THE USE OF VISIBLE COLOR DIFFERENCE IN THE QUANTITATIVE EVALUATION OF COLOR IMAGE SEGMENTATION

Hsin-Chia Chen and Sheng-Jyh Wang

Department of Electronics Engineering, Institute of Electronics,
National Chiao Tung University, Hsin-Chu, Taiwan
E-mail address: shinga@archer.ee.nctu.edu.tw and shengjyh@cc.nctu.edu.tw

ABSTRACT

In this paper, we propose the use of visible color difference in a new quantitative evaluation scheme for color image segmentation. With visible color difference, two measurements, named intra-region visual error and inter-region visual error, are defined to evaluate the quality of segmentation results. To fit for human's visual perception, segmentation results with an excessive amount of intra-region error are regarded as under-segmentation, while segmentation results with too many inter-region errors are regarded as over-segmentation. Based on these two measurements, a complete scheme for the evaluation of color image segmentation is proposed. The simulation results demonstrate that this new scheme may provide a reliable and efficient way to automatically select the parameter settings for a given segmentation algorithm and to compare the performance between various segmentation algorithms.

1. INTRODUCTION

Automatic image segmentation has long been thought as one of the most difficult tasks in image analysis. For automatic image segmentation, how to develop an objective and quantitative way to evaluate the performance of segmentation is a crucial subject. Although the development of segmentation algorithms has already attracted a significant amount of efforts, relatively fewer efforts have been spent on the subject of performance evaluation [1].

According to the classification made by Zhang [2, 3], most evaluation methods could be roughly classified into tree categories: 1) analytical methods; 2) discrepancy methods; and 3) goodness methods. In general, analytical methods directly evaluate the segmentation algorithms by analyzing their principles, requirements, utilities and complexity, etc. On the contrary, both discrepancy methods and goodness methods evaluate the performance of segmentation by judging the quality of segmentation results. Especially, discrepancy methods measure the difference between the segmentation result and a reference segmentation result, which is usually an expected result or a ground truth. On the other hand, goodness methods evaluate the segmentation results with certain quality measures without the use of any reference result.

Due to the lack of a general theory for image segmentation, analytical methods work well only with some particular models or for some desirable properties of the algorithms. For discrepancy methods, the reference result is essential for the evaluation of segmentation. However, the acquirement of reference results is usually non-trivial and the acquired reference results are usually user-dependent [4]. Hence, in normal circumstances, goodness methods tend to be more practical. For this type of methods, a given algorithm can be evaluated by simply computing some goodness measures over the segmentation results. So far, plenty of goodness measures have already been proposed. For example, in [5], an evaluation function based on a color difference defined in the RGB space and the total number of segmented regions is proposed to measure the difference between the original image and the segmented image. In [6], an evaluation composed of the normalized standard deviation within the segmented regions and the intensity difference between adjacent regions is proposed.

In this paper, we propose a new evaluation scheme which is basically a goodness approach. In our approach, the visible color difference is defined and used to assist the evaluation of color segmentation. With visible color difference, two measurements are designed. A so-called intra-region visual error is designed to measure the visible color difference within the segmented regions. This measure can be used to estimate the degree of under-segmentation. On the other hand, another measurement named inter-region visual error is designed to measure the invisible color difference between every adjacent pairs of segmented regions. This measure can be used to estimate the degree of over-segmentation. Based on these two measures, a complete scheme is then proposed to evaluate the performance of color segmentation algorithms.

2. VISIBLE COLOR DIFFERENCE AND ERROR MEASUREMENTS

To evaluate the quality of color segmentation, we first propose the use of "visible color difference". Among various definitions of color difference, we choose the CIE

ΔE^*_{ab} definition as the basis of color difference. This definition is defined over the CIE L*a*b* color space, which is a roughly uniform color space. In this CIE L*a*b* color space, the color difference between two colors, (L_1^*, a_1^*, b_1^*) and (L_2^*, a_2^*, b_2^*), is defined as

$$\begin{aligned}\Delta E^* ab &\equiv \left\|(L_1^*, a_1^*, b_1^*) - (L_2^*, a_2^*, b_2^*)\right\|_{L^*a^*b^*} \\ &= \sqrt{(L_1^* - L_2^*)^2 + (a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2}\,. \end{aligned} \quad (1)$$

As mentioned in [7], the value of ΔE^*_{ab} is perceptually analogous to human's visual perception of color difference. Moreover, the values of ΔE^*_{ab} can be roughly classified into three different levels to reflect the degrees of color difference perceived by human. As shown in Table 1, the color difference is hardly perceptible when ΔE^*_{ab} is smaller than 3; is perceptible but still tolerable when ΔE^*_{ab} is between 3 and 6; and is usually not acceptable when ΔE^*_{ab} is larger than 6 [7]. Hence, in this paper, we define a color difference to be "visible" if its ΔE^*_{ab} value is larger than 6.

Table 1 [7]

ΔE^*_{ab}	*Effect*
< 3	*Hardly perceptible*
3 < 6	*Perceptible, but acceptable*
> 6	*Not acceptable*

Based on the definition of visible color difference, we then define two measurements to evaluate the quality of color segmentation. The first measurement, named "intra-region visual error", is designed to evaluate the degree of under-segmentation. In each segmented region, these pixels with ***visible*** color difference away from the average color of that region are regarded pixels with visible color errors. Intuitively, a properly segmented region should contain as few visible color errors as possible. Given an $N \times M$ color image f, we first denote $\hat{f}$ as the segmented color image, with the color of each segmented region being filled with the average color of that region. We then define the intra-region error as:

$$E_{\text{intra}} = \frac{\left(\sum_{x=1}^{N}\sum_{y=1}^{M} u(\left\|f(x,y) - \hat{f}(x,y)\right\|_{L^*a^*b^*} - th)\right)}{N \times M}\,, \quad (2)$$

where $\| \ \|_{L^*a^*b^*}$ denotes the color difference in the CIE L*a*b* space, th denotes the threshold for visible color difference, and u(.) denotes the step function:

$$u(t) = \begin{cases} 1, & t > 0 \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

Here, we choose the threshold "*th*" to be 6.

On the other hand, the second measurement, named "inter-region visual error", is designed to evaluate the degree of over-segmentation. Given a color segmentation result, we take into account these boundary pixels with ***invisible*** color difference across the boundary. Intuitively, these pixels are not supposed to be treated as boundaries. Hence, the inter-region visual error of a segmented image is defined as:

$$E_{\text{inter}} = \frac{\sum_{i=1}^{R}\sum_{\substack{j=1 \\ j \neq i}}^{R} w_{ij} \times u(th - \left\|\hat{f}_i - \hat{f}_j\right\|_{L^*a^*b^*})}{C \times N \times M}\,, \quad (4)$$

where R denotes the number of segmented regions, w_{ij} denotes the joined length between Region i and Region j and is equal to zero if Region i and Region j are not connected, and C denotes a normalization factor. Here, C = 1/6, which was determined empirically.

Note that, for a segmented image, a large value of intra-region visual error means plenty of pixels may be mistakenly merged and this image could have been under-segmented. On the other hand, a large value of inter-region visual error means plenty of boundary pixels may be mistakenly generated and the image could have been over-segmented. Moreover, there is a reciprocal relationship between intra-region error and inter-region error. As we adjust the controlling parameters of a segmentation algorithm to merge more regions together, the inter-region error decreases while the intra-region error increases. On the contrary, as we segment an image into more regions, the intra-region error decreases while the inter-region error increases. An illustration of this reciprocal property is shown in Fig. 1. Note that, the number of segmentation regions, which is commonly used in some evaluation methods, is subtly excluded in the evaluation of segmentation performance. This is because the number of homogeneous regions could be very different in different images. It would be difficult to develop a fair evaluation scheme based on such an image-dependent measurement.

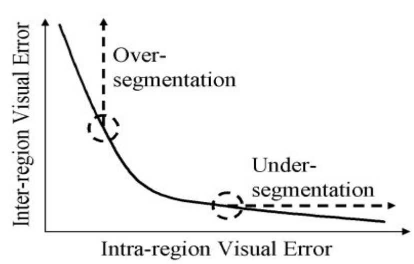


Fig.1. The plot of intra-region visual error v.s. inter-region visual error as we adjust the controlling parameters of a segmentation algorithm.

3. EVALUATION OF SEGMENTATION

In this paper, we focus on the evaluation of color segmentation for these images without complex texture. In this section, the use of the Inter-Region-Error/Intra-Region-Error plot in the evaluation of color segmentation is to be introduced. First, the evaluation of segmentation results for a given segmentation algorithm is described. Then, the evaluation method for the comparison of various segmentation algorithms is to be presented.

3.1 Evaluation of segmentation results

Fig. 2 (a) shows a color image and Fig. 2 (b)(c)(d)(e)(f) show several segmentation results of Fig. 2(a) produced by the JSEG algorithm [8], with different parameter settings. Subjectively, Fig. 2(e) is preferable. In comparison with Fig. 2(e), Fig.2 (b)(c)(d) are over-segmented, while Fig. 2(f) is under-segmented. As shown in Fig. 2(m), from left to right, five blue circles represent the intra-region visual error and the inter-region visual error pairs of Fig.2(b)(c)(d)(e)(f), respectively. It can be easily seen that, with a similar intra-region error, Fig. 2(b)(c)(d) have larger inter-region error values than that of Fig. 2 (e). On the other hand, with a similar inter-region error, Fig. 2(f) has a larger intra-region error value than that of Fig. 2(e). Hence, in the selection of parameter setting, a weighted sum of E_{intra} and E_{inter} may serve as a suitable criterion for the evaluation of segmentation performance. As the weighted sum reaches a smaller value, the parameter setting is regarded as achieving a better segmentation. In Fig. 2(m), we use the simplest combination $E_{intra} + E_{inter}$ to illustrate this idea. Here we use gray straight lines to denote the lines $E_{intra} + E_{inter}$ = constant. It can be easily seen that Fig. 2(e) does have the smallest weighted sum if compared with the other four.

In Fig. 2 (g), we show another example of color image. Fig. 2(h)(i)(j)(k)(l) show its segmentation results produced by the Mean-Shift algorithm[9], with different parameter settings. Similarly, in Fig.2 (m), from left to right, the intra-region error and the inter-region error pairs of Fig. 2(h)(i)(j)(k)(l) are represented in five red stars, respectively. It can be easily seen that Fig. 2(i) has the smallest weighted error sum and the segmented result in Fig. 2(i) does correspond to a preferred result.

In summary, the above two simulation results demonstrate how the Inter-Region-Error/Intra-Region-Error plot can be used to automatically select the parameter setting based on the performance of segmentation results. Actually, E_{intra} and E_{inter} can be combined in different forms to fit for user's requirements. So far, we found that the simple form $E_{intra} + E_{inter}$ performs pretty well and reliably when applied to various types of color images.

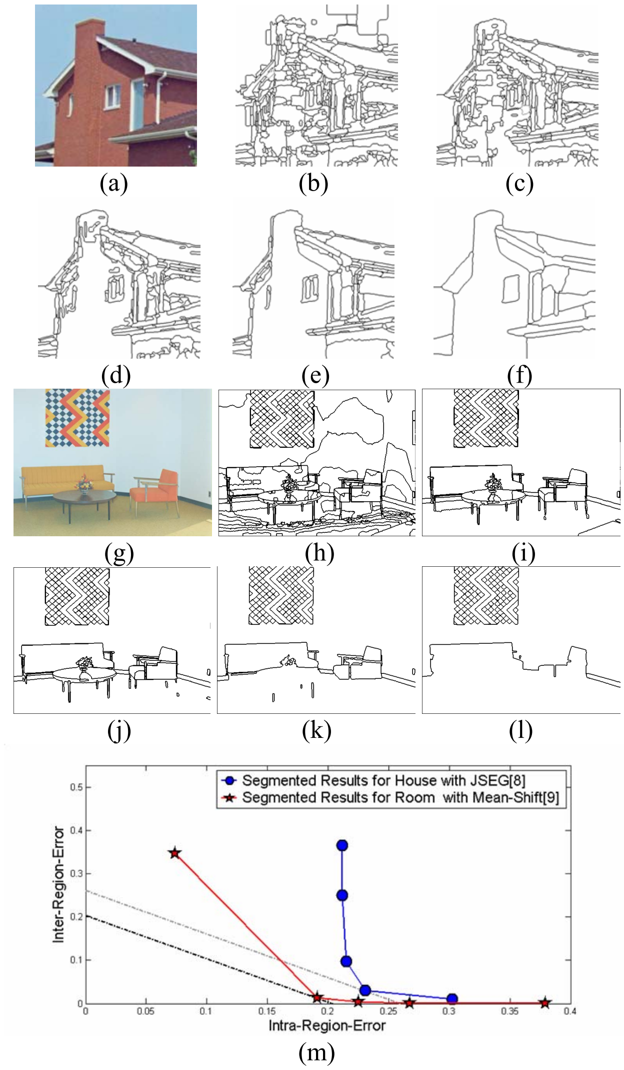


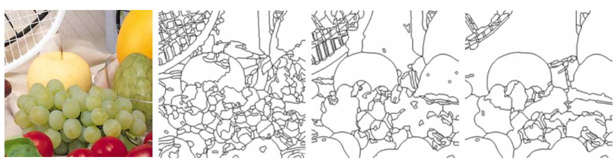
Fig.2. Evaluation of segmented results.
(a)(g) Original color images.
(b)(c)(d)(e)(f) Segmented results of (a) using JSEG algorithm [8].
(h)(i)(j)(k)(l) Segmented results of (g) using Mean-Shift algorithm [9].
(m) Intra-Region-Error vs. Inter-Region-Error of (b)(c)(d)(e)(f) and (h)(i)(j)(k)(l).

3.2 Evaluation of segmentation algorithms

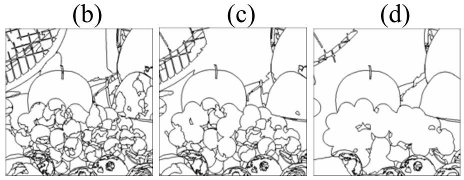
The performance comparison between different segmentation algorithms is also an important issue in image segmentation. In the evaluation of segmentation algorithms, an objective and quantitative evaluation is essential. In this section, we will demonstrate how to use the Inter-Region-Error/Intra-Region-Error plot to compare the performance of various segmentation algorithms.

Fig. 3(a) shows a color image and several segmented results produced by both the JSEG algorithm [8] and the

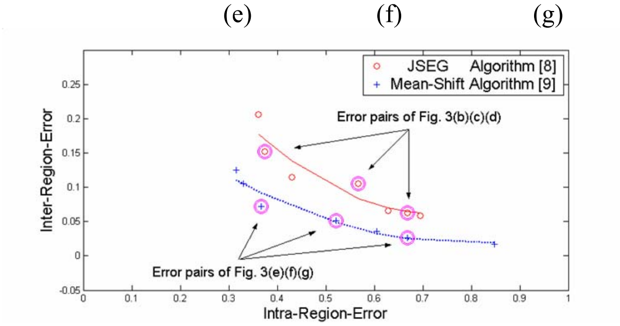
Mean-Shift algorithm [9]. In Fig. 3(h), the red circles denote the inter-region error and the intra-region error pairs of 7 segmentation results produced by the JSEG algorithm, while the blue crosses represent the error pairs of 7 segmentation results produced by the Mean-Shift algorithm. Fig. 3(b)(c)(d) show three of these seven segmentation results produced by the JSEG algorithm, while Fig. 3(e)(f)(g) show three of these seven segmentation results produced by the Mean-Shift algorithm. With these sample pairs, a 2nd-order curve fitting is adopted to estimate the tendency between the inter-region error and the intra-region error in Fig. 3(h). Here, the red curve shows the performance tendency of the JSEG algorithm, while the blue curve shows the performance tendency of the Mean-Shift algorithm. With these two tendency curves, it can be easily seen that with a similar intra-region error, the segmentation results produced by the Mean-Shift algorithm tend to have a smaller inter-region error. That is, with a similar intra-error value, the segmentation results produced by the JSEG algorithm tend to be more over-segmented. Similarly, with a similar inter-region error, the segmentation results produced by the Mean-Shift algorithm tend to have a smaller intra-region error.



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



(e) (f) (g)



(h)

Fig.3. Comparison with different algorithms.
(a) Original image.
(b)(c)(d) Segmentation results using the JSEG algorithm [8].
(e)(f)(g) Segmentation results using the Mean-Shift algorithm [9].
(h) Comparison of the JSEG algorithm and the Mean-Shift algorithm.

4. CONCLUSION

In this paper, we describe a new goodness evaluation method for color segmentation, based on the use of visible color difference. To evaluate the quality of segmentation results, we define intra-region visual error and inter-region visual error to measure the degrees of under-segmentation and over-segmentation, respectively. With these two measures, a complete evaluation scheme is proposed to evaluate the parameter settings of a given segmentation algorithm and to compare the segmentation performance between different algorithms. The simulation results have demonstrated the potential of this approach in providing reliable and efficient evaluations over the performance of color image segmentation.

5. REFERENCES

[1] L. Lucchese and S. K. Mitra, “Color Image Segmentation: A State-of-the-Art Survey,” in *Proc. the Indian National Science Academy* (*INSA-A*), New Delhi, India, vol. 67, A, no. 2, pp. 207-221, Mar. 2001.

[2] Y. J. Zhang, “A Survey on Evaluation Methods for Image Segmentation,” *Pattern Recognition*, vol. 29, no. 8, pp. 1335-1346, Aug. 1996.

[3] Y. J. Zhang, “A Review of Recent Evaluation Methods for Image Segmentation,” in *Proc. 6th Int. Symp. on Signal Processing and Its Applications*, pp. 148-151, Kuala Lumpur, Malaysia, Aug. 2001.

[4] D. Martin, C. Fowlkes, D. Tal and J. Malik, “A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics,” in *Proc. IEEE Int’l. Conf. Computer Vision*, vol. 2, pp. 416-423, Vancouver, Canada, July 2001.

[5] M. Borsotti, P. Campadelli and R. Schettini, “Quantitative Evaluation of Color Image Segmentation Results,” *Pattern Recognition Letters*, vol. 19, no. 8, pp. 741-747, June 1998.

[6] C. Rosenberger and K. Chehdi, “Genetic Fusion: Application to Multi-components Image Segmentation,” in *Proc. ICASSP*, vol. 4, pp. 2223-2226, Istanbul, Turkey, June 2000.

[7] J. Y. Hardeberg, “Acquisition and Reproduction of Colour Images: Colorimetric and Multispectral Approaches,” A dissertation submitted in partial fulfillment of the degree of “Docteur de l’Ecole Nationale Supérieure des Télécommunications,” Paris, France, Feb. 1999.

[8] Y. Deng and B.S. Manjunath, “Unsupervised Segmentation of Color-Texture Regions in Images and Video,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 8, pp. 800-810, Aug. 2001.

[9] D. Comaniciu and P. Meer, “Mean Shift: A Robust Approach toward Feature Space Analysis,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 603-619, May 2002.