Fifth International Symposium on Signal Processing and its Applications, ISSPA '99, Brisbane, Australia, 22-25 August, 1999
Organised by the Signal Processing Research Centre, QUT, Brisbane, Australia

PERCEPTUAL GROUPING OF NATURAL IMAGES FOR CBIR

Aster Wardhani and Ruben Gonzalez

School of Information Technology, Griffith University
PMB 50 Gold Coast Mail Center, QLD 4217, Australia
{a.wardhani, r.gonzalez}@gu.edu.au

ABSTRACT

In current developments of CBIR system, it is desirable to design a computer program that can index the content of an image automatically into a set of perceptually significant components. Results from existing image segmentation techniques are not sufficient to represent the content of an image. Further grouping is required to produce more meaningful segmentation.

This paper describes our approach to implement Gestalt principles for region grouping. Results from image segmentation are further grouped into regions representing major components of image content.

1. INTRODUCTION

The effectiveness of CBIR systems depends on the mechanisms used to describe images. Currently, the most prevalent mechanism is to describe the images content by manual annotation and using low level statistical features. Providing manual annotation is not only tedious but also limited by the expressive power of the annotation language which is most cases [4] in a form of keyword index. Low level features such as colour and texture, can describe the content of an image, unfortunately, it is often difficult to translate an image retrieval requirement into a statistical distribution of low level features. Additionally, the low level features provide descriptions of image content which do not necessarily provide a meaningful representation of an object.

Identifying objects automatically and providing any high level description from an image is still a difficult problem. Image contents can be very complex. While objects in an image have various shapes, sizes, orientation and shadings, image data is still represented in a low level and unstructured representation. Image segmentation techniques alone cannot produce satisfactory results. Current research in computer vision has not yet providing a general solution to the problem.

While extracting image content using object recognition is still infeasible for arbitrary images, we can however extract various segments contained in the image. Using these segments, in [5], we proposed a system that performs further grouping of image segments into main components of the image. This system is shown in Figure 1.

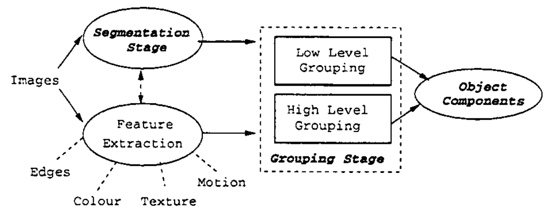


Figure 1: Image Content Analysis System

2. IMAGE CONTENT ANALYSIS SYSTEM

To form perceptually significant parts in an image, the grouping system needs to follow human perceptual grouping processes. The questions involved are: *(i) how do we determine which elements belong to the same object? (ii) What is the measure of similarity?*.

Gestalt laws state that elements are grouped based on the principles of: *proximity*, *similarity*, *good continuation*, *closure*, *common fate*, *surroundedness*, *relative size* and *symmetry* [6], hence provide an explanation to the questions above. Using Gestalt laws, we can assume that *segments will be grouped as the same region when they satisfy Gestalt principles*. This is the main assumption of our region grouping system, shown in Figure 2.

Initially an image is segmented using region growing algorithm directed by object boundaries. The results are then further grouped using five Gestalt grouping procedures shown.

3. IMAGE SEGMENTATION STAGE

3.1. Region Growing Procedure

Image segmentation process is the first step in our image information extraction system. For better results, luminance

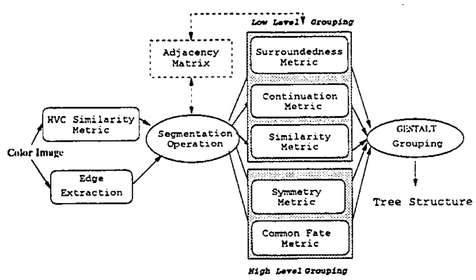
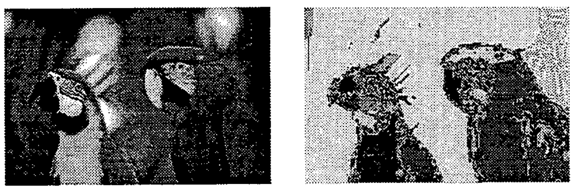


Figure 2: Perceptual Grouping System



(a) Original (b) Results

Figure 3: Region Growing Results

component alone may not be sufficient. Colour information provides more information. Hence, the HVC colour space is chosen for its correspondence to human colour perception. To implement region growing algorithm for HVC colour space, as follows.

FOR all image pixels $y(i, j)$ LOOP
  create new region = R
  WHILE pixels satisfy $P(y(i, j), R)$ THEN
    ADD pixels to R
    get next pixel
  END WHILE
END LOOP

A HVC similarity predicate required, as follows.

$$P(y(i,j),R) = \begin{cases} \text{true} & \text{if } dist < \tau \\ \text{false} & \text{otherwise} \end{cases} \quad (1)$$

$$\begin{aligned} dist &= \sqrt{(\Delta v)^2 + (\Delta(ch)_1)^2 + (\Delta(ch)_2)^2} \\ \Delta(ch)_1 &= c_i \cos h_i - c_j \cos h_j \\ \Delta(ch)_2 &= c_i \sin h_i - c_j \sin h_j \end{aligned} \quad (2)$$

Where $dist$ is a Euclidean distance of HVC space, Δv is the difference of luminance values between values at i and j, c_i and h_i are the saturation and hue values at the previous pixel location i, and j is the next location. τ is the threshold, described next. The segmentation results using the HVC metric and edge information is shown in Figure 3.

3.2. Automatic Threshold

The success of a region growing technique is often linked to the selection of an appropriate threshold. Currently we provide a fixed global threshold for an entire image. Unfortunately for different images, the required threshold varies, depending on the image content. To calculate the threshold automatically, we investigate the properties of image content which includes the textural properties and the variation of pixel changes in the image. The particular threshold used here is derived as a linear combination of the mean and standard deviation of the source image [2]. We propose to apply this formulation, but applied on gradient images to estimate the suitable step change required for the region growing procedure. For a given difference image $\mathbf{d}$ obtained by applying a gradient operator, $\mathbf{d} \in \mathbb{R}^x$, where x is an $m \times n$ grid, the mean, μ, and standard deviation, σ, of $\mathbf{d}$ are given by

$$\mu = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{d}(i,j) \quad (3)$$

$$\sigma^2 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{d}(i,j) - \mu)^2 \quad (4)$$

The threshold level τ is set at

$$\tau = k_0 + k_1\mu + k_2\sigma \quad (5)$$

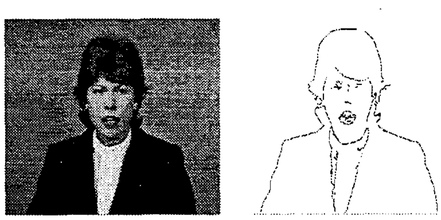
Where the constants k_0, k_1 and k_2 are obtained experimentally using Least Square optimisation.

3.3. Combining Edge Information

Lines are known to be important psychologically and can capture the high level meaning of the image content. To provide a more accurate results, the growing process needs to be controlled from leaking out to different regions. To prevent this, edge information is used. The values of the edge pixels at the original HVC image are controlled, so that the difference of values at the boundaries are always maximum.

Edge information is also used later for good continuation grouping and some requirements are imposed. The lines should be a connected pixels with 1 pixel width, unnecessary edges should be deleted and the points should be ordered with possible gaps filled. To achieve this, edge detection is performed using Robert operator [1], followed by 3x3 edge thinning algorithm [3]. Edge linking is performed by fitting second order polynomials to every edge ends in the image. The procedure is as follows.

FOR all edge ends $e(x, y)$ LOOP
  obtain an edge segment E_n of length n
  WHILE mean square error $< \tau$ THEN
    solve A and B using Eq.(7)



(a) Original (b) Results

Figure 4: Results from Edge Extraction

$n = n + 1$
END WHILE
extend $e(x, y)$ until it intersects
END LOOP

The equation for the edges (x_t, y_t) are described in Eq.(6) and to obtain the fitting coefficients A and B, Eq.(7) and (8) are used.

$$x_t = x_0 + at; \quad y_t = y_0 + bt \tag{6}$$

$$T * A = X; \quad T * B = Y \tag{7}$$

Matrices X and Y contain the true edge coordinates for an edge segment of length n.

$$T = \begin{bmatrix} 1\,1\,1\,..\,n^0 \\ 1\,2\,3\,..\,n^1 \\ 1\,4\,9\,..\,n^2 \end{bmatrix} \quad A, B = \begin{bmatrix} a_0, b_0 \\ a_1, b_1 \\ a_2, b_2 \end{bmatrix} \tag{8}$$

After the extension procedure, the lines need to be pruned, so that all the gaps are filled. The edge linking results is shown in Figure 4.

3.4. Comparison of Segmentation Results

We compare results from *region growing*, *quad-tree split and merge*, *histogram based*, *edge-based region growing* and *watershed* techniques. Table 1 shows the comparison of the number of regions produced compared with the reference sets. These sets are the average number of regions produced by 10 viewers manually segmenting the image tests.

We then calculate a confidence level as a measure of closeness between results from automatic and manual segmentation techniques. We define a measure in Eq.(9).

$$\text{confidence} = \frac{|\bar{N_M} - \bar{N_A}|}{\bar{N_M}} \tag{9}$$

where $\bar{N_M}$ is the average number of regions recorded for all responses from manually segmented each image, and $\bar{N_A}$ is the number of regions produced by each automatic segmentation results. From this confidence measure, we

Image	A	B	C	D	E	F
Split	43	92	81	141	84	87
SR-71	10	12	37	58	50	103
Pepper	30	77	68	119	82	163
Fish	14	68	52	82	56	71
House	24	133	68	107	48	230
Colours	10	36	7	52	18	13
Bridge	16	91	66	78	31	103
Balls	13	11	25	5	27	28
Araras	13	43	37	64	35	57
Airplane	13	90	63	87	55	140
Claire	11	18	11	4	7	25
Flowers	17	70	38	110	20	150
Suzie	13	56	48	23	104	83

Table 1: Comparison: A = Manual Test, B = Split & Merge, C = Region Grow, D = Hist. Split, E = Edge based Region Grow, F = Watershed

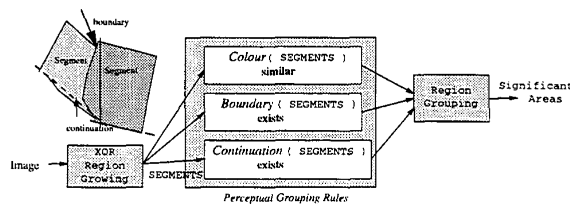


Figure 5: Good Continuation Grouping

obtained the following confidence levels for each segmentation technique as shown in Table 2. Table 2 shows that results from region growing technique are the closest to the reference sets with a confidence level of 1.8.

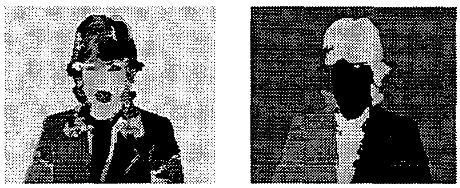
4. REGION GROUPING

Prior to Gestalt grouping operation, smaller segments are grouped based on their size. The same basic textured elements commonly have similarity in size. An example of results from size grouping is shown in Figure 6(a). Segments are then grouped by good continuation. This is illustrated in Figure 5.

Good continuation is implemented to group segments that share the same continuous line together[5]. If the line is crossing between the regions, they should not be merged. To allow the two operation above, we need to be able to compute where the relative position of a region is with respect to a line. In other words, we need to compute a polarity of region with respect to a line. To provide this computation, we use accumulative cross product of points at a region with points at an edge. The region grouping procedure using polarity is described as follows.

Image	**A**	**B**	**C**	**D**	**E**
Split	1.157	0.899	2.307	0.970	1.040
SR-71	0.233	2.803	4.962	4.140	9.588
Pepper	1.543	1.246	2.930	1.708	4.384
Fish	3.704	2.597	4.672	2.874	3.911
House	4.541	1.833	3.458	1.000	8.583
Colours	2.882	0.245	4.607	0.941	0.401
Bridge	4.687	3.125	3.875	0.937	5.437
Balls	0.141	0.950	0.609	1.106	1.184
Araras	2.307	1.846	3.923	1.692	3.384
Airplane	6.021	3.914	5.787	3.290	9.921
Claire	0.596	0.024	0.645	1.217	3.967
Flowers	3.207	1.284	5.612	0.202	8.016
Suzie	3.307	2.692	0.769	7.000	5.384
Average	2.641	**1.801**	3.397	2.083	5.016

Table 2: Confidence Level:A = Split / Merge, B = Region Grow, C = Hist. Split, D = Edge Based Region Grow, E = Watershed



(A) Size Grouping (b) Good Continuation

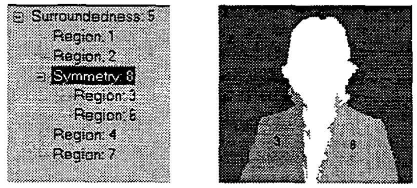
Figure 6: Grouping Results

FOR all pixels $y(i,j)$ in an edge E LOOP
obtain an edge direction vector $\Delta G(i,j)$
obtain a point in a region, $R(i,j)$
$CP+ = CP(\Delta G(i,j), R(i,j))$
END LOOP

Results from good continuation grouping is shown in Figure 6(b).

To obtain the surroundedness for every region, we scan each region outwards in many directions. In each direction, we check for the same region index found in scanning. The first common region found is placed at the higher node of the current region node in the tree, describing the surroundedness property.

For symmetry, initially, we assume that the region candidate pair are symmetrical to each other. This will allow us to estimate the axis of symmetry. The axis is defined as a line that is perpendicular to the line that crosses the center points of the two regions. Then for each region, we transform its boundary coordinate into polar coordinates and quantise the boundary by some angle magnitude. Then the reflectivity of each sample points from both regions are inspected, by comparing the distance from the points to the line axis. For all sample points the total error is computed and used as an estimate for the grouping decision. Common fate grouping is performed by merging segments that have similar motion vectors. Currently, this is still in experimentation stage. The grouping results from surroundedness, symmetry and common fate are stored as a tree structure, showing all the grouping decisions made. A structure obtained is shown in Figure 7.



(a) Tree Structure (b) Symmetry Displayed

Figure 7: Tree Structure from Gestalt Grouping

5. CONCLUSION AND FUTURE WORK

We have presented our current work in image structure content analysis. We implement Gestalt laws in attempt to integrate various sources of information that an image can have (*The whole is greater than the sum of its parts*). Some grouping results were presented and shown to be better than the results from initial segmentation. We showed a possible image content structure that can be derived using Gestalt laws.

6. REFERENCES

[1] R. C. Gonzalez and Woods R. E. Digital image processing. 1993.

[2] N. Hamadani. Automatic target cueing in ir imagery. December 1981. WPAFB.

[3] L. O'Gorman. kxk thinning. *Computer Vision, Graphics, and Image Processing*, 51:195–215, 1990.

[4] Rohini K. Shihari. Use of captions and other collateral text in understanding photos. *Artificial Intelligence Review, Special Issue on Integrating Language and Vision*, 8:409–430, 1994-95.

[5] A. W. Wardhani and R. Gonzalez. Automatic image structure analysis. *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, pages 180–188, 1998.

[6] M. Wertheimer. Laws of organization in perceptual forms. *Psychologische Forschung*, 1923. 6.