COLOR SEGMENTATION AND FIGURE-GROUND SEGREGATION OF NATURAL IMAGES

Swee-Seong Wong and Wee Kheng Leow

School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543
{wongss,leowwk}@comp.nus.edu.sg

ABSTRACT

To recognize the objects in an image and to understand the image content, a computer system has to first separate the foreground objects from the background. Image segmentation and figure-ground segregation are, therefore, essential for computer image understanding. This paper describes a system called OLAG (Object-LAyer Grouping) for image segmentation and figure-ground segregation. OLAG consists of several incremental refinement steps which use colour and other visual cues such as size and compactness for grouping the image pixels. It produces, as an end result, a set of layers each containing an object or object part. Figure and ground relationships among the objects are inferred, giving their relative depths. It is shown that interesting and useful segmentation results can be obtained from the system.

1. INTRODUCTION

To recognize the objects in an image and to understand the image content, a computer system has to first separate the foreground objects from the background. Image segmentation and figure-ground segregation are, therefore, essential for computer image understanding. This paper describes a system called OLAG (Object-LAyer Grouping) for image segmentation and figure-ground segregation. In the present implementation, OLAG focuses on using colour to segment an input image into regions. Next, it uses colour as well as other visual features, such as size and compactness, to group regions that are likely to belong to the same object into a cluster of regions. Finally, foreground and background relationships between the clusters are inferred, giving the relative depths of the objects. OLAG also provides a framework in which other visual features such as texture and shape can be incorporated to improve the segmentation results.

Recent works on colour segmentation include a variety of techniques such as seed growing with global optimization [1], region growing with colour-texture-based criteria [2] and clustering [3]. On the other hand, figure-ground segregation has not been adequately addressed in computer vision research. Although there are theories about human perception, and neural network and mathematical models of figure-ground segregation [4, 5], no actual system has been applied to the segmentation of natural images. In contrast, OLAG integrates image segmentation and figure-ground segregation into a single system.

OLAG focuses on finding objects that are visually obvious and distinct from the background and neighbouring objects. These objects should have significant sizes compared to the image and have visual properties such as colour and texture that differ significantly from the background and neighbouring objects. An object may consists of several disjoint parts and may be occluded by other objects. It is necessary to group the disjoint parts of an object together and identify them as parts of a single object.

2. THE OLAG SYSTEM

OLAG follows a conservative approach in segmenting an image into regions. Pixels are grouped into regions, and regions into objects only when there is enough confidence that the grouping results are correct. In the initial steps, simple but strict colour similarity criteria are used to quickly group pixels into regions. Due to the strictness of the criteria, many regions are produced (i.e., over-segmentation) but they all have homogeneous colours. In subsequent steps, more complex criteria are used to further group the regions together. With the use of complex grouping criteria, the colour similarity criteria can be relaxed to encourage further merging of regions. This incremental refinement method avoids the need to split regions that are already merged and improves efficiency because complex grouping criteria are applied only to large regions instead of all the pixels.

OLAG consists of four stages, namely *pre-processing*, *colour-based grouping*, *non-colour-based grouping* and *figure-ground inferencing* (Fig. 1). Each of these stages is further divided into several steps.

This research is supported by NUS Research Grant RP3999903.

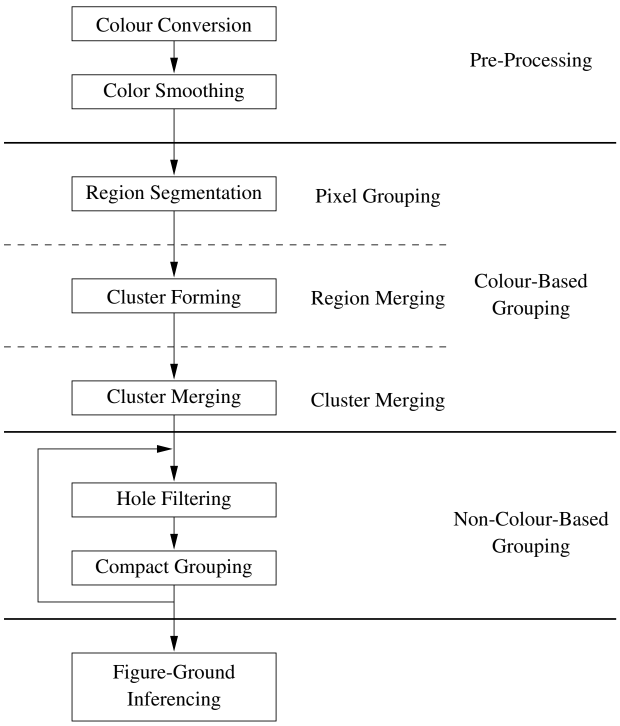


Fig. 1. The OLAG system consists of progressive refinement steps.

The **pre-processing** stage prepares the underlying image for subsequent grouping by transforming the image into a suitable colour domain and removing noise. It first applies the *colour conversion* step to transform image pixels from the RGB colour space to the CIE L*u*v* space [6], which has the advantage that Euclidean colour distance measured in the space is quite consistent with human's perception.

In the second step, *colour smoothing* is performed using non-linear edge preserving filter to remove noise in the image while preserving object boundaries. OLAG uses a hybrid filter which follows the symmetric nearest neighbour filter's pixel selection strategy [7] but the sum of the chosen pixels are weighted according to the colour distance between the chosen pixels and the center pixel. The closer the colour, the higher is the assigned weight. Smoothing is performed iteratively and it terminates when the change in the image's smoothness is small (< 0.05). The smoothness Θ is measured as the mean colour distance D_{ij} between adjacent pairs of pixels P_i and P_j in the image:

$$\Theta = \frac{1}{N} \sum_{i,j} D_{ij} \tag{1}$$

where N is the number of adjacent pairs in the image.

In the second stage, **colour-based grouping** methods merge pixels and regions together based on their colour similarity. The colour-based grouping stage consists of three steps. First, the *region segmentation* step merges adjacent pixels P_i and P_j into the same region if the colour distance D_{ij} between the pixels is smaller than the *pixel grouping threshold* $D(\Gamma_P)$, which is defined as follows. The parameter Γ_P is a percentage of the number of pixels within the neighbourhood of P_i and P_j. The threshold $D(\Gamma_P)$ is a colour distance, say D', such that Γ_P amount of adjacent pairs of pixels in the neighbourhood have colour distances smaller than or equal to D'. Therefore, the threshold varies according to the local colour distribution of the pixels.

In a typical image, the total number of pixels within all the regions should be much larger than the total number of pixels along the regions' boundaries. The colour distance between adjacent pixels in a region should be smaller than that across the region boundary. Therefore, most adjacent pairs of pixels should have small colour distances. This observation indicates that Γ_P should be much larger than 50%. Empirical studies show that a Γ_P value of 80% consistently produces good initial segmentation.

Next, the *cluster forming* step combines possibly disjoint regions that are likely to belong to the same object into a single cluster. Two regions R_i and R_j are grouped into the same cluster if the following criteria are satisfied:

- The colour difference D_{ij} between the two regions is smaller than or equal to the *region grouping threshold* $D(\Gamma_R)$. The colour of a region is defined as the mean colour of the pixels in the region. The threshold $D(\Gamma_R)$ is defined in a similar manner as $D(\Gamma_P)$. For $D(\Gamma_R)$, the neighbourhood of the regions is used to define the threshold.

- The two regions are close to each other. Closeness is defined relative to the size of the regions because regions that correspond to larger object parts can be located further apart than those corresponding to smaller parts. Two regions are considered close to each other if their *extended bounding boxes* overlap. A region's extended bounding box is derived by extending its minimum bounding box in all four directions by an amount equal to the width and height of the minimum bounding box.

- The neighbourhoods of the two regions have similar colours (as defined by the threshold $D(\Gamma_R)$).

The parameter Γ_R is typically set at 85%, which gives rise to a more relaxed threshold than $D(\Gamma_P)$ (with Γ_P = 80%). That is, more pixels can be grouped together because $D(\Gamma_R) > D(\Gamma_P)$. The threshold at this step can be relaxed because of the introduction of two additional grouping criteria. The cluster forming step is performed iteratively. At the end of this step, regions are grouped into clusters, each containing a set of regions that most likely belong to the same object.

The third step, *cluster merging*, further groups adjacent clusters together by relaxing the colour similarity criteria. The *cluster grouping threshold* $D(\Gamma_C)$ is slowly raised, in each iteration, from a typical Γ_C value of 85% to 95%. Two adjacent clusters C_i and C_j are merged if their colour distance D_{ij} is smaller than or equal to $D(\Gamma_C)$.

The third stage of OLAG, **non-colour-based grouping** stage merges clusters together based on non-colour visual features to improve the grouping result. It consists of two iterative steps. First, the *hole filtering* step removes small regions that are insignificant compared to its neighbouring regions. This step is in line with OLAG's emphasis of finding significantly large objects in the image. Empirical studies reveal that the significance of a region is a nonlinear function of its neighbouring regions. A region R_i is insignificant if its area S_i is smaller than or equal to $14(A_i)^{\frac{1}{4}}$ where A_i is the total size of its neighbouring regions. A hole is removed by absorbing it into the adjacent region that shares the longest boundary.

The second step, *compact grouping*, combines adjacent clusters to form more compact clusters. This step illustrates the ease of incorporating other visual features into OLAG. The compactness M_i of cluster C_i is defined as the ratio of the square of the cluster's perimeter and its area. The smaller is the value of M_i, the more compact is the cluster. A circle is the most compact shape with a compactness measure of 4π. Two adjacent clusters are merged if the compactness of the merged cluster is significantly smaller than those of the individual clusters. In the non-colour-based grouping stage, the hole filtering and compact grouping steps iterate until all the holes are removed (Fig. 1).

The final stage of OLAG, **figure-ground inferencing**, applies inference rules to determine the depth relationship among the clusters identified:

- Background: A cluster that covers the full width or length of an image is identified as a background because backgrounds typically extend over large portions of an image.
- Enclosure: Suppose that a cluster C_i is fully enclosed in another cluster C_j. Then, the smaller cluster C_i is regarded as locating in front of and occluding the enclosing cluster C_j, conforming to human perception.
- Occlusion: A cluster may contain two or more regions separated by smaller regions in another cluster. The cluster with smaller regions is regarded as locating in front of and occluding the cluster with larger regions, again conforming to human perception.

Background clusters are placed in a layer behind all other layers. An occluding cluster is placed in a layer in front of the layer that contains the occluded cluster. The clusters of regions are thus separated into different layers that represent the relative depths of the regions.

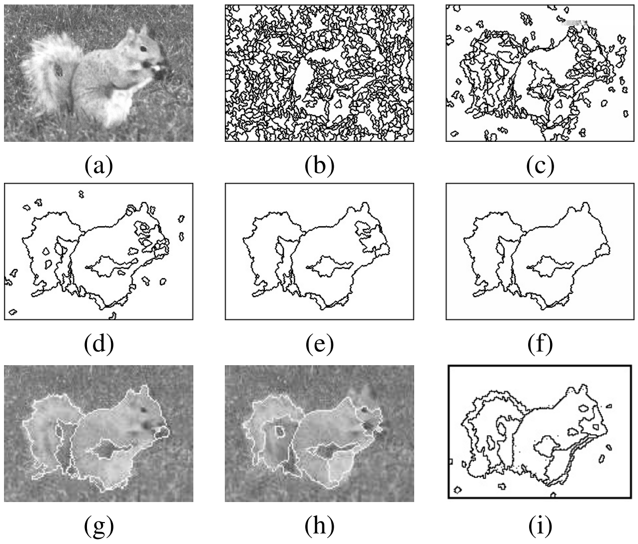
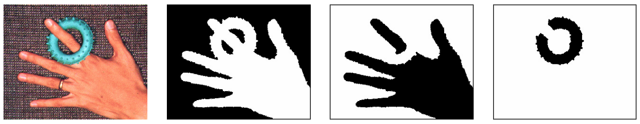


Fig. 2. Segmentation results. (a) The input image and the results of (b) region segmentation, (c) cluster forming, (d) cluster merging, (e) hole filtering, and (f) the final result after compact grouping. In comparison, OLAG's final segmentation result (f, g) is better than those of (h) JSEG and (i) the Region Competition method.

3. EXPERIMENTAL RESULTS

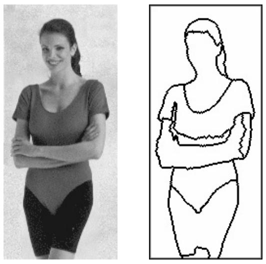
Experiments were conducted to test OLAG in segmenting natural images. Figure 2 shows the intermediate results of each processing steps of OLAG. The input image was initially over-segmented in the region segmentation step and the segmentation result was refined in successive steps. The final step produced a good segmentation of the image into major significant object parts. OLAG's segmentation result (Fig. 2f, g) compares favourably with those of JSEG [2] (Fig. 2h) and the Region Competition method [1] (Fig. 2i). In particular, OLAG can segment the foreground object into a small number of large regions without losing parts of the object or including many small regions.

Figure 3 illustrates OLAG's segmentation and figure-ground segregation results. JSEG and Region Competition produced similar segmentation results as OLAG does for this image except that they do not perform figure-ground segregation. The input image is segmented into 3 clusters corresponding to the background, the rubber ring and the hand, ordered in decreasing relative depth. The background has an extensive coverage across both the width and length of the image and thus can be easily identified. Note that the background cluster consists of 3 disjoint regions, all of which are correctly grouped into the same cluster. Similarly, the two disjoint regions of the hand cluster are correctly grouped into the same cluster. The ring cluster consists of a small region which separates the large regions of

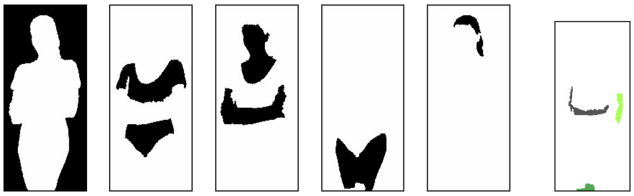


(a) (b) (c) (d)

Fig. 3. Figure-ground inferencing: (a) The input image, (b) the background and (c, d) the foreground objects.



(a) (b)



(c) (d)

Fig. 4. (a) The input image, (b) the segmented result, (c) the 5 largest clusters and (d) the remaining clusters shown in a single image.

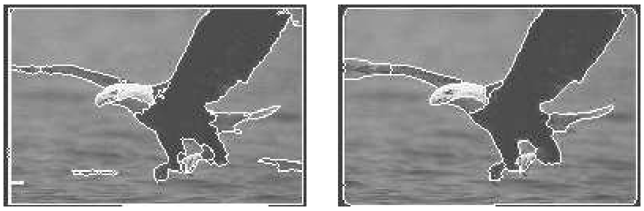
the hand cluster. Therefore, it is correctly identified as occluding the hand cluster. Actually, a part of the hand (the index finger) also occludes the ring. However, due to the layering scheme used in the current OLAG implementation, such a cyclical occlusion relationship is not identified.

Figure 4 further illustrates the grouping of disjoint regions corresponding to different parts of the same object into the same cluster. In particular, the face and the arms are identified as disjoint part of the same object (the body) whereas the light-coloured garments are identified are two disjoint parts of the same object (the garment).

Figure 5 further compares the segmentation results of OLAG and JSEG. OLAG segments the eagle's wing correctly but it produces some small regions in the sea area (Fig. 5a). JSEG's segmentation results are less fragmented but the eagle's wing tip is merged with the sea (Fig. 5b).

4. CONCLUSION

The objective of the OLAG system is to perform segmentation and figure-ground segregation of natural scene images based on colour and other visual information. It provides



(a) (b)

Fig. 5. Comparison of segmentation results. (a) OLAG's segmentation results. (b) JSEG's segmentation results.

a framework consisting of several processing steps, each refining the results produced by the previous step. OLAG can not only group disjoint regions of a potential object together but can also infer the relationships among the objects. OLAG's segmentation results can be further improved by including other visual cues such as texture, boundary continuity, symmetry of object parts, etc. into appropriate processing stages in OLAG.

5. REFERENCES

[1] S.C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing and Bayes / MDL for multiband image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. vol. 18, no. 9, pp. 884–900, 1996.

[2] Y. Deng, B.S. Manjunath, and H. Shin, "Color image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1999.

[3] T. Uchiyama and M.A. Arbib, "Color image segmentation using competitve learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. vol. 16, no. 12, pp. 1197–1206, 1994.

[4] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological Review*, vol. vol. 94, no. 2, pp. 115–147, 1987.

[5] S. Grossberg, "A solution of the figure-ground problem for biological vision," *Neural Nerworks*, vol. vol. 6, no. 4, pp. 463–483, 1993.

[6] Hunt, R. W. Gainer, *Measuring Colour*, Ellis Horwood series in applied science and industrial technology. New York: Ellis Horwood, 2nd edition, 1991.

[7] M. Pietikäinen and D. Harwood, "Segmentation of colur images using edge-preserving filters," in *Advances in Image Processing and Pattern Recognition*, V. Cappellini and R. Marconi, Eds., pp. 94–99. North-Holland, 1986.