2L-5

Perception-based Evaluation of Segmentation Algorithms

David GAVILAN† Hiroki TAKAHASHI† Suguru SAITO† Masayuki NAKAJIMA†‡

†Graduate School of Information Science & Engineering, *Tokyo Institute of Technology*

‡ *National Institute of Informatics*

1 Introduction

A lot of segmentation methods have been proposed from the very beginning of Computer Vision, successful under certain conditions for certain tasks. However, global purpose segmentation in uncontrolled environments, usually based on human perception, has gained attention in recent days. In this paper, we propose an evaluation method to measure how much close are these segmentation results to human perception. Such a comparison is very important in applications that require human interaction. For instance, if we want to allow the user to search image databases by means of sketches, the representation in our database should be close to that sketch. Therefore, a method to evaluate how close are the representations in the computer to those of human brains is needed. Such a method will be able to judge the goodness of a Computer Vision application, a Human-Computer interface, or a Non-Photo-realistic Rendering algorithm.

2 Related Work

Traditionally, there has been two ways to judge good segmentations: by showing examples or by measuring performance in a certain task. The first one is very common when the segmentation algorithm has not yet been applied to any particular case, and the evaluation is left for the critical eye of the reader [2]. When an application is provided, the performance of the segmentation algorithm in that particular application is shown, such as the ability to retrieve images in an image database, the performance in a tracking system, or similar measurements.

With the aim of obtaining a more precise characterization, in this paper we propose an evaluation method based on a collection of drawings made by users. By doing this, the data collected pertains to the same domain as the output of what we want to evaluate, that is, to the domain of digital pictures.

3 Evaluation Method

Given an original image w, let us call $p(w)$ the *percept* that represents that image in the user's mind. Since there is no way of knowing $p(w)$, it will be approximated by the the sketch drawn by the user $s(m)$, where m is the mind state that contains the set $\mathcal{P} = \{p_1(w), p_2(w), p_3(w), ...\}$ (variations of $p(w)$ across time). For the sake of simplicity, let us note it as $s(w)$. Let us characterize the perception of image w by $\boldsymbol{h}_{s(w)}$, being the feature vector $\boldsymbol{h}$ of an image l

$$\boldsymbol{h}_l \equiv (\#regions(l), \frac{\sum(\nabla l > t)}{size(l)}), \qquad (1)$$

where t is a threshold to decide which levels of the gradient count as contours.

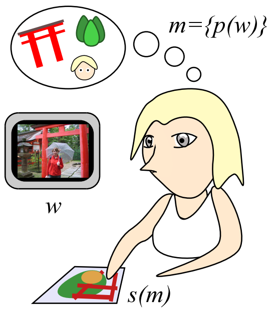

Figure 1: Visual Perception model.

If $G(w)$ is a segmentation of the image w, we define a segmentation close to human perception as one that minimizes the distance between $\boldsymbol{h}_{G(w)}$ and $\boldsymbol{h}_{s(w)}$. Since the two features of $\boldsymbol{h}$, the contours and the number of regions, are correlated, let us use the Mahalanobis distance

$$((\boldsymbol{h}_{G(w)} - \boldsymbol{h}_{s(w)})^T \mathbf{C}^{-1}(\boldsymbol{h}_{G(w)} - \boldsymbol{h}_{s(w)}))^{\frac{1}{2}}. \qquad (2)$$

The covariance matrix $\mathbf{C}$ will be estimated with the feature vectors extracted from the sketches. Then, we will define the distance between two segmented images simply as the Mahalanobis distance between their feature vectors,

$$d(G(l), G'(l)) \equiv d(\boldsymbol{h}_{G(w)}, \boldsymbol{h}_{G'(w)}). \qquad (3)$$

The distance between a segmented image and the sketch associated with the original image will be used as an evaluation function for a given segmentation algorithm.

4 Data collection

18 users of ages from 20 to 40 participated in the experiment, making the total number of photographs used as stimuli 210, from a database of 1476 images. The users were prompted to select one of them and draw an sketch representing it. In figure 2 there are some examples of the sketches collected.

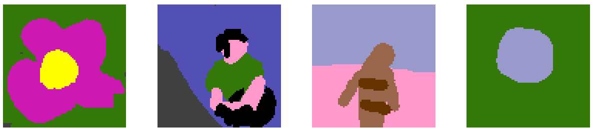

Figure 2: Some of the sketches collected from users.

5 Results

With the $\boldsymbol{h}_l$ values of the sketches collected in our experiment, the covariance matrix $\mathbf{C}$ needed for equation (2) is estimated. Using the set of 210 sketches, this matrix is

$$\mathbf{C} = \begin{pmatrix} 25.1711 & 0.2359 \\ 0.2359 & 0.0041 \end{pmatrix} \tag{4}$$

Three different color segmentation algorithms were chosen to test the evaluation function; the mean shift based color segmentation algorithm [2], the Blobworld system [1], and our Color Blobs approach [3].

The mean shift algorithm has two basic parameters we have to set; the bandwidth parameter $\mathbf{h} = (\mathbf{h_s}, \mathbf{h_r})$, and the minimum number of pixels M needed to form a region. We will set them to minimize our function.

The Blobworld system uses the Expectation-Maximization (EM) algorithm to determine the maximum likelihood parameters of a mixture of K Gaussians. These values of K are iterated and the best one is selected automatically.

The Color Blobs algorithm depends on the neural network used to learn color categories. We will use the network presented in [4].

Figure 3 is a plot of the $\boldsymbol{h}_l$ vectors of the sketches and the three different methods, using the minimizing parameters for each one.

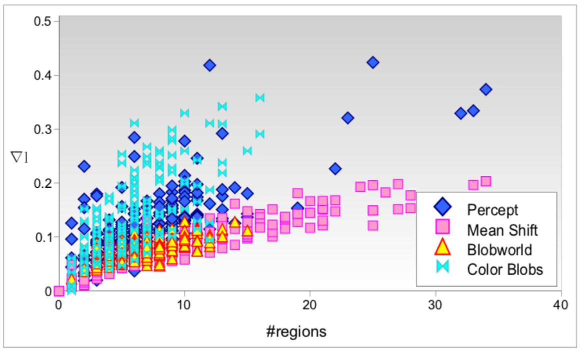

Figure 3: Feature points $\boldsymbol{h}_l$ for different segmentation algorithms.

The distribution of distances to the original set of images is given by the following averages:

$d(G,s)$	Mean Shift	Blobworld	C.Blobs
$Var(d)$	1.64303	0.73269	0.63074
$E(d)$	1.87935	1.24231	1.07811

From this table, we can observe that the variance of the mean-shift algorithm is very high. That is due to the varying nature of the pictures, but also to the inability of the algorithm to adapt if its parameters are fixed. The similar results of Color Blobs and Blobworld not only show that they are indeed closer to human perception, but also shows that is not necessary a complex model to obtain these results (check [4] for a speed comparison between the two methods). Figure 4 is a visual example of the different segmentations, compared to a sketch.

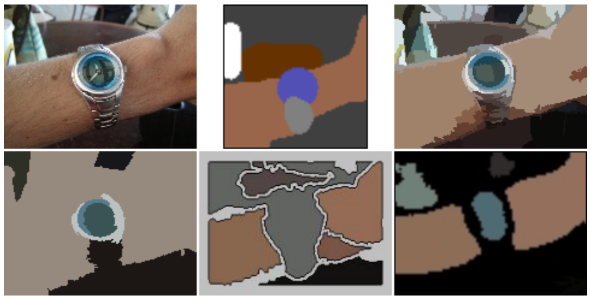

Figure 4: From left to right, up-down: original image, sketch drawn by an user, segmentation using Mean Shift $(4, 4, 50)$, Mean Shift $(4, 16, 192)$, Blobworld, and Color Blobs.

6 Conclusions and future work

A basic test to evaluate whether or not an segmentation method is congruent with human perception has been proposed. A comparison between different segmentation algorithms shows that our previously presented segmentation method is closer to human perception, thus, more appropriate to be used in applications that require user's interaction. This test can be extended to contain other features, such as the graphs representing the objects in the scene. Furthermore, with the aid of this evaluation method we plan to improve our Color Blobs segmentation method.

References

[1] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, August 2002.

[2] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.

[3] D. Gavilan, H. Takahashi, and M. Nakajima. Image categorization using color blobs in a mobile environment. *Computer Graphics Forum (EG 2003)*, 22(3):427–432, September 2003.

[4] D. Gavilan, H. Takahashi, and M. Nakajima. Color blobs-based image retrieval in broad domains. In *IWAIT*, January 2005.