Which color similarity measure is most effective for background-frame differencing ?

Michael Hild and Toshinobu Emura
Osaka Electro-Communication University, Dept. of Engineering Informatics
Neyagawa, Osaka, Japan

Abstract

We examine twelve known color similarity measures with regard to their effectiveness for the background-frame differencing task. The RGB sensor space and the CIE $L^* a^* b^*$ color space is used to represent color. Based on experiments conducted on twelve different scenes (indoor as well as outdoor scenes under various lighting conditions and photographed with two different types of CCD cameras) we show that the *Absolute-value exponent method* is superior.

1. Introduction

In recent years, research on image sequences analysis for the purpose of moving object tracking and recognition has steadily increased in the fields of computer vision and pattern recognition. In this context, many researchers have used background-frame differencing as a first, low-cost processing step to extract regions of moving objects [1, 2, 3]. The camera is kept stationary and a photo of the background is taken as a reference image at the start of the image sequence. Then the color difference between the image frames of the sequence and the reference image is computed, and if this difference exceeds a preset value at a pixel, it is considered to be an object pixel.

Although this kind of background-frame differencing plays an important role in image sequence processing because of its simplicity and high processing speed, it has the drawback that (a) similar background and object colors often cannot be distinguished (i.e. motion region extraction fails), (b) shadows cast onto the scene by the moving object appear as object regions (i.e. background extraction fails), and (c) random color deviations due to sensor noise can cause either one of these phenomena to occur.

The simplest background-frame differencing methods use only the lightness part of color, in which case the computation is carried out on scalar values, but more sophisticated methods compute the color difference based on vector representations of color. Another way of realizing background-frame differencing is to compute the similarity between two colors: if color vectors are sufficiently similar, the pixel would be judged as belonging to the background. Here the question of which color similarity measure might be best suited for this task arises. Even more sophisticated methods try to model noise and textural statistics within small image neighborhoods, and methods have been proposed that normalize the images with respect to various image properties (e.g. mean lightness) before background-frame differencing is carried out.

In this paper we focus on the color similarity approach and examine twelve known color similarity measures with regard to their effectiveness for the background-frame differencing task. As color spaces we use (a) the RGB sensor spaces of (off-the-shelf) CCD cameras, and (b) the CIE $L^* a^* b^*$ color spaces which are derived from these RGB spaces through a non-linear transformation. Based on experiments conducted on twelve different scenes (indoor as well as outdoor scenes under various lighting conditions and photographed with two different types of CCD cameras) we are able to show that among the twelve similarity measures there is one that clearly is superior.

2. Color similarity measures

The twelve color similarity measures all have been proposed in the literature [4, 5]. All of them take two vectors as input and compute a real number in the range (0.0, 1.0), where value 1.0 indicates "identical" and "0.0" is synonymous with "not similar at all." The similarity measures used are listed below. The similarity measures are represented as functions $S_k(\mathbf{x}_i, \mathbf{x}_j), \quad k = 1, 2, ..., 12$, where $\mathbf{x}_i, \mathbf{x}_j$ are the two p-dimensional color vectors. The arguments $(\mathbf{x}_i, \mathbf{x}_j)$ are skipped in the formulae shown below.

Measure 1

$$S_1 = \frac{\mathbf{x}_i \mathbf{x}_j^t}{|\mathbf{x}_i|\,|\mathbf{x}_j|} = \cos(\theta) \tag{1}$$

This measure uses the angle between the two color vectors to represent similarity.

Measure 2

$$S_2 = \left(\frac{\mathbf{x}_i \mathbf{x}_j^t}{|\mathbf{x}_i||\mathbf{x}_j|}\right)\left(1 - \frac{||\mathbf{x}_i|\,|\mathbf{x}_j||}{\max(|\mathbf{x}_i|\,,|\mathbf{x}_j|)}\right) \tag{2}$$

This similarity measure takes into account both the angle between the vectors and their magnitudes.

Measure 3

$$S_3 = \frac{|\mathbf{x}_i| \cos(\theta) + |\mathbf{x}_j| \cos(\theta)}{\left(|\mathbf{x}_i|^2 + |\mathbf{x}_j|^2 + 2|\mathbf{x}_i||\mathbf{x}_j| \cos(\theta)\right)^{\frac{1}{2}}} \quad (3)$$

When $|x_i| = |x_j|$ is satisfied, this measure reduces to

$$S_3 = \cos(\theta) / \cos\left(\frac{\theta}{2}\right) \quad (4)$$

Measure 4

$$S_4 = \frac{\cos(\theta)\left(|\mathbf{x}_i|^2 + |\mathbf{x}_j|^2 + 2|\mathbf{x}_i||\mathbf{x}_j| \cos(\theta)\right)^{\frac{1}{2}}}{|\mathbf{x}_i| + |\mathbf{x}_j|} \quad (5)$$

This similarity measure has some affinity to S_3. When $|\mathbf{x}_i| = |\mathbf{x}_j|$ is satisfied, it reduces to

$$S_4 = \cos(\theta) \cos\left(\frac{\theta}{2}\right) \quad (6)$$

Measure 5

$$S_5 = 1 - \frac{\left(|\mathbf{x}_i|^2 + |\mathbf{x}_j|^2 - 2|\mathbf{x}_i||\mathbf{x}_j| \cos(\theta)\right)^{\frac{1}{2}}}{\left(|\mathbf{x}_i|^2 + |\mathbf{x}_j|^2 + 2|\mathbf{x}_i||\mathbf{x}_j| \cos(\theta)\right)^{\frac{1}{2}}} \quad (7)$$

In this measure emphasis is on the dissimilarity between the two vectors.

Measure 6: Correlation coefficient method

$$S_6 = \frac{\sum_{k=1}^{p} |\mathbf{x}_{ik} - \bar{\mathbf{x}}_i| |\mathbf{x}_{jk} - \bar{\mathbf{x}}_j|}{\left(\sum_{k=1}^{p} (\mathbf{x}_{ik} - \bar{\mathbf{x}}_i)^2\right)^{\frac{1}{2}} \left(\sum_{k=1}^{p} (\mathbf{x}_{jk} - \bar{\mathbf{x}}_j)^2\right)^{\frac{1}{2}}} \quad (8)$$

where $\bar{\mathbf{x}}_i = \frac{1}{p} \sum_{k=1}^{p} \mathbf{x}_{ik}$.

Measure 7: Exponential similarity method

$$S_7 = \frac{1}{p} \sum_{i=1}^{p} \exp\left(\frac{-3}{4} \cdot \frac{(\mathbf{x}_{ik} - \mathbf{x}_{jk})^2}{\beta_k^2}\right) \quad (9)$$

where $\beta_k^2 > 0$ is a parameter that is determined experimentally.

Measure 8: Absolute-value exponent method

$$S_8 = \exp\left(-\beta \sum_{k=1}^{p} |\mathbf{x}_{ik} - \mathbf{x}_{jk}|\right) \quad (10)$$

where $\beta > 0$.

Measure 9: Absolute-value reciprocal method

$$S_9 = 1 - \beta \sum_{k=1}^{p} |\mathbf{x}_{ik} - \mathbf{x}_{jk}| \quad (11)$$

β is selected through experiments.

Measure 10: Maximum-minimum method

$$S_{10} = \frac{\sum_{k=1}^{p} \min(\mathbf{x}_{ik}, \mathbf{x}_{jk})}{\sum_{k=1}^{p} \max(\mathbf{x}_{ik}, \mathbf{x}_{jk})} \quad (12)$$

Measure 11: Arithmetic-mean minimum method

$$S_{11} = \frac{\sum_{k=1}^{p} \min(\mathbf{x}_{ik}, \mathbf{x}_{jk})}{\frac{1}{2} \sum_{k=1}^{p} (\mathbf{x}_{ik} + \mathbf{x}_{jk})} \quad (13)$$

Measure 12: Geometric-mean minimum method

$$S_{12} = \frac{\sum_{k=1}^{p} \min(\mathbf{x}_{ik}, \mathbf{x}_{jk})}{\sum_{k=1}^{p} (\mathbf{x}_{ik} \mathbf{x}_{jk})^{\frac{1}{2}}} \quad (14)$$

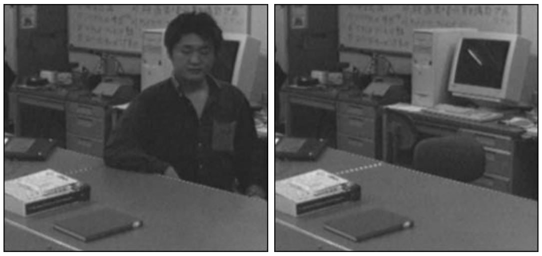
3. Test method

For testing the twelve color similarity measures the following procedure is used: The scene is photographed as the *background* reference and immediately after that a person in front of the background is photographed as the *frame* using the same camera settings (see example of Fig.1(a),(b)). In order to reduce sensor noise, the images are convolved with a Gaussian filter mask of size 7x7 pixels at all image locations.

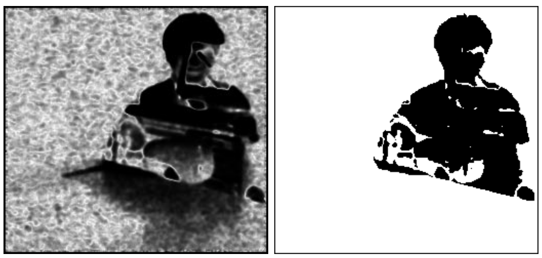
Then color similarity is computed for all twelve background-frame image pairs at every image location, using all twelve similarity measures, and the results are recorded. An example is shown in Fig.1(c) where the values of computed similarity are coded as gray values between white and black.

The decision of whether a pixel belongs to the background or the object region is made by counting all pixels with similarity value above a threshold as object region pixels. We could have chosen this threshold to be a fixed value set by the user, but this would make the test method somewhat arbitrary. We therefore decided to use an automatic threshold selection method. The threshold selection method we use was proposed in [6], and it determines the threshold such that the variances of values in the two classes resulting from the thesholding operation are minimized and the inter-class variance is maximized. This way, threshold selection is based on a sound statistical principle. An example of the result of thresholding is shown in Fig.1(d).

In order to evaluate the effectiveness of each similarity measure for background-frame differencing we prepared



(a) Frame (Scene 1) (b) Background



(c) Gray value-coded color similarity (white := 0.0, black := 1.0) (d) Extracted object pixels

Figure 1: *Processing steps of test method*

a ground-truth image by hand for the object regions in each scene. An example is shown in Fig.2. The *object region extraction rate* is determined by applying a pixel-wise AND-operation to the ground-truth and thresholded similarity value images and dividing the obtained number of true-pixels by the total number of pixels contained in the ground-truth object region.

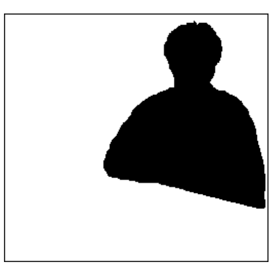


Figure 2: *Ground-truth for scene in Fig.1*

4. Experimental results

The scene images used in the experiment are displayed in Fig.1(a) and Fig.3. They include indoor as well as outdoor scenes photographed under various lighting conditions. The moving objects are always persons. The camera used for taking indoor scenes was a 3CCD camera for industrial applications made by Sony Corporation and the camera used for taking outdoor scenes was an electronic still camera (fully automatic dial setting) made by Kyocera Corporation. The background colors included largely monochrome as well as highly saturated colored surfaces, both man-made and natural.

As an example, object region extraction results using the RGB color spaces are shown in Fig.4 for Scene 1. It is obvious that the color similarity measures have a very wide spread of effectiveness, ranging from "bad" to "good." Results for the $L^*a^*b^*$ color spaces (not shown) were similar. Graphs displaying the object region extraction rates for all twelve scenes, all twelve measures, and both color spaces are shown in Fig.6. A graph displaying the object region extraction rate averaged over all twelve scenes for each similarity measure is shown in Fig.5.

These results allow to make the following observation regarding the effectiveness of color similarity measures for background-frame differencing:

1. The on-average most effective color similarity measure is the *Absolute-value exponent method* (Measure 8), followed by the *Exponential similarity method* (Method 7). Especially Measure 8 does well for both RGB and $L^*a^*b^*$ color spaces. Measures 2 and 10 have quite acceptable performance.

2. The on-average least effective color similarity measures are Measures 1, 3, and 4 for both RGB and $L^*a^*b^*$ color spaces, and Measure 12 is not effective for the $L^*a^*b^*$ color space. The bad performance of Measure 1 is due to neglecting the lightness part of color.

3. The performance of Measures 5, 6, 9, 11, and 12 is highly volatile, strongly depending on the scenes being processed.

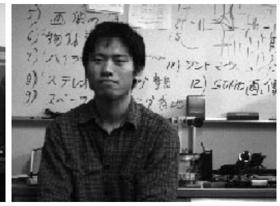
Although Measure 8 is best overall, one still has to decide which one of the two color spaces is best when Measure 8 is used. Based on results not included in this paper it appears that the $L^*a^*b^*$ color space is more appropriate because (a) on average fewer background pixels are extracted, and (b) fewer shadow pixels are extracted than if the RGB color space was used.

5. Conclusion

In this paper we have raised the question of which color similarity measure is most effective for the task of background-frame differencing. Background-frame differencing is important for image sequence analysis because it is

(a) Scene 2

(b) Scene 3



(c) Scene 4

(d) Scene 5

(e) Scene 6

(f) Scene 7

(g) Scene 8

(h) Scene 9

(i) Scene 10

(j) Scene 11

(k) Scene 12

Figure 3: *Test scenes*

relatively low-cost in terms of processing speed. Based on experiments involving twelve indoor and outdoor scenes, we conclude that among the twelve color similarity measures tested the *Absolute-value exponent method* in combination with the $L^*a^*b^*$ color space is most effective for the task at hand.

References

[1] I. Haritaoglu, D. Harwood, and L. S. Davis, W4: Who? when? where? what? a real-time system for detecting and tracking people, Proc. The Third IEEE Internationall Conference on Automatic Face and Gesture Recognition, Nara Japan, 222-227 (1998)

[2] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, Pfinder: Real-time tracking of the human body, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.7, 780-785 (1998)

[3] M. Yamada, K. Ebihara, and Jun Ohya, A new robust real-time method for extracting human silhouettes from color images, Proc. The Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara Japan, 528-533 (1998)

[4] K. N. Plataniotis and A. N. Venetsanopoulos, Color Image Processing and Applications, Springer-Verlag, Berlin (2000)

[5] Y. Leung, Spatial Analysis and Planning under Imprecision, North-Holland Publishers, The Netherlands (1988)

[6] N. Otsu, Discriminant and Least Squares Threshold Selection, Proc. 4th Conference on Pattern Recognition, 592-596, Kyoto Japan (1978)

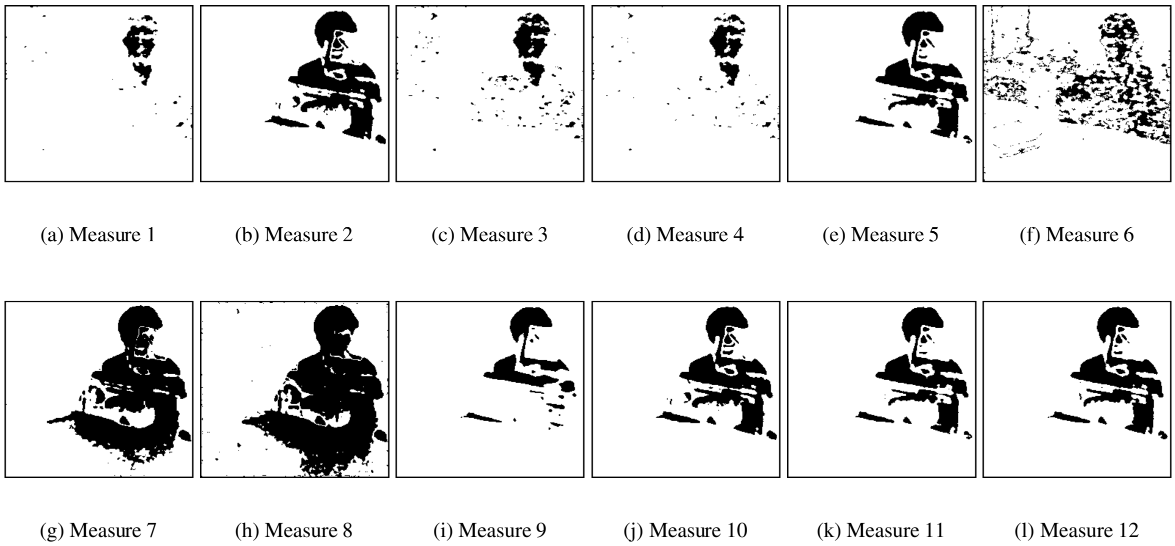
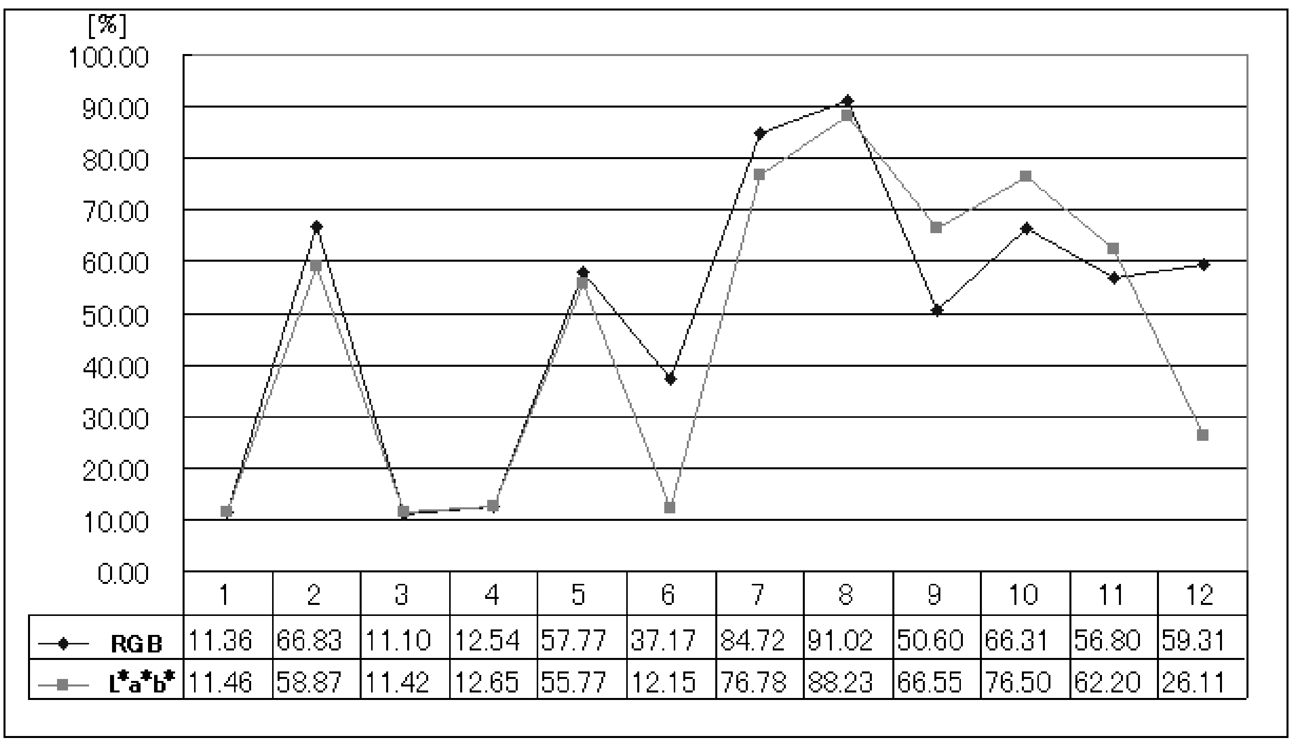
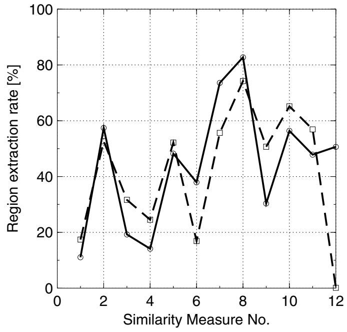


Figure 4: *Extracted regions of Scene 1 using different similarity measures and RGB color space*

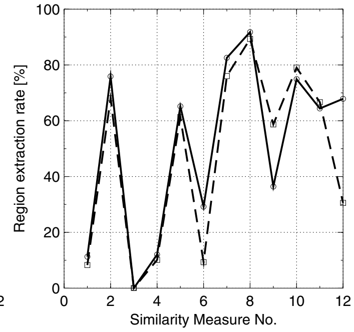


	1	2	3	4	5	6	7	8	9	10	11	12
RGB	11.36	66.83	11.10	12.54	57.77	37.17	84.72	91.02	50.60	66.31	56.80	59.31
L*a*b*	11.46	58.87	11.42	12.65	55.77	12.15	76.78	88.23	66.55	76.50	62.20	26.11

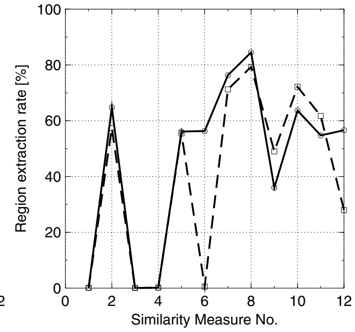
Figure 5: *Object region extraction rates averaged over all twelve scenes*



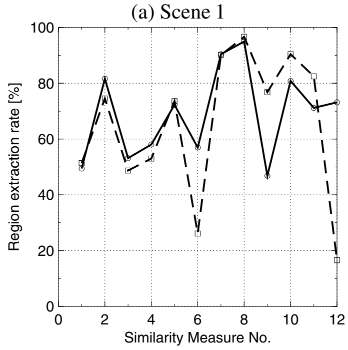
(a) Scene 1



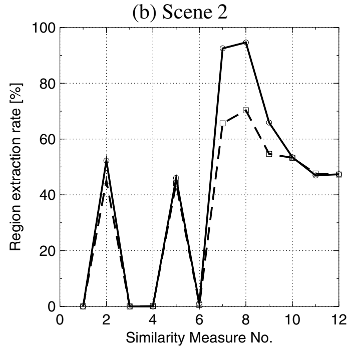
(b) Scene 2



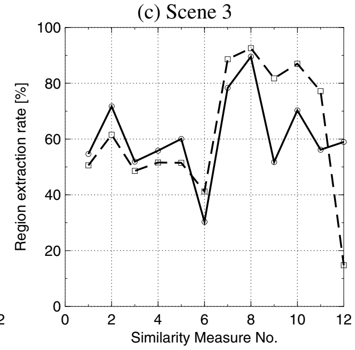
(c) Scene 3



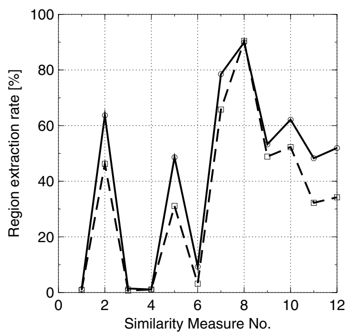
(d) Scene 4



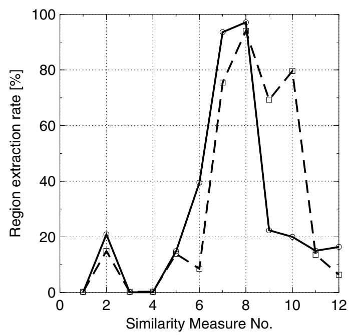
(e) Scene 5



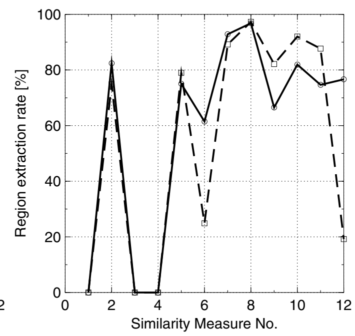
(f) Scene 6



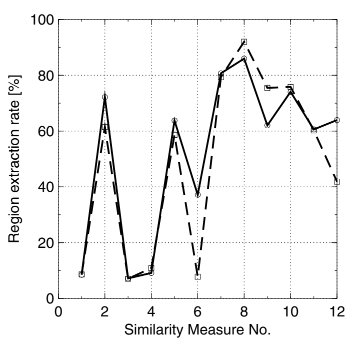
(g) Scene 7



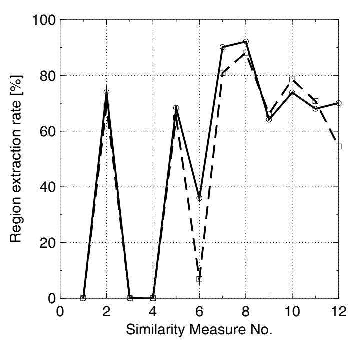
(h) Scene 8



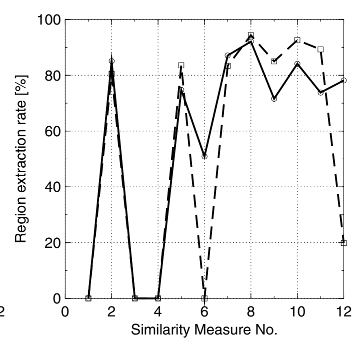
(i) Scene 9



(j) Scene 10



(k) Scene 11



(l) Scene 12

Figure 6: *Object region extraction rates for all twelve scenes, all twelve measures, and both color spaces (dashed line: $L^*a^*b^*$)*