USING HIGH LEVEL INFORMATION FOR REGION GROUPING

A.W Wardhani and R Gonzalez
School of Information Technology
Gold Coast Campus Griffith University
e-mail: easaster@davros.eas.gu.edu.au
R.Gonzalez@eas.gu.edu.au

ABSTRACT

Effective labeling for an image indexing system requires all objects in the image to be identified. This identification process can be performed by extracting components of the objects and grouping these components together. We propose the use of image segmentation techniques as a first step to solve the problem of extracting these components automatically. The difficult task is how the grouping of these components is performed.

This paper presents an approach in region grouping using high level information. This information permits image segments grouped into "more meaningful" regions. In this paper, we present the issues and problems involved in region grouping. Some experiment results will be presented.

1. INTRODUCTION

The MPEG committee has recently approved a new work item entitled "Multimedia Content Description Interface", named MPEG-7. This work aims at providing support for labeling media to support content based retrieval of the multimedia data. However, before supporting an interface to describe content, there needs to be some method of identifying objects in the media which are to be labeled and described. This paper reports on a recent experiment in image region grouping using high level information which attempts to automatically identify these objects.

Traditionally for image and video data, this identification problem has been approached by two main techniques: manual annotation [1] and object recognition [2]. Apart from being tedious, providing textual annotation to every image and video segment is limited to annotator's choice of vocabulary and context. This limits the scope of retrieval. Object recognition techniques also suffer from drawbacks such as computational expense and its restriction to simple predefined polyhedra in highly constrained environments [3].

The current work in the area of automatic index generation for content based retrieval, can automatically generate object descriptions using statistical attributes such as colour and texture, [4] but cannot identify the objects themselves. These statistical methods are unable to support object based queries, without relying on manual or highly constrained semantic methods.

In this paper, we present an approach in region grouping in attempt to provide automatic identification schemes. This paper is organised as follows. Section 2 discusses the method to provide automatic object identification. Three issues will be

described in this section: image segmentation, feature extraction and region grouping. In section 3, the experiment procedure and results will be presented. Section 4 describes the conclusion and future works.

2. AUTOMATIC OBJECT IDENTIFICATION TECHNIQUES

As explained in section 1, existing methods of identification are not suitable for current needs. The requirement for a new approach in object identification is the ability to provide a technique that allows unsupervised (automatic) identification but does not require prior knowledge. We propose a scheme illustrated as follows.

Fig. 1. The Proposed Object Identification Technique

First, components of the object are obtained from the image. Image segmentation techniques can solve the problem of extracting these components automatically. Then, these components are grouped together based on some grouping rules. Some features from the image need to be extracted to measure the similarity or "togetherness" between the components. This is illustrated in Figure 2. Initially, the image is segmented to obtain segments 1, 2 and 3. Then, using the region grouping procedure, these segments are combined producing "An Apple"¹.

Fig. 2. Object Formation Using Image Segments

To summarise the discussion, the proposed automatic object identification scheme can be divided into three topics: image segmentation, feature extraction and region grouping.

¹ Note that the description of "An Apple" is not interpreted by the program. This is used only to describe the illustration.

2.1. Image Segmentation

Image segments can be defined as group of pixels which are statistically correlated. These segments can be obtained by using image segmentation techniques. These techniques can be broadly categorized into four main groups: region based segmentation, color quantisation, texture based segmentation and motion based segmentation.

Region based segmentation produces segments from grouping some pixels according to some constraints such as spatial proximity and intensity similarity [3]. For regions which do not have smooth homogenous luminance distribution, texture-based segmentation is used. This technique groups pixel which has consistent texture properties. For video data, the pixels in each frame can also be grouped based on the similarity in their motion characteristics, using motion-based segmentation techniques. Various reviews for all these techniques are available, such as in [3] and [5].

For this experiment, the colour quantisation technique using JND (Just Noticeable Difference) threshold [6] is implemented. This technique reduces the number of colours or intensity values present in an image. Using this method, the pixels with a small difference in colour can be grouped as one segment under one colour representation. This process is illustrated in Figure 3, and explained as follows.

Fig. 3. Quantisation Using JND Threshold

Before the quantisation process is taken place, all values are transformed to Gamma function domain, which is described as follows:

$$F(y) = y^k \quad (1)$$

Where y is luminance value and k ranges from 2.2 - 2.5. In this experiment, we select the value $k = 2.3$.

This domain is used to model what we “visually” perceived. The values in the new domain are quantized by 5 bits. Values that fall below this threshold are visually indistinguishable. After each value is assigned according to this threshold, it is transformed back to original domain, using the following equation.

$$y = F(y)^{1/k} \quad (2)$$

The segments resulted using this method is shown in Figure 9(b).

2.2. Feature Extraction

After the image is segmented, the resulted regions need to be grouped to form descriptions of the object in the image. To perform region grouping, some information is required to be able to measure the similarity and “togetherness” between these regions. This information is extracted from each segment in the form of various type of features. In most existing methods these have been statistical (low level) features, such as average

luminance, luminance variance, colour and texture. These features are compared and the grouping decision is made based on how close the values between features of each region are. An example of this operation can be seen in [7].

To extract these low level features does not require an intensive computation. The disadvantage of using these features is the grouping will not follow the underlying object structure, therefore being semantically inconsistent. Hence, we need to use higher level information that can reveal the semantic or structural relations between segments.

In this paper, we investigate a region grouping method that uses not only low level information, but also high level information. The problem now lies on which features possess high level properties that can be extracted without any intensive computations and can be done automatically.

As illustrated in Figure 4 below, line information in an image possesses these high level properties. Lines can describe the structure of an object, hence describing what the content of the image is. Also lines are relatively simple to extract from the image.

Fig. 4. Lines as high level information

Therefore, lines are used in our region grouping experiment as high level information. Using lines, we can make suggestions as whether two segments can be grouped together. This will be described in the next section. To extract lines from the image, the following extraction process is performed.

Fig. 5. Line Extraction Process

Before the edge detection process, the image is sub-sampled by 2, to reduce some unnecessary line details. The image edges are detected using Sobel operators [8]. The image is then interpolated back to its original size. The resulted edges are thinned using 3x3 thinning algorithm [9] and linked. The edge direction is used to determine which edges should be linked. This method is illustrated in Figure 6.

Fig. 6. Edge Linking and Pruning

The linking process is described as follows. The edges are initially extended by 5 pixels according to the direction of the edges. When these extended edges meet other edges, the two edges are linked. The extended lines that are not connected to any edges are deleted. Finally in the last step, edges that have length less than some threshold, are "pruned". In this case, only lines that are longer than 20 pixels are considered to be useful and will be used for in grouping. The result of this line extraction process is shown in Figure 9(c).

2.3. Region Grouping

2.3.1. Related Work

While image segmentation techniques are available quite extensively, this is not the case for region grouping schemes. Work related to region grouping mostly concentrate on block images, such as the work performed by Guzman in [10]. This technique performs grouping on simple and uniform block objects. Objects are formed by studying the characteristics of the line junction information.

For natural images, the research in region grouping is very limited. The grouping results have been limited only to produce a better segmentation. These results are still low level and preliminary. Nazif and Levine in [7] incorporate a rule-based system to segment an image into uniform regions and connected lines. Feldman and Yakimovsky in [11] use decision theory and semantic information to produce image segments. So far, there has been no automatic techniques that can successfully group segments into an object.

2.3.2. Proposed Region Grouping Schemes

The difficulty in performing region grouping is centred in the following questions: how do we determine which segments belong to the same object? In other words, what is the measure of similarity of segments as one object?

To answer these questions, there are some issues involved. The measure of similarity depends on two factors. One factor is the selection of features used in grouping. As discussed in Section 2.2, low level features are mostly used in existing region grouping schemes. These features alone are insufficient to provide good grouping results. Lines as high level features were selected to solve this problem. The second factor that needs consideration are the rules used in the grouping procedure. To generate robust grouping rules, we are investigating the area of human perception.

The definition of an object and the elements of the object depends on the way humans perceive objects. This is explained

by the theory of perceptual organisation, known as Gestalt theory. This theory reflects the general properties of world objects and explains how certain organisations are more likely to occur. The grouping of elements into an object is explained as having these properties between elements: *proximity, similarity, common fate, closure, good continuation, symmetry, surroundedness and relative size* [12]. When the relations between segments satisfy these properties, according to Gestalt theory, the segments are grouped to the same object. Considering this information, Gestalt theory is used as the basis for our proposed region grouping schemes. The difficulty in using this theory would be, how to translate this theory into region grouping rules. Some rule examples are described in [13].

In this experiment, we attempt to implement Gestalt theory in grouping rules. These rules accept two forms of information from the image: line and region information, as shown in Figure 7.

Fig. 7. Region Grouping Implementation

The lines are extracted using the method described in section 2.2 and the regions are obtained using the segmentation method described in section 2.1. The grouping rules generated for this experiment are described as follows.

- (1) IF intensity of region A and B are similar THEN
Group(A,B)
- (2) IF region A and B share a common continuous line THEN
Group(A,B)
- (3) IF region A and B is separated by a line THEN
Do not group(A,B)

Rule (1) is based on Gestalt theory of similarity and Fisher's test [15] is used as the measure of similarity. This test is defined as follows.

$$F = \frac{(n_1 + n_2)(\hat{\mu}_1 - \hat{\mu}_2)^2}{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2} \quad (3)$$

Rule (2) and (3) are based on Gestalt good continuity principle

Fig. 8. Using line information in grouping and are illustrated in Figure 8.

Using these grouping rules, the following grouping procedure is implemented:

```

Get Image Segments;
Get Image Line information;
FOR all segments in the image DO:
  Get Next Segment Pair (A,B);
  Measure the statistical properties of region A, B;
  IF Similarity Test < Threshold THEN
    IF region A and B is not separated by a line THEN
      Group(A,B);
    ELSE
      Get Next Segment Pair (A,B);
    ELSE
      IF region A and B is not separated by a line THEN
        IF region A and B share a common cont. line THEN
          Group(A,B);
        ELSE
          Get Next Segment Pair (A,B);
        ELSE
          Get Next Segment Pair (A,B);
      END
    END
  END
END procedure

```

After the image is segmented, component labeling algorithm [14] is applied to label the regions. This label information is stored in an array called region map. The line information is stored in a separate array called edge map. When two regions are compared, the region map array is inspected to measure the intensity distribution of the regions. The appearance of lines around the regions is checked by inspecting the edge map information.

Using this region grouping procedure, an experiment is conducted and the results are described in the next section.

3. EXPERIMENT RESULTS

The following are the experiment results in region grouping schemes proposed in this paper. The luminance image "Claire" is used in this experiment, as shown in Figure 9(a). This image is segmented and the segmentation results are shown in the image in Figure 9(b). An image in figure 9(c) describes the line extracted from the original image.

Using the above similarity measures and the edge information, the grouping is performed. The grouping results are shown in Figure 9(d). The number of segments produced are reduced from around 300 to 4 main object components. The threshold used in similarity test is currently set to 4.

By looking at the reduction of the number of regions, we can see that the grouping reduces the segmented image into more meaningful areas. For labeling and indexing purpose, these areas can provide better query results and definition of the content of the image.

Fig. 9(a) Original Image

Fig. 9(b) Segmented Image

Fig. 9(c) Line Information

Fig. 9(d) Grouping Result

5. FUTURE WORK AND CONCLUSION

In this paper, the study of issues involved in providing automatic object identification is presented.

In the future, we will implement region grouping using other Gestalt principles. The aim is to combine the results from this experiment with other Gestalt rule implementations to produce an object description in an image.

REFERENCE

- [1] 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
- [2] W.L. Grosky, R. Mehrotra, "Index based Object Recognition in Pictorial Data Management," *Computer Vision, Graphics and Image Processing*, Vol.52, No.3, 416-436, 1990
- [3] R. Gonzalez and A.W. Wardhani, "Resource Discovery and Content Based Retrieval of Visual Data," *Proc. ATNAC*, Vol.2, 369-374, 1996.
- [4] M. Flicker, et. Al., "Query by Image and Video Content: The QBIC system," *IEEE Computer*, 23-31, 1995
- [5] D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice-Hall, New Jersey, 1982.
- [6] G. Murch, N. Weiman, "Assessing Visual Grey Scale Sensitivity On A CRT," *Proc. of the SPIE*, Vol. 1249, Human Vision and Electronic Imaging: Models, Methods and App. 1990, pp.214-223.
- [7] A. M. Nazif, M. D. Levine, "Low Level Image Segmentation," *IEEE PAMI*, Vol. PAMI-6 No.5, 555-577, 1984.
- [8] Hart and Duda, *Pattern Classification and Scene Analysis*, Wiley-Interscience Publication, 1973.
- [9] L. O'Gorman, "k x k Thinning," *Computer Vision, Graphics and Image Processing*, 51, 195-215, 1990.
- [10] A. Guzman, "Decomposition of a Visual Scene into three dimensional bodies," *Proc. FJCC* 33, 291-304, 1968.
- [11] J.A. Feldman, Y. Yakimovsky, "Decision Theory and Artificial Intelligence: I.A Semantics-Based Region Analyser," *Artificial Intelligence*, 5, 349-371, 1974
- [12] V. Bruce and P.R. Green, *Visual Perception: Physiology, Psychology and Ecology*, 2nd Ed., Lawrence Erlbaum Assoc., Hove and London, 1990
- [13] A.W. Wardhani, R. Gonzalez, "Automatic Object Extraction for Content Based Indexing," *Proc. ICT '97*, Vol.2, 1005 - 1010, 1997.
- [14] A.F. Lochovsky, "Algorithms for Realtime Component Labelling of Images," *Image and Vision Computing*, Vol.6, No.1, 21-27, 1988.
- [15] T.R. Reed, "Segmentation-based image processing," *Proc. Int'l conf. Digital Signal Processing*, Vol.1, pages 308-313, 1995.