

Block-Based Error Measure for Object Segmentation

Tsorng-Lin Chia and Chaur-Heh Hsieh*

Department of Computer and Communication Engineering, Ming Chuan University

ABSTRACT

Object segmentation is an important research topic in computer vision. The assessment of the quality of the segmentation results is of crucial importance. The conventional performance measure for object segmentation is mainly based on the pixel error between the segmented object and ground truth. The pixel-based error measure does not consider the spatial distribution of segmentation errors, which is essential in semantic processing. This paper presents a novel block-based error measure for evaluating the performance of object segmentation. We first analyze the spatial distribution of segmentation errors and classify them into scattered error and region error. Then we develop a block-based error measure that enhances the contribution of the region error. The mathematical analysis of both error measures is also presented to demonstrate the advantages of the proposed block-based measure.

Keywords: block-based error measure, object segmentation, pixel-based error measure, relative evaluation

用於物件分割之區塊錯誤量測準則

賈叢林 謝朝和*

銘傳大學 資訊傳播工程學系

摘 要

物件分割是電腦視覺中重要的研究主題，而如何評估分割結果的品質是非常重要的議題。傳統的物件分割效能的評估主要都是建立在分割物件與標準物件之畫素錯誤上。這種以畫素錯誤為基礎的量測技術沒有考慮分割錯誤的空間分佈(spatial distribution)，而此空間分佈在影像意涵處理上是很重要的。本文提出一用於評估物件分割效能之區塊錯誤為基礎的量測新技術。首先，本文分析分割錯誤的空間分佈，然後將其分類為分散型錯誤與集中型錯誤，最後提出區塊錯誤為基礎的量測技術，此技術強化集中型錯誤的貢獻度。我們也以數學分析顯示此新量測技術的優異性。

關鍵詞：區塊錯誤量測，物件分割，畫素錯誤量測，相對評估

文稿收件日期 106.11.27;文稿修正後接受日期 107.6.4;*通訊作者

Manuscript received November 27, 2017; revised June 4, 2018;*Corresponding author

I. INTRODUCTION

Object segmentation (or detection) from images or videos has received wide attention recently since it has a large variety of applications such as object recognition, video surveillance, human-machine interface (HMI), traffic monitoring, object-based video coding [1], [2]. Extensive efforts have been made to develop object segmentation (detection) techniques either for images or videos, but much less attention has been paid for the performance evaluation of these techniques [3], [4]. Generally, for still images, the performance evaluation is to calculate the error between the segmented object and a reference (ground-truth) object. The reference object is often obtained manually, and the segmented object is from a segmentation algorithm. This approach is referred to as relative evaluation [5]. For video object segmentation, an alternative called stand-alone evaluation has been presented, in which the reference object is not available [6], [7]. In this work, we adopt the relative evaluation approach.

The methods based on relative evaluation in the literature mainly calculate the error pixel counts between the segmented object and the reference object. The errors can be categorized into two types: miss detection and false alarm. Miss detection error refers to that a pixel belonging to the object is classified as background. False alarm error means that a pixel belonging to the background is classified as foreground (object). Using the two types of classification errors, the receiver operating characteristic (ROC) curve can be generated [8]. The ROC curve shows the tradeoff between the true positive rate (TPR) and the false positive rate (FPR). The FPR is also called false alarm rate, and TPR denotes the detection rate which is equal to $1 - \text{MDR}$ (miss detection rate). The ROC has been widely used in low-level applications such as skin segmentation (detection) [9] – [17].

Most related evaluation methods mentioned above are constructed by considering image segmentation as a process of pixel labeling. Consequently, they are not appropriate for object-level evaluation. Various metrics which extend the pixel error measure have been

proposed for higher-level applications such as image segmentation at object level [18], [3], video object segmentation and tracking, and image understanding [2], [6].

Martin et al. [18] proposed an object level error measure, including global consistency error (GCE) and local consistency error (LCE). The error metrics are very useful to quantify the consistency between segmentations manually performed by different people. However, this error measure is insensitive to over (under)-segmentation; thus it is not appropriate in segmentation applications in which the exact boundaries or sizes of the fragments are important.

To attack this problem, Polak et al. presented the object-level consistency error (OCE) [3]. The OCE quantifies the discrepancy between a segmented image and the ground truth image at the object level that takes into account the existence, size, position, and shape of each fragment and penalizes both over-segmentation and under-segmentation. The OCE is suitable for specific applications in which the many small objects exist in a scene, and exact object size is critical in segmentation. The typical applications are the detection of crown canopies of trees, and segmentation of tar sands [3]. However, the OCE does not consider the false alarm error, which is essential in many higher-level applications such as tracking of objects.

All the metrics based on the relative evaluation usually treat each pixel error independent. We refer to the performance metric as pixel-based error measure. This type of error metrics does not take the spatial distribution of segmentation errors into account. The spatial distribution of segmentation error pixels will affect the shape of the segmented object. Moreover, this would yield the error of the subsequent higher level semantic processing such as tracking and/or recognition of objects.

In this paper, we first analyze the spatial distribution of the segmentation pixel errors and classify them into scattered error and region error accordingly. We then use an example to explain the limitation of the pixel-based error measure in the evaluation of object segmentation performance. Next, we propose a novel block-based error measure which assigns the

higher weight to the region error and lower weight to the scattered error. Finally, analysis of the pixel-based measure and block-based measure are presented to justify the advantages of the proposed block-based error measure.

II. ANALYSIS OF PIXEL-BASED ERROR MEASURE

The pixel-based error measure does not consider the spatial distribution of the detection error pixels. This work classifies the segmentation errors into *scattered error* and *region error* according to their distributions in two-dimensional (2-D) space. The scattered error means that the locations of the error pixels are sparse in the 2-D space, yet the *region error* refers to the type that the error pixels locate densely in a particular region.

The object segmentation is a preprocessing step in vision systems [13]. The following process after the preprocessing is often to group those pixels with similar property into an object using segmentation techniques like connected-component labeling, region growing, etc. [19]. Before segmentation, morphological operations such as erosion and dilation are usually used to remove segmentation errors. The erosion/dilation is far more powerful in removing scattered error than in removing region error. The underlying assumption of this work is that the *region error* will cause more serious destruction than the *scattered error* in a semantic sense. Therefore, the segmentation performance criterion should not consider the region error and scattered error in equal importance like the conventional pixel-based measure.

A set of hand palm images shown in Fig. 1 is used to explain our proposed idea. Fig. 1(a) is the ground truth image. Figs. 1(b) to 1(e) are the images that the ground-truth was corrupted by various errors. Figs. 1(b) and 1(d) are region errors, yet Fig. 1(c) and Fig. 1(e) are scattered errors. The pixel-based measure obtains the error pixel counts for the four cases, and the results are listed in Table 1. The error pixel count of Fig. 1(b) and Fig. 1(c) are the same (540). Similarly, the error pixel counts of Fig. 1(d) and Fig. 1(e) are the same (165) as well. However, the effects

of these two types of errors on semantic object segmentation are significantly different. We apply some morphological operations to Fig. 1(b) to Fig. 1(e), and the results are shown in Fig. 2(a) to Fig. 2(d), respectively. It is obvious that the scattered errors can be reduced significantly (see Fig. 2(b)), or be remedied (see Fig. 2(d)). However, the morphological operations do not work well for the region errors as shown in Fig. 2(a) and Fig. 2(c). For the following higher level processing, the star pattern noise in Fig. 2(a) may be regarded as another semantic object; similarly, the missed finger in Fig. 2(c) may cause the recognition error of a posture. In short, for the same number of error pixels, the effect of the region error is far more significant than the scatter error. Thus, the pixel-based error measure

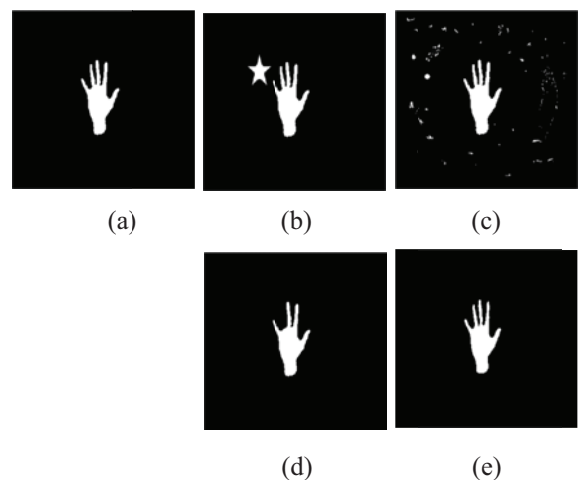


Fig. 1. Four types of errors: (a) Ground truth, (b) Region false alarm error, (c) Scattered false alarm error, (d) Region miss error, (e) Scattered miss error

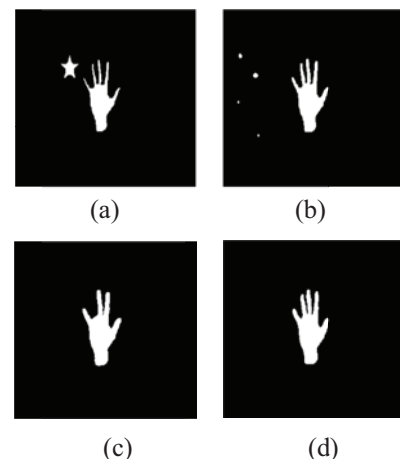


Fig. 2. Results of performing morphological operations

is not a good criterion for evaluating the performance of

Table 1. Error pixel counts for the segmentation examples in Fig. 1.

Segmentation examples	Error pixel count	Segmentation error category
Fig. 1(b)	540	Region false alarm error
Fig. 1(c)	540	Scattered false alarm error
Fig. 1(d)	165	Region miss error
Fig. 1(e)	165	Scattered miss error

object segmentation methods because it only considers the number of error pixels, but cannot distinguish *region error* from *scattered error*. We need to seek a new measure criterion that treats the *region error* and *scattered error* differently.

III. BLOCK-BASED ERROR MEASURE FOR OBJECT SEGMENTATION

As mentioned before, an error measure should be able to classify the region error and scattered error and treat the two errors differently. However, it is rather difficult to segment the area with region error since its size and shape are unknown a priori. To solve the problem, we present a novel block-based error measure which enhances the error score of the region-error block and suppresses that of the scattered error block.

The presented measure first divides foreground and background into small blocks separately and calculates the error ratio of each block. Next, it calculates the error score of each block by transforming the error ratio with a nonlinear increasing function. This function enlarges the error score for large error blocks (region-error blocks), yet reduces the error score for small error blocks (scattered-error blocks). To further enhance the importance of region error with a large area, the error score of each region-error block is further modified using block correlation. Fig. 3 shows the block diagram of our algorithm. The details of the proposed error measure are described in the following subsections.

3.1. Block Partition for Foreground and

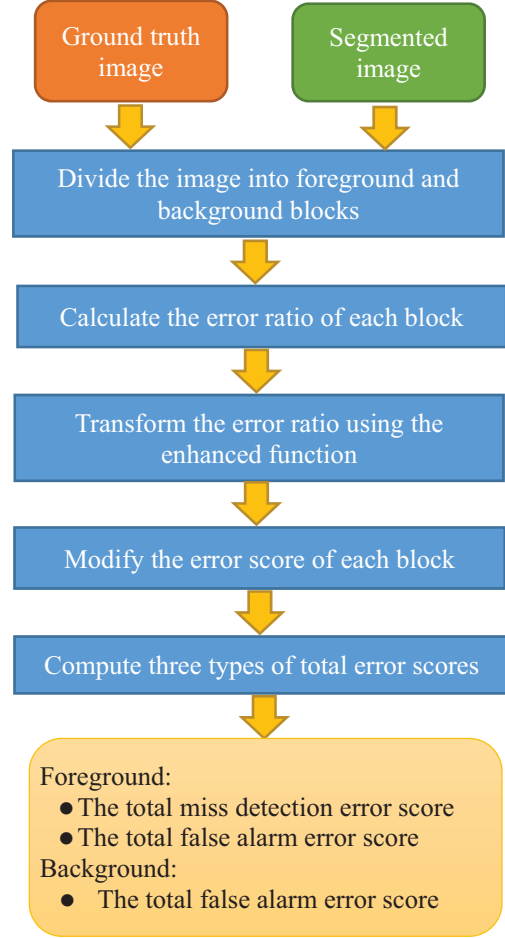


Fig. 3. Block diagram of our method

Background

Assume I_g is a reference (ground-truth) image and I_s the segmented image, both with the size of $M \times N$. The images contain foreground (object) part and background part. The background part of an image is divided into non-overlapping blocks with a size of $m \times n$, and the foreground region is partitioned into non-overlapping blocks with a size of $k \times l$. In practice, the size of a foreground block, $k \times l$, is smaller than that of a background block, $m \times n$. It indicates that the foreground block is more important than the background block in measuring errors. The foreground block is dependent upon on the smallest sub-object we considered. For example, for hand palm processing, the foreground block size is

designed approximately the same to the finger size; the maximal allowed size of a background block is limited by the size of a hand palm in an image. In our experiments, the block sizes of the foreground and background are set to be 8×6 and 16×12 , respectively.

3.2. Types of Block Segmentation Error

Assume a reference image I_g is divided into F foreground blocks and B background blocks. Thus, the image can be represented as

$$I_g = \{FB_i\} \cup \{BB_j\}, \quad i=1,2,\dots,F, j=1,2,\dots,B. \quad (1)$$

$$k \cdot l \cdot F + m \cdot n \cdot B = M \cdot N$$

A foreground block contains object pixels and non-object pixels (denoted as 1 and 0 respectively) and is represented as (refer to Fig. 4)

$$FB_i = FO_i \cup FNO_i \quad (2)$$

where,

$$FO_i = \{I_g(x, y) | (x, y) \in FB_i, I_g(x, y) = 1\}$$

$$FNO_i = \{I_g(x, y) | (x, y) \in FB_i, I_g(x, y) = 0\} \quad (3)$$

$$|FB_i| = |FO_i| + |FNO_i| = k \cdot l$$

The background block contains non-object pixels only, which is represented as

$$BNO_j = \{I_g(x, y) | (x, y) \in BB_j, I_g(x, y) = 0\}$$

$$|BB_j| = |BNO_j| = m \cdot n \quad (4)$$

Similarly, we can define the segmented image I_s as a binary image which contains the object and non-object pixels and represented by

$$I_s = SO \cup SNO \quad (5)$$

where

$$SO = \{I_s(x, y) | I_s(x, y) = 1\}$$

$$SNO = \{I_s(x, y) | I_s(x, y) = 0\}$$

In measuring segmentation errors, we superimpose the reference image and segmented image; then each foreground block is represented as (refer to Fig. 4).

$$FO_i = \{a_i\} \cup \{b_i\} \quad FNO_i = \{c_i\} \cup \{d_i\} \quad (6)$$

$$FB_i = FO_i \cup FNO_i = \{a_i\} \cup \{b_i\} \cup \{c_i\} \cup \{d_i\} \quad (7)$$

where

$\{a_i\}$: The set of miss detected part of the foreground block FB_i with some pixels of a_i ,

$\{b_i\}$: The set of correctly detected part of the foreground block FB_i with some pixels of b_i ,

$\{c_i\}$: The set of correct detected of background pixel in the foreground block FB_i with some pixels of c_i , and

$\{d_i\}$: The set of false alarm of the foreground block FB_i with some pixels of d_i .

Similarly, each background block can be represented as (refer to Fig. 5)

$$BNO_j = \{e_j\} \cup \{f_j\} \quad (8)$$

where

$\{e_j\}$: The set of correctly detected part of the background block BB_j with some pixels of e_j ,

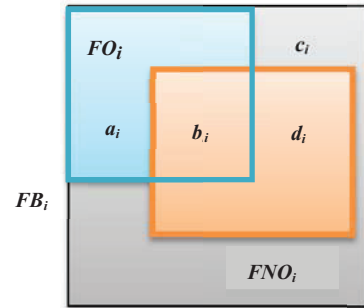


Fig. 4. Definition of segmentation error in a foreground block. The blue rectangle and orange rectangle denote the original object in I_g and segmented object in I_s , respectively

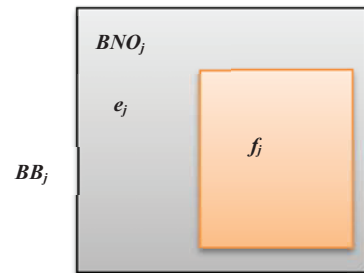


Fig. 5. Definition of segmentation error in a background block. The orange rectangle denotes the segmented background region in the background block BB_j

and

$\{f_i\}$: The set of incorrectly detected part of the background block BB_j with some pixels of f_j .

3.3. Design of Enhanced Function

Based on the block error definition above, three types of error ratios can be defined in the following.

(a) Miss detection ratio for the foreground block FB_i

$$R_{F_md}^i = \frac{N_{F_md}^i}{N_{F_t}^i} = \frac{a_i}{a_i + b_i} \quad (9)$$

(b) False alarm ratio for the foreground block FB_i

$$R_{F_fa}^i = \frac{N_{F_fa}^i}{N_{F_tn}^i} = \frac{d_i}{c_i + d_i} \quad (10)$$

(c) False alarm ratio for the background block BB_j

$$R_{B_fa}^j = \frac{N_{B_fa}^j}{N_{B_tn}^j} = \frac{f_j}{e_j + f_j} = \frac{f_j}{mn} \quad (11)$$

In general, if the error ratio (miss or false alarm) of a block is small, it implies that the possibility of scattered error is high for this block. On the other hand, if the error ratio is very high, it indicates the block suffers from region error with a high possibility; thus the error of the block should be enhanced.

However, for the case of the very little foreground, i.e., $(a + b)$ is tiny; the small a may result in large miss detection error ratio. Similarly, the small d may also yield large false alarm ratio if $(c + d)$ is small. Fig. 6 illustrates the phenomenon by showing the miss detection ratio under different object pixel ratios. It shows a small miss detection ($a = 5$) from the foreground block containing a small object $((a + b) / 48 = 0.2)$ yields the same miss detection ratio (0.6) as the large miss detection ($a = 20$) from the foreground block containing a large object $((a + b) / 48 = 0.7)$. It is obvious that a

small ground-truth object in the foreground block often generates a large miss detection ratio.

To consider the error ratio and small foreground (or background) region size

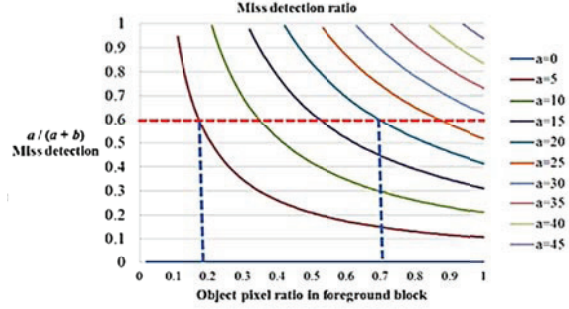


Fig. 6. The phenomenon of the miss detection ratio under different object pixel ratios

simultaneously, two requirements should be met in the design of enhanced function.

- For the tiny error ratio ($< \text{threshold } T$), the error is very likely a noise. Thus it is neglected. If the error ratio is higher than T , the error should be amplified with different factors. Generally, the higher the error ratio, the larger the amplification factor needs.
- The amplification factor should be related to the size of foreground object region $(a_i + b_i)$ or background region $(c_i + d_i)$, or f_j . The factor should be reduced to a small region. The region sizes relative to the block sizes below are used to determine the amplification factor:

$$g_{F_md}^i = \frac{a_i + b_i}{kl}, g_{F_fa}^i = \frac{c_i + d_i}{kl}, g_{B_fa}^j = \frac{f_j}{mn} \quad (12)$$

This work designs three enhanced functions for the three types of block errors defined before. The block errors in Eqs. (9) to (11) is amplified by the following transfer functions:

$$ER_k = \begin{cases} \left(\frac{R_k^i - T}{\left(1 - \frac{1}{T}\right) \cdot (g_k^i - (1+T))} \right), & 0 < T < 1, T \leq R_k^i \leq 1, g_k^i \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where $FK = \{F_md, F_fa\}$, and

$$ER_{fk}^i = \begin{cases} \left(\frac{R_{fk}^i - T}{\left(1 - \frac{1}{T}\right)(g_{fk}^i - (1+T))} \right), & 0 < t < 1, T \leq R_{fk}^i \leq 1, g_{fk}^i \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

where $B_K = \{B_{fa}\}$. The factor t is a scale to control the level of enhancement. Fig. 7 shows the enhanced curves for various values of R and g . It is seen that the larger the R and g , the output of the transfer function is higher; which obviously meets requirement above. For a small error ratio ($R < 0.1$), the error is neglected

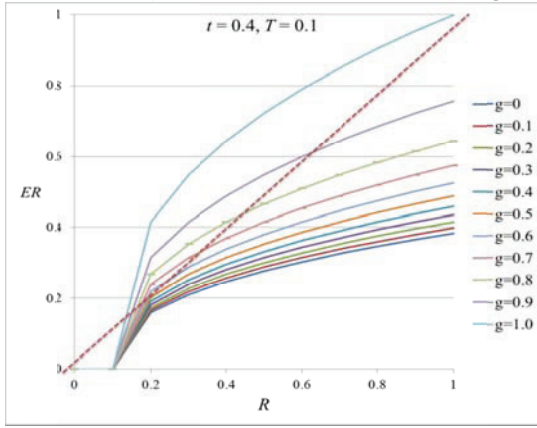


Fig. 7. The enhanced curves for various values of g . ($k \times l = 8 \times 6$, $m \times n = 16 \times 12$)

because it is easily removed by morphological operations.

3.4. Modification of Block Error Score

The error enhanced function processes each block separately. Thus, it is suitable for small region error that may contain a single region-error block. However, in some cases, a big error region which covers several blocks often exists. The significant error region will introduce more harmful effect on the following higher-level processing. Therefore, it should be further enhanced. To put the effect into measure criterion, we further propose an error modification scheme for each region-error block using block correlation as follows.

If the enhanced error score of a block is less than a threshold T , no further enhancement is required because it is not a great region error. For the block with significant error larger than T , we search its eight neighboring blocks one by one. If its error value is higher than that of the

neighbor block, the current block is further enhanced by adding the error score of that block to the current block as defined in (15) for foreground block, and (16) for background block. The first term in (15) and (16) denotes the error score of the current block, and the second term is the score contributed from the neighbor blocks. The term in the denominator is used to restrict the output error in the range of $[0, 1]$.

$$ME_{fk}^i = \frac{1}{W_{fk}^i} \left(ER_{fk}^i \cdot g_{fk}^i + \sum_{\substack{F_l \in \text{Neighbor}(F_i) \\ ER_{fk}^i \geq ER_{fk}^l > T}} ER_{fk}^l \cdot g_{fk}^l \right), \quad ER_{fk}^i > T \quad (15)$$

$$\text{where } W_{fk}^i = \sum_{F_l \in \{F_i, \text{Neighbor}(F_i)\}} g_{fk}^l.$$

$$ME_{Bk}^j = \frac{1}{W_{Bk}^j} \left(ER_{Bk}^j \cdot g_{Bk}^j + \sum_{\substack{B_l \in \text{Neighbor}(B_j) \\ ER_{Bk}^j \geq ER_{Bk}^l > T}} ER_{Bk}^l \cdot g_{Bk}^l \right), \quad ER_{Bk}^j > T \quad (16)$$

$$\text{where } W_{Bk}^j = \sum_{B_l \in \{B_j, \text{Neighbor}(B_j)\}} g_{Bk}^l.$$

We explain (15) and (16) in more details as follows. If both the current block and the neighbor block are significant, and the error score of the current block is higher than the neighbor block, these blocks may form a larger area with higher errors. Hence, the current block will be further enhanced. Eqs. (15) and (16), of course, can be replaced with other block correlation measures. However, our experiments indicate that this measure is simple and effective without requiring extra threshold parameters.

After all the blocks of a whole image are complete, we calculate the total miss detection error score, false alarm error score for the foreground, and false alarm error score for background, separately, as:

$$TE_{FK} = \frac{1}{F} \sum_{i=1}^F ME_{FK}^i, \quad (17)$$

$$TE_{Bk} = \frac{1}{B} \sum_{j=1}^B ME_{Bk}^j. \quad (18)$$

IV. COMPARISON OF PIXEL-BASED METHODS

Table 2. Number of blocks and number of neighbors for three types of areas

	Corner	Border	Center
Number of neighbors	3	5	8
Number of blocks	4	$2[(p-2) + (q-2)]$	$(p-2)(q-2)$

Any of the three types of block segmentation errors defined in the previous section may belong to the *region error* (Case 1) or *scattered error* (Case 2) defined in Section 2. Table II illustrates the six possible combinations of error types. Each type has the same error amount for the two cases. We mentioned before that *region error* is more critical in a semantic sense than *scattered error*. In the following, we will compare the conventional pixel-based error metric with the proposed block-based method error metric for the error types listed in Table 2.

4.1 False alarm error in background block

Assume background block is denoted as BB_j , $j = 1, 2, \dots, B$. Consider two cases: region error and scattered error. For a fair comparison, we assume the two cases have the same error amount $k(mn)$, $m \times n$ is the block size.

(A) Pixel-based measure

The false alarm error value can be directly written as

$$k(mn) / B(mn) = k / B \quad (19)$$

(B) Block-based measure

Case 1 : Region false alarm error

Assume the region error contains l background blocks (called error coverage hereafter). For convenience, assume the error is uniformly distributed in these blocks, and the region is a rectangle with size $p \times q$ ($p \times q = l$). Then we can assume BB_j , $j = 1, 2, \dots, l$ are background blocks containing errors, and the remaining blocks BB_j , $j = (l+1), \dots, B$ are background blocks with no error.

Thus we have

$$R_{B_fa}^j = \begin{cases} \frac{k(mn)/l}{mn} = \frac{k}{l}, & j=1, \dots, l \\ 0, & j=l+1, \dots, B \end{cases} \quad (20)$$

After enhancement of Eq. (14), we have

$$ER_{B_fa}^j = \begin{cases} \left(\frac{T(k-lT)}{(T-1)(k-l(1+T))} \right)^t = ER_{B_fa}, & j=1, \dots, l \\ 0, & j=(l+1), \dots, B \end{cases} \quad (21)$$

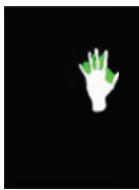
The modification scheme in Section 3.3 applies eight neighbors to further enhance the region error. The error score contribution of the neighbor blocks for each block varies with the number of neighbors of the block. The error region is a rectangle with the size of $p \times q$. The number of neighbor blocks is related to which area of the error region the current block locates. There are three types of areas including corner, border, and center areas. These areas have different numbers of neighbor blocks, and the numbers of blocks belonging to different areas are also not the same. Therefore, there are four corner blocks in the rectangle, each of which has three neighbor blocks. There are $2[(p-2) + (q-2)]$ blocks at the boundary (with corners deducted) and have five neighbor blocks, and there are $(p-2)(q-2)$ blocks with eight neighbor blocks inside the rectangle. The relation is listed in Table 3. Therefore, the average error value for an image can be derived as

$$\begin{aligned} TE_{B_fa} &= \frac{1}{B} \sum_{j=1}^B ME_{B_fa}^j = \frac{1}{l} \sum_{j=1}^l ME_{B_fa}^j \\ &= \frac{1}{l} \sum_{j=1}^l \left(\frac{1}{W_{B_fa}^j} \left(ER_{B_fa}^j \cdot g_{B_fa}^j + \sum_{\substack{BB_j \in \text{Neighbor}(BB_j) \\ ER_{B_fa}^j \geq ER_{B_fa}^j > T}} ER_{B_fa}^j \cdot g_{B_fa}^j \right) \right) \\ &= \frac{1}{l} \left(4 \cdot \frac{1}{4 \cdot \frac{k}{l}} \left(ER_{B_fa}^j \cdot \frac{k}{l} + 3 \cdot ER_{B_fa}^j \cdot \frac{k}{l} \right) \right. \\ &\quad \left. + 2((p-2) + (q-2)) \cdot \frac{1}{6 \cdot \frac{k}{l}} \left(ER_{B_fa}^j \cdot \frac{k}{l} + 5 \cdot ER_{B_fa}^j \cdot \frac{k}{l} \right) \right. \\ &\quad \left. + ((p-2) \cdot (q-2)) \cdot \frac{1}{9 \cdot \frac{k}{l}} \left(ER_{B_fa}^j \cdot \frac{k}{l} + 8 \cdot ER_{B_fa}^j \cdot \frac{k}{l} \right) \right) \\ &= \frac{1}{l} \left(p \left(ER_{B_fa}^j \cdot \frac{k}{l} \right) \right) = ER_{B_fa}^j \cdot \frac{k}{l} = \left(\frac{T(k-lT)}{(T-1)(k-l(1+T))} \right)^t \cdot \frac{k}{l} \end{aligned} \quad (22)$$

Eq. (19) indicates that the error value calculated by pixel-based error measure is

proportional to the error amount (k), but not related to the error coverage (l) (l blocks containing errors). The result can also be seen in Fig. 8, which shows the error value is constant over l when k is kept fixed. On the contrast, Eq. (22) shows that the error value computed by the proposed block-based method is determined by k and l , as demonstrated in Fig. 8. The figure indicates that the error measure raises with the increase of error amount k when error coverage (l) is fixed. The more important thing is

Table 3. Six possible combinations of error types

	Case 1	Case 2
Type 1		
Type 2		
Type 3		

Type 1: false alarm error in background block
Type 2: false alarm error in foreground block
Type 3: miss detection error in foreground block
Case 1: region error
Case 2: scattered error

that the error measure increases with the decrease of l when k is kept fixed.

By subtracting block-based error measure (Fig. 9) from pixel-based error measure (Fig. 8), we obtain the difference curve in Fig. 10. The case that the error amount (k) is small and the error coverage (l) is substantial belongs to scattered error. It is observed from the bottom-left area of the figure, the difference is negative, which indicates the scattered error is suppressed in our block-based measure. On the contrast, in the diagonal area, the difference is positive, which means the error is enhanced. This is because this area belongs to region error. Besides, for the fixed

error amount, the difference value is increasing, which means amplification factor is becoming larger, with the decreasing of the error coverage. In this case, the average number of error pixels of a block is becoming larger, which indicates the region error is becoming more compact. Thus the amplification factor is increased accordingly, as expected.

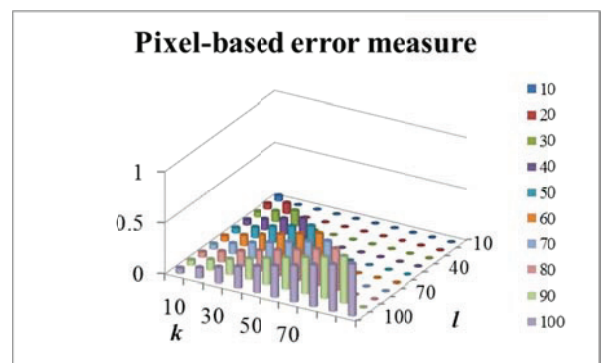


Fig. 8. The pixel-based error measure with the error amount $k(mn)$ and l blocks containing errors

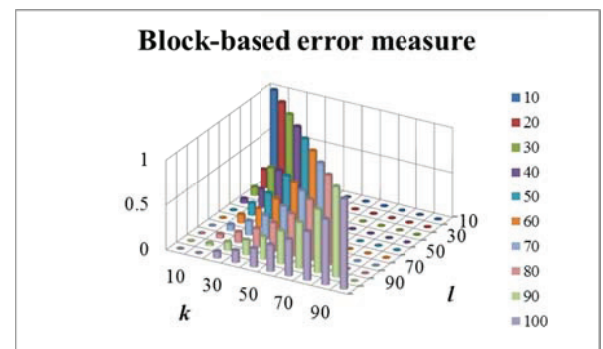


Fig. 9. The block-based error measure with the error amount $k(mn)$ and l blocks containing errors

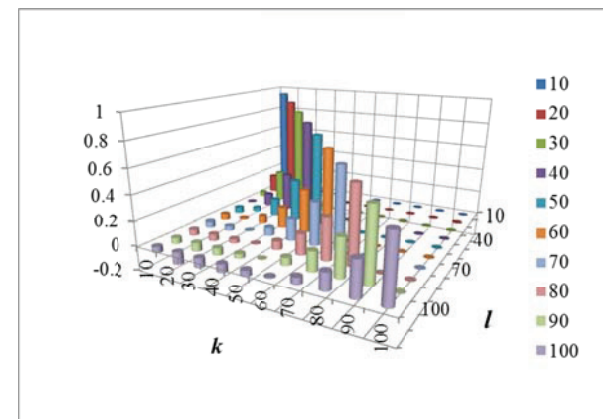


Fig. 10. The difference between the block-based error measure (Fig. 9) and the pixel-based error measure (Fig. 8)

Case 2 : Scattered false alarm error

Assume that errors are uniformly distributed in the B background blocks. Thus the error measure is simply represented as

$$R_{B_fa}^i = \frac{k(mn)}{B(mn)} = \frac{k}{B} \quad (23)$$

The result is very similar to that of the pixel-based method in Eq. (19). The minor differences are when $(k / B) < T$, the error measure is set to zero, and for small k / B , the error is suppressed by the enhanced function.

4.2 False alarm error in foreground block

Assume that all foreground blocks are denoted as $FB_j, j = 1, 2, \dots, F$. They also can be classified into two cases: region false alarm error and scattered false alarm error. Again, we assume the two cases have the same error amount $k(mn)$.

(A) Pixel-based measure

The false alarm error value can be simply written as

$$k(mn) / F(mn) = k / F \quad (24)$$

(B) Block-based measure

Case 1 : Region false alarm error

Assume the region error contains l foreground blocks. Again, assume the errors are uniformly distributed in these blocks, and the region is a rectangle with size $p \times q$ ($p \times q = l$). Then we can assume $FB_i, i = 1, 2, \dots, l$ are foreground blocks containing errors, and $FB_j, j = (l + 1), \dots, F$ are foreground blocks with no error.

$$R_{F_fa}^i = \begin{cases} \frac{k(mn)/l}{mn} = \frac{k}{l}, & i=1, \dots, l \\ 0, & i=l+1, \dots, F \end{cases} \quad (25)$$

After enhancement of Eq. (13), we obtain

$$ER_{BF_fa}^i = \begin{cases} \left(\frac{T(k-lT)}{(T-1)(k-l(1+T))} \right)^l = ER_{F_fa}, & i=1, \dots, l \\ 0, & i=(l+1), \dots, F \end{cases} \quad (26)$$

Then we apply the modification scheme in Section 3.3 to further raise the region error based on eight neighbor blocks. Using the similar analysis in Case 1 of Section 4.1, we can derive the average error value for an image as

$$\begin{aligned} TE_{F_fa} &= \frac{1}{F} \sum_{i=1}^{BF} ME_{F_fa}^i = \frac{1}{l} \sum_{i=1}^l ME_{F_fa}^i \\ &= \frac{1}{l} \sum_{i=1}^l \left(\frac{1}{W_{F_fa}^i} \left(ER_{F_fa}^i \cdot g_{F_fa}^i + \sum_{\substack{FB_j \in \text{Neighbor}(FB_i) \\ ER_{F_fa}^j \geq ER_{F_fa}^i > T}} ER_{F_fa}^j \cdot g_{F_fa}^j \right) \right) \quad (27) \\ &= \frac{1}{l} \sum_{i=1}^l \left(\frac{4 \cdot \frac{1}{k} \left(ER_{F_fa}^i \cdot \frac{k}{l} + 3 \cdot ER_{F_fa}^i \cdot \frac{k}{l} \right)}{4 \cdot \frac{k}{l}} \right) \\ &\quad + 2 \left((p-2) + (q-2) \right) \cdot \frac{1}{6 \cdot \frac{k}{l}} \left(ER_{F_fa}^i \cdot \frac{k}{l} + 5 \cdot ER_{F_fa}^i \cdot \frac{k}{l} \right) \\ &\quad + \left((p-2) \cdot (q-2) \right) \cdot \frac{1}{9 \cdot \frac{k}{l}} \left(ER_{F_fa}^i \cdot \frac{k}{l} + 8 \cdot ER_{F_fa}^i \cdot \frac{k}{l} \right) \\ &= \frac{1}{l} \left(pq \left(ER_{F_fa}^i \cdot \frac{k}{l} \right) \right) = ER_{F_fa}^i \cdot \frac{k}{l} = \left(\frac{T(k-lT)}{(T-1)(k-l(1+T))} \right)^l \cdot \frac{k}{l} \end{aligned}$$

Case 2 : Scattered false alarm error

Assume that errors are uniformly distributed in the F background blocks. Thus the error measure is simply represented as

$$R_{F_fa}^i = \frac{k(mn)}{F(mn)} = \frac{k}{F} \quad (28)$$

The result is also similar to that of the pixel-based method in Eq. (24).

The miss detection error in foreground block can be analyzed in a way similar to the procedure above, and the result is also similar. Therefore we neglect the analysis here.

V. CONCLUSIONS

In this paper, we investigated the segmentation error of object segmentation and classified it into *scattered error* and *region error*. We concluded that the *region error* is more critical than the *scattered error* for segmentation since it significantly affects the performance of the following higher-level processing modules. We developed a block-based error measure in which the *region error* is enhanced, but the *scattered error* is suppressed. Mathematical analysis proved that the block-based error measure achieves expected a result and it is more reasonable than the conventional

pixel-based error measure in the semantic scene.

ACKNOWLEDGEMENTS

This work was supported in part by Ministry of Science and Technology granted MOST 104-2221-E-130 -014 -MY2.

REFERENCES

- [1] Shantaiya, S., Verma, K., and Mehta, K., "A Survey on Approaches of Object Detection," *International Journal of Computer Applications*, Vol. 65, No. 18, pp.14-20, 2013.
- [2] Kim, C., and Hwang, J. N., "Fast and Automatic Video Object Segmentation and Tracking for Content-Based Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, No. 2, pp.122-129, 2002.
- [3] Polak, M., Zhang, H., and Pi, M. H., "An Evaluation Metric for Image Segmentation of Multiple Objects," *Image and Vision Computing*, Vol. 27, pp. 1223-1227, 2009.
- [4] Erdem, C. E., Tekalp, A. M., and Sankur, B., "Metrics for Performance Evaluation of Video Object Segmentation and Tracking without Ground-Truth," *Proc. of IEEE International Conference on Image Processing*, Vol. 2, pp. 69-72, 2010.
- [5] Correia, P. L., and Pereira, F., "Objective Evaluation of Video Segmentation Quality," *IEEE Transactions on Image Processing*, Vol. 12, No. 2, pp. 186-200, 2003.
- [6] Hemery, B., Laurent, H., and Rosenberger, C., "Evaluation Metric for Image Understanding," *Proc. of the International Conference on Image Processing*, pp. 4381-4384, 2009.
- [7] Correia, P. L., and Pereira, F., "Stand-Alone Objective Segmentation Quality Evaluation," *EURASIP Journal on Applied Signal Processing*, Vol. 2002, Issue 4, pp. 389-400, 2002.
- [8] Zhou, X. H., Obuchowski, N. A., and McClish, D. K., *Statistical Methods in Diagnostic Medicine*, New York, NY: Wiley & Sons, 2002.
- [9] Brand, J., and Mason, J. S., "A Comparative Assessment of Three Approaches to Pixel Level Human Skin-Detection," *Proc. of the International Conference on Pattern Recognition*, pp. 1056-1059, 2000.
- [10] Sebe, N., Cohen, I., Huang, T. S., and Gevers, T., "Skin Detection: A Bayesian Network Approach," *Proc. of IEEE International Conference on Pattern Recognition*, Vol. 2, pp. 903-90, 2004.
- [11] Jedynek, B., Zheng, H., and Daoudi, M., "Statistical Models for Skin Detection," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, Vol. 8, pp. 92-97, 2003.
- [12] Phung, S. L., Bouzerdoum, A., and Chai Sr, D., "Skin Segmentation Using Color Pixel Classification: Analysis and Comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 1, pp. 148-154, 2005.
- [13] Ramamoorthy, A., Vaswani, N., Chaudhury, S., and Banerjee, S., "Recognition of Dynamic Hand Gestures," *Pattern Recognition*, Vol. 36, No. 9, pp. 2069-2081, 2003.
- [14] Jones, M. J., and Rehg, J. M., "Statistical Color Models with Application to Skin Detection," *International Journal of Computer Vision*, Vol. 46, No. 1, pp. 81-96, 2002.
- [15] Yang, J., Lu, W., and Waibel, A., "Skin-Color Modeling and Adaptation," *Proc. of Asian Conference on Computer Vision*, pp. 687-694, January 1998.
- [16] Zhu, X., Yang, L., and Waibel, A., "Segmenting Hands of Arbitrary Color," *Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 446-453, 2000.
- [17] Sigal, L., Sclaroff, S., and Athitsos, V., "Skin Color-Based Video Segmentation under Time-Varying Illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 7, pp. 862-877, 2004.
- [18] Martin, D., Fowlkes, C., Tal, D., and Malik, J., "A Database of Human Segmented Natural Images and Its

Application to Evaluating Algorithms and Measuring Ecological Statistics,” Proc. of the International Conference on Computer Vision, pp. 416–42, 2001.

- [19] Gonzalez, R. C., and Woods, R. E., Digital Image Processing, 3rd ed., Prentice Hall Inc., 2008.