.0048 | 0.1176      |
| B       | 66.5698  | 0.0225      | 0.0116 | 0.2154      |
| C       | 418.2400 | -0.5141     | 0.0100 | 0.1861      |
| D       | 572.4856 | -0.0694     | 0.0036 | 0.1412      |
| E       | 493.6790 | 0.4522      | 0.0062 | 0.1711      |
| F       | 381.2909 | -0.5687     | 0.0091 | 0.1945      |
| G       | 26.8088  | 0.0582      | 0.0147 | 0.2817      |
| H       | 29.0652  | -4.7587     | 0.0109 | 0.3055      |
| I       | 22.8810  | 2.4965      | 0.0238 | 0.3252      |
| J       | 428.6495 | -0.5507     | 0.0103 | 0.1829      |
| K       | 23.2927  | 0.0999      | 0.0122 | 0.3833      |
| L       | 408.8119 | -0.4770     | 0.0099 | 0.2181      |
| M       | 47.0571  | 0.0689      | 0.0095 | 0.2849      |
| N       | 26.7971  | 0.0477      | 0.0145 | 0.3360      |
| O       | 31.3200  | 0.0150      | 0.0133 | 0.2432      |
| P       | 361.2222 | -0.5784     | 0.0085 | 0.1944      |
| Q       | 312.5507 | -0.4707     | 0.0072 | 0.1996      |
| R       | 311.1752 | -0.5207     | 0.0073 | 0.2396      |
| S       | 20.0000  | -0.0147     | 0.0175 | 0.3810      |
| T       | 21.6800  | -2.0245     | 0.0200 | 0.3133      |
| U       | 395.0472 | -0.5338     | 0.0094 | 0.1999      |
| V       | 414.9100 | -0.5478     | 0.0100 | 0.2398      |
| W       | 47.5000  | 0.0831      | 0.0104 | 0.2741      |
| X       | 20.8030  | 0.2052      | 0.0152 | 0.4225      |
| Y       | 20.5500  | 0.0738      | 0.0167 | 0.3772      |
| Z       | 379.0090 | -0.5311     | 0.0090 | 0.2010      |
Figure(5): The properties a co-occurrence matrix (contrast, correlation, energy, and homogeneity) for English letters

The proposed algorithm applied on same image (X1, X2, X3, X4, X5) but in different format (gif, png, bmp mono, bmp 16 color, bmp 256 color) as shown in (Fig. 6), then obtained results of the feature extracted as in Table(3) which clearly seen in (Fig. 7).

Table 3: Represented feature extraction by Gloms for image in different format

| Image | Contrast | Correlation | Energy | Homogeneity |
|-------|----------|-------------|--------|-------------|
| X1    | 4.8245   | 0.0070      | 5.6774 | 0.0490      |
| X2    | 4.8224   | 0.0074      | 5.5599 | 0.0490      |
| X3    | 5.1024   | 0.0201      | 1.9455 | 0.0495      |
| X4    | 4.7924   | 0.0060      | 5.4473 | 0.0490      |
| X5    | 4.7736   | 0.0022      | 5.8510 | 0.0494      |

Figure(6)): Represented the same image in different format (X1, X2, X3, X4, X5)
The proposed algorithm can be applied on any image of letters to obtain a big database used later in recognition or classification. By applying the proposed algorithm on character(A) image with varying sizes to show the effect of each of the four properties, the energy, homogeneity and contrast characteristics give a high recognition specially when the size changed as shown in (Fig. 8).

Figure 7: The properties a co-occurrence matrix (contrast, correlation, energy, and homogeneity) for images(X1,X2,X3,X4,X5)

Figure 8: The properties a GLCM (contrast, correlation, energy, and homogeneity) for A images in different size
But the correlation property is fluctuating and as shown in (Fig.8). The algorithm applied on the same character (A) in different font and same sizes show the four properties are fluctuating as shown in (Fig. 9).

Figure (9): The properties a GLCM (contrast, energy, and homogeneity) for A image in different font

8. Conclusion:
From this present work, we can concluding that GLCM prove to be effective method to quantify the surface texture of the images, spacially the contrast energy and homogeneity characteristics are better than the other characteristics because they are give high clear difference (for the images and characters which have different sizes) compared to the correlation factors and other properties.

9. References:
B. Sebastian, A. Unnikrishnan, and K. Balakrishnan, "Grey level Co-occurrence Matrices: Generalization and some new features".
F. Roberti, W. Robson, and H. Pedrini, "Multi-Scale Gray Level Co-Occurrence Matrices for Texture Description".
ComputerVision : Mar2000/ https://courses.cs.washington.edu/courses/cse576/book/ch9.pdf.
S. Yamuna, S. Abirami, 2015, "Feature Extraction of Face Value Through Gray Level Co-Occurrence Matrix", International Journal of Open Information Technologies ISSN: 2307-8162 vol. 3, no. 6.
A. Gebejes, R. Huertas, 2013, "Texture Characterization based on Grey-Level Co-occurrence Matrix", Conference of Informatics and Management Sciences, ICTIC.
Clausi, David A., 2002, “An analysis of co-occurrence texture statistics as a function of gray level quantization,” Can. J. Remote Sensing; Vol. 28, No.1, pp. 45-62.
Rose F. Walker, Paul Jackway, I. D. Longstaff, 1995 "Improving co-occurrence matrix feature discrimination", proceedings of DICTA '95, pp.643-648.
Howard D., Mark B., "Neural Network Toolbox User’s Guide", The MathWork, Inc, 2013.
Haralick, R.M., K. Shanmugam, and Dinstein, 1973, “Textural Features for Image Classification,” IEEE Transactions on Systems Man, and Cybernetics; Vol. Smc-3, No.6, pp. 610-621.
Haralick, R.M. and L.G. Shapiro, 1992, "Computer and Robot Vision", Addison-Wesley Publishing Co.
Xin Z., Jintian C., Weisheng W. and Chao L., 2017, "A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm", Sensors 17, 1474; doi: 10.3390/s17071474.

Neelima B. and Punit K.J., 2015, "A comparative study on feature

Maayad K. Al-Naymi, 2010, "region detection using Contourlet transformation in image", M.sc. thesis in computer science department, college of computers and mathematics, university of Mosul, Iraq. extraction using texture and shape for content based image retrieval", International Journal of Advanced Science and Technology Vol.80.