An Advanced Computational Method to Determine Co-occurrence Probability Texture Features

David A. Clausi, Yongping Zhao
Department of Systems Design Engineering
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
Email : dclausi@uwaterloo.ca, Tel: (519) 888 – 4567 x2604

***Abstract*-A critical shortcoming of determining co-occurrence probability texture features using Haralick's popular grey level co-occurrence matrix (GLCM) is the excessive computational burden. Here, a more robust algorithm (the grey level co-occurrence integrated algorithm or GLCIA) to perform this task is presented. The GLCIA is created by integrating the preferred aspects of two algorithms: the grey level co-occurrence hybrid structure (GLCHS) and the grey level co-occurrence hybrid histogram (GLCHH). The GLCHS utilizes a dedicated 2-d data structure to quickly generate the probabilities and apply statistics to generate the features. The GLCHH uses a more efficient 1-d data structure to perform the same tasks. Since the GLCHH is faster than the GLCHS yet the GLCHH is not able to calculate features using all available statistics, the integration of these two methods generates a superior algorithm (the GLCIA). The computational gains vary as a function of window size, quantization level, and statistics selected. The GLCIA computational time relative to that of the standard GLCM method ranges from 0.04% to 16%. The GLCIA is a highly recommended technique for anyone wishing to calculate co-occurrence probability texture features, especially from large-scale digital imagery.**

Index Terms-Texture features, data structures, co-occurrence probabilities, remote sensing imagery, computational efficiency.

I. INTRODUCTION

A popular method for texture feature extraction in remote sensing image interpretation is the use of co-occurrence probabilities, first introduced by Haralick et al. using the grey level co-occurrence matrix (GLCM) [1]. Since the GLCMs are usually sparse, excessive computation is required to generate the co-occurrence texture features. Other methods have been used to improve on the computational demands of the GLCM approach [2][3][4][5].

Unser [2] investigated sum and difference histograms to store the co-occurring data, which are efficient relative to using matrices. Clausi and Jernigan [3] improved on the GLCM by presenting a grey level co-occurrence linked list (GLCLL) structure that stores the non-zero co-occurring probabilities in a sorted linked list. An advancement on the GLCLL is illustrated by Svolos and Todd-Pokropek [4] where a tree data structure is used to store the co-occurring probabilities. This is an improvement since the tree structure search has order O(log m) compared to O(m) of the GLCLL approach. Another computational improvement on the GLCLL is the grey level co-occurrence hybrid structure (GLCHS) [5]. The GLCHS is more efficient than the GLCLL since it avoids the need for maintaining sorted linked lists by using a combined hash table and linked list data structure. The GLCHS is also an improvement on the tree data structure since the search order in the GLCHS is based on a hash table ie. O(1).

3	3	3	1	1
1	3	3	1	1
1	3	2	3	1
1	2	2	1	1
2	2	2	2	1

a

$i \backslash j$	0	1	2	3
0	0	-	-	-
1	0	0.15	-	-
2	0	0.15	0.2	-
3	0	0.25	0.1	0.15

b

Fig. 1. a. 5 x 5 image window with four grey levels values (0-3)
b. Corresponding GLCM given δ =1 pixel and θ = 0° and 180°. Only lower triangle matrix required.

In this paper, Unser's sum and difference histograms are implemented using a GLCHS framework. The resulting implementation (the grey level co-occurrence hybrid histogram or GLCHH) is demonstrated to improve the computational performance relative to the GLCHS alone. However, the sum and difference histograms are not able to act as a basis to apply all of Haralick's statistics exactly [2]. For these statistics, the GLCHS is used. Consequently, an integrated method called the grey level co-occurrence integrated algorithm (GLCIA) is created, which offers a dramatically improved solution by combining the GLCHS and GLCHH methods.

Determination of co-occurrence probability texture features requires two steps. First, the co-occurring probabilities for the given window are determined. The probabilities are a function of the number of quantized grey levels (G), the window size, and the relative spatial pixel separation distance. Fig. 1 shows the creation of the co-occuring probabilities stored in a GLCM. Second, statistics are applied to these probabilities to produce the texture features. Dissimilarity, contrast, entropy, uniformity, inverse difference moment, maximum probability, inverse difference, and correlation are eight statistics often used among the fourteen original statistics mentioned by Haralick et al. [1].

This paper will first briefly discuss Unser's sum and difference histograms. A description of a GLCHH implementation of the histogram method follows. From this, the design and implementation of the GLCIA is presented. Computational speed testing completes the paper.

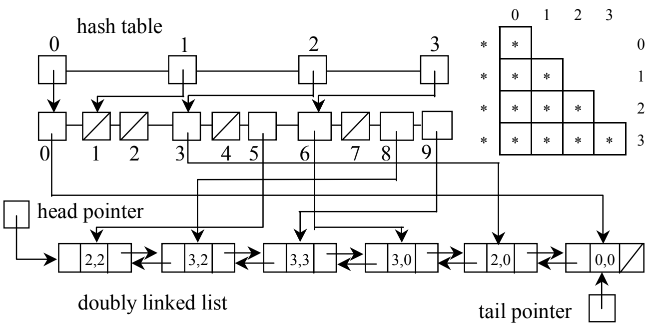

Fig. 2. GLCHS structure for determining image texture features. Nodes created in linked list and hash table based on sample image in Fig. 1A.

II. Sum and Difference Histograms

Unser developed a means of using vectors in the form of sum and difference histograms for the purpose of generating co-occurrence texture features [2]. The normalized sum and difference histograms can be used to determine co-occurrence texture features. The computational advantage is that summations over vectors of length $2G - 1$ are used as opposed to summations over matrices of size GxG in the GLCM approach.

The drawback of the sum and difference histograms is that they cannot be used to determine all of Haralick's statistics exactly [2]. For example, the statistics uniformity, entropy, and maximum probability cannot be determined exactly using sum and difference histograms (although estimates of two of these terms, namely uniformity and entropy, are provided by Unser). Another shortcoming is that, with increasing G and/or decreasing window size, the histograms become sparse, which reduces the computational effectiveness of the method.

III. Algorithm Development

Given the computational shortcoming of the sum and difference histogram method, implementation of this method using the GLCHS data structure was performed. The GLCHS provides a more efficient means of storing the sum and difference histograms and determining the texture features. Fig. 2 displays the basic structure. A hash table is used to instantly access an existing, unique grey level. Non-null hash table pointers point to a linked list which provides a fast means of applying the various statistics.

A. GLCHH

Given the computational shortcoming of the sum and difference histogram approach, these vectors may be implemented using the hybrid data structure used by the GLCHS. These pointer-based data structures will run faster than standard vectors. This implementation is referred to as the grey level co-occurrence hybrid histogram (GLCHH). The normalized sum and difference histograms must each be represented using a hybrid data structure. These are referred to as the grey level co-occurrence hybrid sum histogram (GLCHSH) and the grey level co-occurrence hybrid difference histogram (GLCHDH).

This concept is illustrated in Fig. 3. Fig. 3A represents the sum histogram using a hash table with pointers pointing to nodes on a linked list and Fig. 3B represents the comparable implementation for a difference histogram. Since a histogram is represented, only a one-dimensional hash table is created which is accessed by a single key (either the sum or the difference). To accommodate the lower triangle GLCM (ie. $i \geq j$) for the GLCHDH (Fig. 5B) and maintain consistency across the earlier algorithms, the hash table only needs to contain G elements.

B. GLCIA

Since only sums and differences need to be determined using the GLCHH implementation, it represents a faster algorithm to calculate co-occurrence texture features compared to the GLCHS, which requires a two-dimensional hash table with longer linked lists. However, not all the statistics commonly used can be calculated using sum and difference histograms. As a result, the GLCHH method and the GLCHS method can be integrated to produce a preferred method for determining co-occurrence probability texture features. This method is called the grey level co-occurrence integrated algorithm (GLCIA) [6]. Note that the methods compared here calculate identical texture feature values.

Decisions concerning which data structures to create are made based on which statistics have been selected. Algorithm efficiencies (both speed and memory) are introduced by limiting, when necessary, the co-occurring information collected from the window. Not all GLCHH statistics require both sum and difference histograms. For any of dissimilarity, contrast, inverse difference moment, and inverse moment, the GLCHDH data structure is required. For correlation, both the GLCHDH and GLCHSH are required. The GLCHS is used for any of the statistics uniformity, entropy, and maximum probability.

IV. Testing and Results

A. Methodology

Testing is performed on a Sun Sparc Ultra 1 200E (200 MHz, 128 Mbytes RAM, 322 SPECint, 462 SPECfp) workstation using a four-class 128 x 128 Brodatz [7] test image (Fig. 7). Six window sizes (5x5, 10x10, 15x15, 20x20, 25x25, and 30x30) and five quantization levels (128, 64, 32, 16, and 8) are used as parameters. The interpixel displacements (δ_x, δ_y) are selected as {(1, 0), (1, 1), (0, 1), (-1, 1)}.

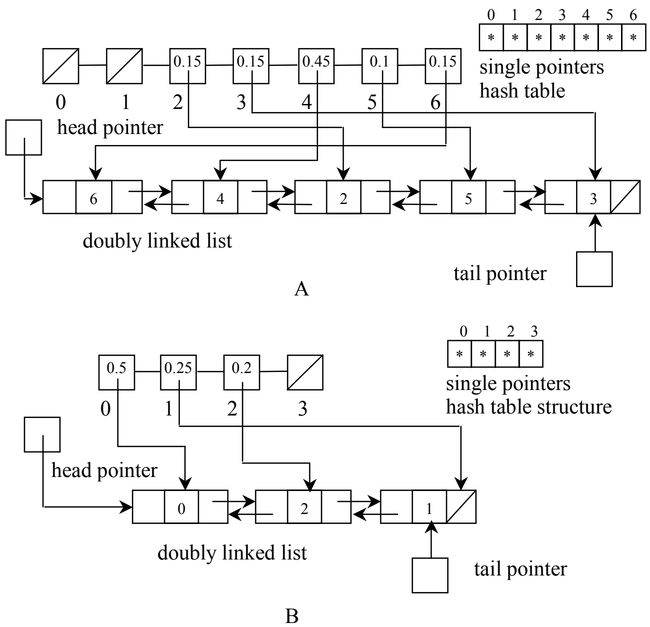

Fig. 3. GLCHH algorithm to determine image texture features. Nodes created in the linked lists and hash tables are based on sample image in Fig. 1A
A. GLCHSH structure for determining image texture features.
B. GLCHDH structure for determining image texture features.

TABLE I
PERCENTAGE RATIOS BASED ON RESULTS FOR GLCIA VS. GLCHS

Comparison [%]	Grey Levels (G)	Window Size (n x n) 5 x 5	10 x 10	15 x 15	20 x 20	25 x 25	30 x 30
i) Percentage of computational time for GLCIA compared to GLCHS (8 statistics)	**128**	53.4	37.9	32.9	29.3	28.8	27.0
	64	50.5	36.1	31.4	29.8	28.2	27.0
	32	48.8	34.7	32.8	30.8	30.2	29.8
	16	47.2	37.0	36.0	36.3	35.4	36.6
	8	53.6	45.9	47.1	48.4	51.0	53.9
ii) Percentage of computational time for GLCIA compared to GLCHS (5 statistics)	**128**	41.4	19.8	12.2	9.0	6.3	5.1
	64	34.2	15.4	9.5	6.9	5.3	4.6
	32	32.2	14.4	10.1	8.6	7.9	8.2
	16	32.5	18.3	16.8	16.3	17.6	18.1
	8	39.3	29.8	31.6	32.9	34.6	38.4
iii) Percentage of computational time for GLCIA compared to GLCHS (4 statistics)	**128**	36.9	16.6	9.7	7.3	5.1	4.1
	64	34.1	13.6	8.3	6.3	5.0	4.6
	32	30.1	13.1	9.7	8.3	8.0	8.2
	16	30.6	18.4	16.3	16.9	17.5	18.2
	8	35.5	29.2	31.0	30.6	32.6	34.6

B. Comparing GLCIA and GLCHS Completion Times

Timed testing is performed for the GLCIA using three scenarios:

i) using all *eight* indicated statistics which requires both the GLCHS and the GLCHH;
ii) using the *five* statistics that require the GLCHH data structure ie. both the GLCHDH and GLCHSH; and
iii) using the *four* statistics that require only the GLCHDH data structure.

Table 1 indicates the percentage ratio of the GLCHS versus the GLCIA completion times for each test. For all cases, GLCIA shows a demonstrated improvement over GLCHS. In the case of calculating all eight statistics, the relative performance ranges from 27.0% to 53.9%. For the cases of four and five statistics, the relative performance ranges are even lower (4.1% to 36.9% and 5.1% to 41.4%), as a result of not having to use the hybrid structure. If one wanted to determine texture features using only the statistic CON, n = 25, and G = 64, then approximately 5% of the GLCHS computing time would be required. If using a GLCM for the same purpose, this would amount to 9.1% * 5% ~ 0.46% of the time required (using Table 4 in [3]).

V. CONCLUSIONS

This paper describes a rapid means of calculating co-occurrence probability texture features. One reason that the co-occurrence approach has not been suitable for operational use was due to the exceptional computation required. By applying the GLCIA, completion times are dramatically reduced compared to traditional and other recent methods to perform the same task. The GLCIA is a highly recommended technique for anyone wishing to calculate co-occurrence probability texture features, especially from large-scale remote sensing images. Generally, the total computation time is reduced from hours to minutes for such imagery.

ACKNOWLEDGEMENTS

Support for this project was provided by Geomatics for Informed Decisions (GEOIDE), a Network of Centres of Excellence (NCE) supported by the Canadian funding agency Natural Sciences and Engineering Research Council (NSERC) (http://www.geoide.ulaval.ca) as well as CRYSYS (CRYospheric SYStem in Canada) (http://www.crysys.uwaterloo.ca).

REFERENCES

[1] R.M. Haralick, K. Shanmugam, and I. Dinstein, “Texture features for image classification,” *IEEE Trans. on Syst. Man Cybern*., vol. 3, pp. 610-621, Nov. 1973.
[2] M. Unser, “Sum and difference histograms for texture classification,” *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 8, no. 1, pp. 118-125, Jan. 1986.
[3] D.A. Clausi and M.E. Jernigan, “A fast method to determine co-occurrence texture features,” *IEEE Trans. Geosci. Remote Sensing,* vol. 36, pp. 298-300, Feb. 1998.
[4] A.E. Svolos and A.Todd-Pokropek, “Time and space results of dynamic texture feature extraction in MR and CT image analysis,” *IEEE Trans. Information Technology in Biomedicine,* vol. 2, no. 2, June 1998.
[5] D.A. Clausi and Y. Zhao, “Rapid co-occurrence texture feature extraction using a hybrid data structure,” *Computers and Geosciences*, in press.
[6] D.A. Clausi and Y. Zhao, “Grey level co-occurrence integrated algorithm (GLCIA): A superior computational method to determine co-occurrence probability texture features”, unpublished.
[7] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, New York: Dover, 1966.