SEGMENTATION EVALUATION
USING ULTIMATE MEASUREMENT ACCURACY

Yu Jin ZHANG and Jan J. GERBRANDS
Information Theory Group, Department of Electrical Engineering
Delft University of Technology, 2600 GA Delft, The Netherlands

ABSTRACT

As a wide range of segmentation techniques have been developed in the last two decades, the evaluation and comparison of segmentation techniques becomes indispensable. In this paper, after a thorough review of previous work, we present a general approach for evaluation and comparison of segmentation techniques. More specifically, under this general framework, we propose to use the ultimate measurement accuracy (UMA) to assess the performance of different algorithms. In image analysis, the ultimate goals of segmentation and other processing are often to obtain measurements of the object features in the image. So, the accuracy of those ultimate measurements over segmented images would be a good index revealing the performance of segmentation techniques. We feel this measure is of much greater importance than, *e.g.*, error probabilities on pixel labeling, or even specially developed figures of merit.

There exist many features describing the properties of the objects in the image. Some of them are discussed here and their applicability and performance in the context of segmentation evaluation are studied. Based on experimental results, we provide some useful guidelines for choosing specific measurements for different evaluation situations, and for selecting adequate techniques in particular segmentation applications.

I. INTRODUCTION

Image segmentation is the process that subdivides an image into its constituent parts and extracts those parts of interest. It is one of the most important tasks in automated image analysis since it is in this critical step that the entities of interest are extracted from an image for subsequent processing. It is obvious that the segmentation results will affect all the following tasks.

Much effort has been devoted to the segmentation problem in the last years. This has already resulted in several hundreds of different segmentation techniques, and this number is still increasing at a rate of dozens per year.

One important fact in the development of segmentation techniques is that no general theory about segmentation exists, so this development has traditionally been an *ad hoc* process. As a result, a wide variety of segmentation algorithms have appeared in literatures. Several survey papers have been published [1, 2, 3].

Although a large number of segmentation algorithms have been proposed, none of them is universally applicable. Many of them are empirically developed and problem-oriented. The performance evaluation and comparison between competing segmentation techniques becomes indispensable for choosing appropriate techniques in specific applications. On the other side, the performance evaluation and comparison is also useful for providing guidelines to make refinements of existing techniques and to help new developments.

Relatively less effort has been devoted to the evaluation and comparison of segmentation techniques than to their development. Among previous papers in this direction, most of them concentrate on edge detection, while few address the techniques for the whole range of segmentation [3, 4, 5, 6, 7]. In a wide sense, detecting edges in an image can be considered as a first step toward separating the image into its components. In the context of automatic image analysis, however, it is also necessary to extract the interest parts, while only detecting edges is not enough. From this point of view, evaluating an edge detection algorithms is not equal to evaluating a complete segmentation technique. In this paper, we consider the segmentation techniques which can really divide images into two parts, object for interest part and background for no-interest part, as well as can extract the object from the images. And we concentrate on the evaluation of these techniques.

The rest part of this paper is organized as follows. In the next section, we give an overlook of previous work in segmentation evaluation and comparison. We will also point out some open questions, because it is the imperfections of these methods that has lead to the study of new methods for segmentation evaluation. Our new approaches for segmentation evaluation will be presented in section III, where the general evaluation system, and especially the new criteria for performance assessment will be described. In section IV, we discuss several object features which have the use in the assessment. Their evaluation abilities are evaluated, and their behavior in different situations are compared. Finally, the advantages of our new approach are summarized and potential further directions are indicated in section V.

II. PREVIOUS WORKS AND OPEN QUESTIONS

In the following, short descriptions of the methods for the evaluation of the complete segmentation techniques are given, with special attention for the criteria used to assess the performance of segmentation algorithms.

Image segmentation is often considered as a pixel classification process. Based on this idea, Yasnoff *et al* [4] proposed to take the number of mis-classified pixels and their position into account for the calculation of two error measures. They used these error measures to assess the performance of algorithms and also studied their correlation with human observation. The first error is self explaining. The sum of the distances between pixels that have been incorrectly assigned to a particular class and the nearest pixels that actually belong to that class is taken as the second criterion (up to a proportionality coefficient). Similar ideas are reflected in Abdou *et al*'s figure of merit (FOM) for edge detection evaluation [8]. Some problems of FOM are discussed by Heyden [9]. We give here some examples for Yasnoff *et al*'s error measures. In fact, many configurations can be found for which the same values of distance measure are obtained, which are quite diverged from each other and may yield different consequences. Some examples are depicted in Fig.1, where the distance measures for (A), (B) and (C) are equal (the numbers of mis-classified pixels are also equal). Without further processing, the three mis-classified pixels in Fig.1(A) will be counted as object pixels while the three mis-classified pixels in Fig.1(B) may have no influence on the size or shape of the object. Moreover, it is evident that the distance measure cannot, for example, distinguish several isolated mis-classified pixels (Fig.1(C)) with a cluster of mis-classified pixels (Fig.1(A) and (B)) although the latter is often more difficult to be treated than the former in practice. In addition, as noted by the original authors, the computation of mis-classified pixel distance does not include any consideration of the shape of object in images. The shape of object, however, plays often an important role in image analysis. Finally, one should note that object pixels mis-classified as background pixels and background pixels mis-classified as object pixels may be of different significance.

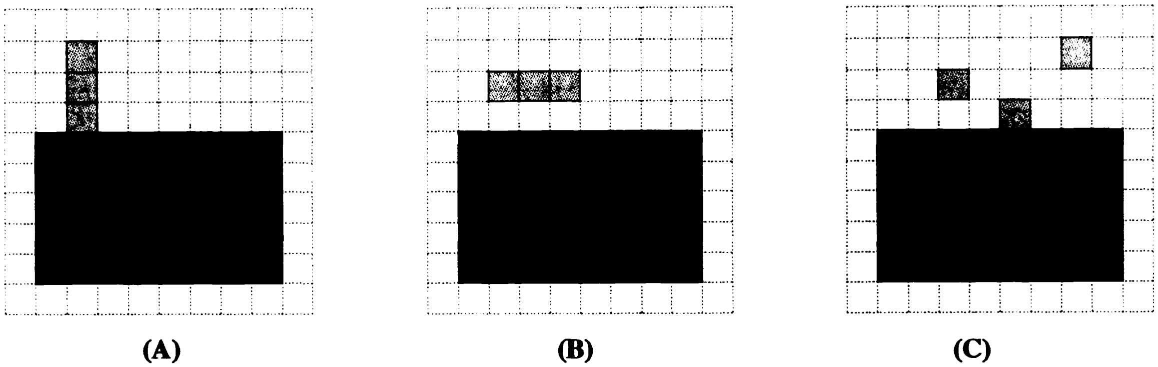


Fig.1 Various example configurations that have equal distance measures and also the same number of misclassified pixels.

Weszka *et al*[5] proposed two "goodness" measures for selecting and evaluating a segmentation threshold. One measure is the discrepancy between the original image and the segmented image, which is actually measured by the classification error. The other is the so-called busyness of the segmented image. The idea behind this measure is the intuition that "we normally want thresholded images to look smooth rather than busy"[5]. Its calculation is based on the gray level co-occurrence matrix whose entries are estimates for the joint probabilities of gray levels for adjacent pixels. Although these two measures are different in appearance, they frequently lead to the selection of the same or very similar threshold values, as remarked by the original authors. One should note here that if one uses these measures for selecting threshold, it is necessary to calculate these measures for all possible gray levels to select the appropriate threshold values. This may be quite time consuming. Furthermore, it seems that other techniques than thresholding can not be evaluated by using this method as the threshold values are needed in the calculation of these measures.

In the context of developing a rule based system for image segmentation using the partition algorithms, Levine *et al*[6] proposed to take the intra-region uniformity (which is similar to Weszka *et al*'s busyness measure[5]), and inter-region contrast into account as performance parameters for the system in segmenting an image into regions. They assume good segmentation should partition the image into regions, some features should be uniform inside each region and should be distinct between adjacent regions. These criteria are also the "goodness" criteria in principal. The authors claimed that by using these criteria, they eliminated the requirements of the knowledge of correct segmentation as required by the methods mentioned above. When there is no *a prior* knowledge about the number of regions in an image, however, confusion can still arise, a "good" segmentation determined by these measures does not necessarily correspond to reality. For example, there is always a transition region between object and background regions in real images[10]. Using the region uniformity criterion, the "better" segmentation (in the sense of more uniformity for each region, *e.g.*, minimum variance within each region) would segment the image into three parts: object, background and transition regions. While in practice, what we seek is to separate objects from the rest of the images, *i.e.*, segment the image into two parts.

Sahoo *et al*[3] also used two measures for the evaluation of thresholding techniques. One is the uniformity criterion adapted from Levine *et al*[6]. Another one is a new measure that they called shape measure. They calculated this measure by the summation of the gradient values over pixels which have gray level higher than the average gray level of its neighbors and also higher than the selected threshold value, as well as pixels which have gray level lower than the average gray level of its neighbors and also lower than the selected threshold value. From this summation, the gradient values over pixels which have gray level higher than the average gray level of its neighbors but lower than the selected threshold value, as well as pixels which have gray level lower than the average gray level of its neighbors but higher than the selected threshold value should be subtracted. Except the vague relation of this measure to object shape, it is evident that this measure is very sensitive to noise, especially at isolated noisy pixels, as they are heavily weighted in the calculation of the shape measure. An interest thing is that Sahoo *et al* also incorporate some visual information to supplement their quantitative evaluation. Some bad ranked segmentation algorithms, according to their two measures, however, apparently provide good binary images in terms of retaining object details. Finally, the involvement of the threshold value in the calculation makes this method only applicable to threshold techniques.

Recently, a comparative study of several histogram based thresholding techniques has been made by Lee *et al*[7]. Although the techniques and images used are different than in other studies, the criteria they utilized are not novel. In fact, the two measures used by Sahoo *et al*[3] and the probability of classification error[4, 5] are employed.

There also exist some papers in the literature, in which after the developing of new segmentation algorithms, a comparative study is made by showing pictures of improved results. These subjective and qualitative human inspection based procedures are not suitable for evaluating large number of images in real applications or for accurate comparison of competing algorithms. Moreover, these methods are often difficult to be applied for a wide range of distinctive segmentation techniques.

Despite of all those efforts, the development of methods for segmentation evaluation is far to its end. Except that each proposed method has its own limitation as discussed above, there are still some open problems which are common for several methods.

As a consequence of the lack of a general segmentation theory, there is no unique judgement of segmentation results. The goodness criteria as defined and used in previous works can, in the best case, only provide some necessary conditions of "good" segmentations. In addition, they are directly related to subjective human quality judgements which often differs from objective measurement. It is probable that a high score of such goodness measure does not correspond to the actual situation. A "pleasing" picture is not necessary reflecting an accurate segmentation.

Some proposed evaluation criteria coincide with the segmentation criteria used inside some particular segmentation algorithms. It is evident these algorithms would be highly ranked if we use these criteria for evaluation. This will introduce some bias when comparing these algorithms with other criteria based algorithms. For example, some algorithms are uniformity oriented in nature as indicated by Sahoo *et al*[3]. If the uniformity criterion is used in evaluation, these algorithms would have more chances showing "better" performance than others. The ranking order may be changed, however, if other criteria are used.

Some methods need human inspection to provide reference. The measures were defined and evaluated in terms of their correlation with human observation. Contrast to certain image processing tasks like, *e.g.*, image enhancement, objective judgements are mandatory in image analysis.

Finally, to show the effectiveness of proposed methods, very often experiment examples are supplied. In all of these works[3, 4, 5, 6, 7] only real images have been used. From one side, these images have the advantage of "reality". Their "random" nature, however, makes the evaluation results not viable for different applications. Moreover, many indeterminate characteristics of real images make an analytic evaluation not possible and an accurate comparison not feasible because various phenomena are mixed, and it is difficult to study them individually. Another problem related to the use of real images is that the reference segmentations should be obtained by hand. Such subjective and imprecise procedure is not appropriate for a quantitative evaluation tasks.

III. NEW APPROACHES

3.1 New criteria for performance assessment

From the discussion in the above section, we know that the appropriate choice of criteria for performance assessment is important and critical. Instead of "goodness" criteria which are related to the human intuition and observation, more objective and quantitative criteria for studying segmentation algorithms are needed. In image analysis, the ultimate goals of segmentation and other processing are often to obtain measurements of the object features in the image (in fact, the quantitative description of objects in an image should be also based on these measurements). The accuracy of those ultimate measurements over segmented images should be of great importance and would be a good index revealing the performance of applied segmentation algorithms. So, we propose to use the Ultimate Measurement Accuracy (UMA) of different object features as the criteria in judging and ranking the performance of segmentation techniques.

The comparative nature of UMA implies some references should be available and this will be discussed in the following subsections. If we denote O_m as the original measurement value from the reference image and R_m as the real measurement value from the segmented image, two types of UMA can be distinguished, *i.e.*, absolute UMA and relative UMA:

$$UMA_{abs} = | O_m - R_m | \tag{1}$$

$$UMA_{rel} = \frac{| O_m - R_m |}{O_m} * 100\% \tag{2}$$

From (2), it seems the UMA_{rel} may be possibly bigger than 100% in certain cases. In practice, we can just set it to 100% as it doesn't made sense to distinguish between such large errors. The calculation of UMA is easy. The calculation of R_m is part of the image analysis task we need to do, and O_m can be obtained by a similar process from reference images. Various object features can be used in UMA calculation so different applications can be covered. The UMA provides a direct index regarding the performance of applied segmentation techniques. In other words, in image analysis, what we are most interested in is to obtain an accurate measurement of object features and UMA is just a measure of it. From this point of view,the UMA should also be the final judgement of a "good" segmentation. Moreover, it is precise as we can calculate it quantitatively, and has no subjective bias as no human observation or segmentation criteria are involved.

3.2 General framework for segmentation evaluation

The choice of criteria for performance assessment is associated with and also determines the segmentation evaluation process. In concrete words, the assessment should be incorporated into a framework for segmentation evaluation. The general framework we propose for segmentation evaluation consists of three related essential components, as depicted in Fig.2. One is the performance assessment in which UMA will be calculated according to formula (1) and (2). The second is the image generation which provides reference images and original measurements. And the last one is the test procedure which produces real measurements from segmented images. The general system works like this: according to the goal of segmentation, the performance assessment block determines which object features should be used and makes requirement to the image generation block. Considering also the real situation (actual application domain), appropriate test images are generated. These images will be segmented in the test procedure block by available algorithms and the pre-selected object features will be measured from the segmented images. Both original and real measurements are gathered in the performance assessment block from which the evaluation results are obtained.

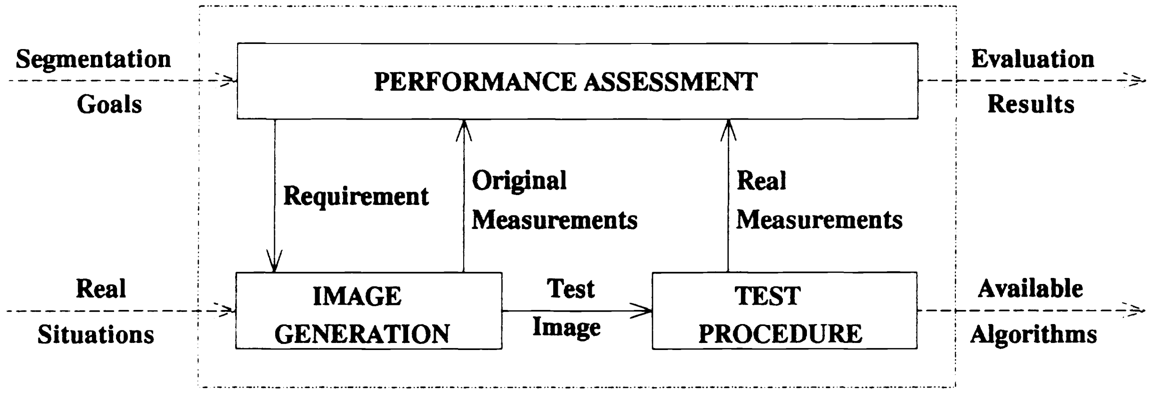


Fig.2 General framework for segmentation evaluation.

3.3 Detailed descriptions

Now we consider the detailed diagram of our segmentation evaluation system that is shown in Fig.3. From the calculation of UMA using formula (1) and (2), we see a requirement appears, *i.e.*, we need know the original measurement values of the interest object in images. One solution is to generate synthetic test images. When using synthetic images, the above requirement can be easily fulfilled. The original measurement values can be accurately determined from syntheticly generated images. Besides, synthetic images are easy to manipulate and to simulate different phenomena. Image generation should not only provide images for testing but also provide images reflecting real situations. Here consists of four parts [11]. The first is to generate a basic image, which can be a simple model of the real images under consideration. This makes all generated images to have a common basis for comparison. In "object variation", the object in the basic image is modified to provide images with a variety of object sizes, shapes, ..., *etc*. Different corruption factors, such as noise, blurring, are simulated, and combined into images with a variety of objects to form different test image sets reflecting different real situations. Before the combination of various corruption factors, the original measurements of object features can be accurately obtained from images with a variety of objects, and these original measurements will be send to "UMA calculation" as references.

The test procedure is a standard image analysis procedure. Its input is an image and its output is the measurement of some features of the object in the image. Here the generated images are first segmented by the algorithms under evaluation. Some appropriate pre- and/or post-processing may be necessary in this step. The interest objects can be extracted from segmented images. The real measurements of object features are then obtained, and will be used in the "UMA calculation".

With the original and real measurement values, the UMA can be calculated. According to the value of UMA, the performance of different segmentation algorithms can be evaluated and these algorithms can be thus ranked. Note here that the type of images generated can be supervised according to the object features being selected, *i.e.*, various images can be designed and generated to meet the conditions required by the use of different object features (one example will be given in the next section). Moreover, the performance evaluation results can be fed back to the testing algorithms for the purpose of modification and refinement.

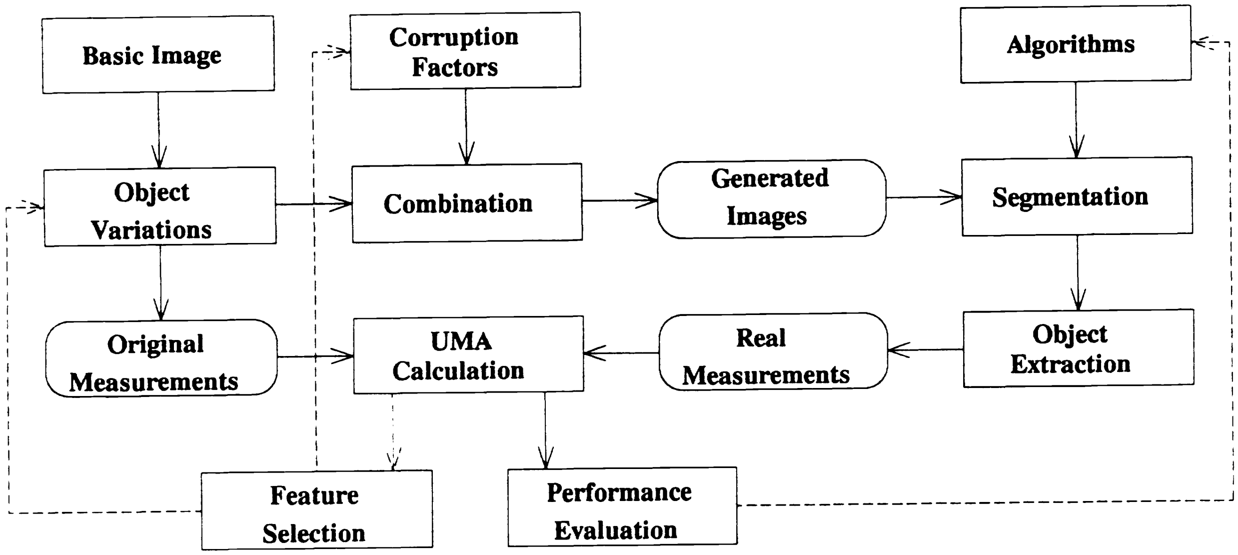


Fig.3 Detailed diagram of segmentation evaluation system.

IV. OBJECT FEATURE STUDY

4.1 Object features

In cell biology and other applications, geometric features are very popular to describe objects. Some examples of geometric features will be considered here like the area and perimeter of objects, as well as so called shape features such as form (shape) factor, sphericity and eccentricity which characterize the contour shape of objects.

The form factor p2a is derived from area and perimeter:

$$\mathbf{p2a} = \frac{\mathbf{perimeter}^2}{\mathbf{4 * \pi * area}} \tag{3}$$

The sphericity is defined as:

$$\mathbf{sph} = \frac{\textbf{radius of inscribed circle}}{\textbf{radius of circumscribed circle}} \tag{4}$$

The eccentricity is obtained on the basis of fitting an ellipse to the object boundary:

$$\mathbf{ecc} = \frac{\textbf{major axis of fitting ellipse}}{\textbf{minor axis of fitting ellipse}} \tag{5}$$

These five features have different capability and behavior in the evaluation of segmentation techniques. This can be illustrated by the following experiments.

4.2 Method

The principal goal of this study is to show the different behavior of these features when used in UMA calculation for assessing the performance of different algorithms. The procedure for studying the behavior of different object features can be similar to that for comparing different segmentation techniques [12]. Now we do not vary over the segmentation techniques, while the various object features in the calculation of UMA are compared.

For this study, one set of synthetic images is generated using the system for test image generation [11]. We use images of size 256*256, with 256 grey levels. The basic image is composed of a centered circular disc object (diameter 128) with grey level 144 on a homogeneous background of grey level 112. Since most of the features under consideration are related to the shape of objects, a set of objects with different shapes (eccentricity 1.0, 1.5, 2.0 and 2.5) are designed. We try to keep the area of those objects to be constant (see table 1). The transition region [10] is produced by filtering the noise-free images with a 3*3 uniform filter. The noise effect is produced by adding zero-mean gaussian noise to the noise-free images. The standard deviations of the gaussian noise were varied over 4, 8, 16, 32. Because the contrast between object and background in the noise-free image is 32, the Signal-contrast-to-Noise Ratios (SNR) of the images correspond to 64, 16, 4, and 1.

Table 1. Object variation of generated images.

object #	1	2	3	4
eccentricity	1.0	1.5	2.0	2.5
major axis length (pixel)	64	78	90	101
minor axis length (pixel)	64	52	45	40
area of object ($pixel^2$)	12868	12742	12723	12692

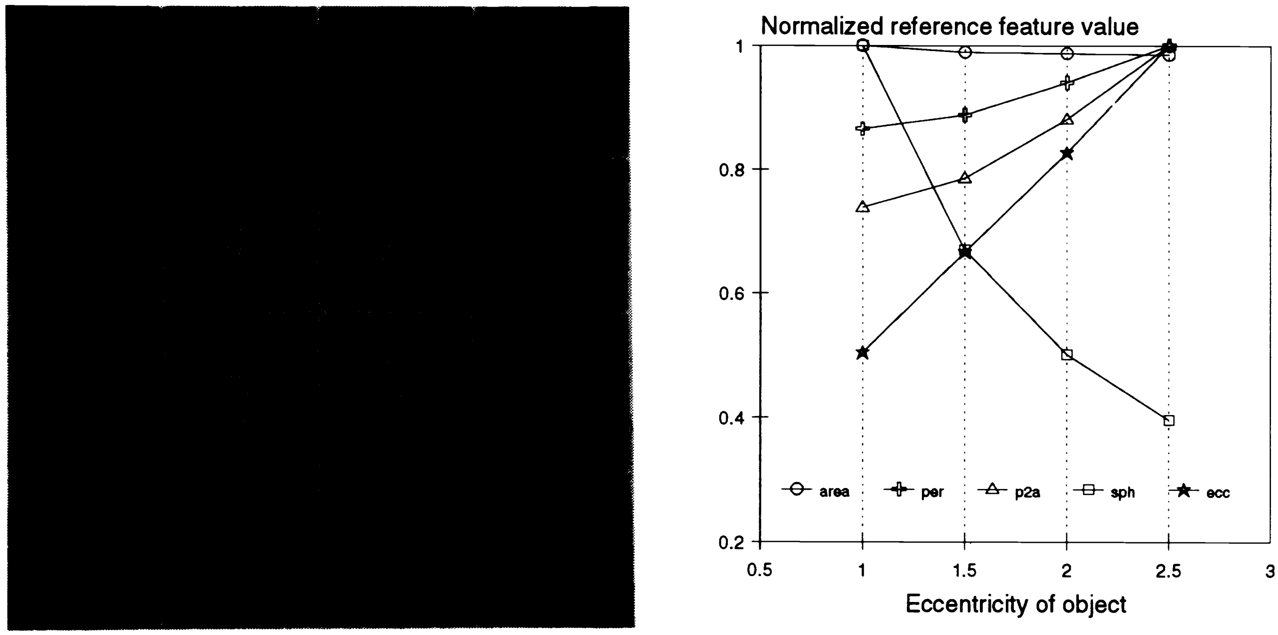
The images thus generated are shown in Fig.4(A). Along the horizontal axis, we have images with object #1 to #4. The vertical axis presents the SNR. We have images of, from bottom to top, SNR = 1, 4, 16 and 64. The normalized values of area, perimeter, form factor, sphericity and eccentricity for these objects are shown in Fig.4(B). Except area, all other features have monotone and notable change along the object axis.

These generated images are segmented by thresholding them with all threshold values between the original grey levels of object and background. After thresholding, the biggest object is chosen and the holes inside of it are filled. One opening step is then carried out for the selected object. To cope with the random nature of noise, for each SNR level, ten samples of noise are generated, so ten measurements are made. The average value of the ten measurements is taken as the final one. In the presentation of results, the horizontal axis shows the threshold value, while the vertical axis is the measurement accuracy.

4.3 Results and discussions

The following three experiments are made.

(1) The first experiment is to look at the behavior of the features under the same image condition, *i.e.*, the same object and same SNR. As an example, we show the results of the five features for the image with object #3 and SNR 16 in Fig.5.



(A) (B)

Fig.4 (A) One set of generated images;
(B) Normalized reference feature values of different objects in generated images.

From Fig.5, we see that all feature curves have a valley located at some threshold value between the gray levels of object and background. This means that UMA (calculated for all these five features) does exhibit the ability to distinguish the differently segmented images. Two deviations of these curves, however, can be marked from Fig.5. The first is that the valleys are not located at the same value. This means using different features, different optimal threshold values (in the sense of minimum UMA) will be reported. The second is that the curves have different forms. This implies that the behavior of different features is different, too. In other words, for the same images, these features show different behavior. For example, their ability to test a small change in the parameters of a segmentation technique is different. The area curve has quite consistent decrease and increase regions, *i.e.*, it steadily follows the change of the threshold value, while others are not always so sensitive to this change. From this point of view, area is a good candidate feature to be used in UMA calculation. Among the others, eccentricity is relatively better than sphericity, followed by p2a.

(2) The second experiment is to look at the behavior of the features in association with different object shapes. For this purpose, we calculate the UMA for images having different objects but with the same SNR (along the horizontal axis in Fig.4(A)). The results for sphericity and eccentricity are shown in Fig.6 (A) and (B), respectively.

Comparing Fig.6(A) with (B), we see that sphericity and eccentricity have different sensitivity with respect to object shape variations. When segmenting images with more elliptical objects, the expected accuracy for measuring sphericity becomes better while for measuring eccentricity becomes worse. We have also found that the perimeter and p2a curves exhibit similar tendency as the eccentricity curve but less noticeable. It is clear that this factor must be taken into account when using them for images with objects of different shapes.

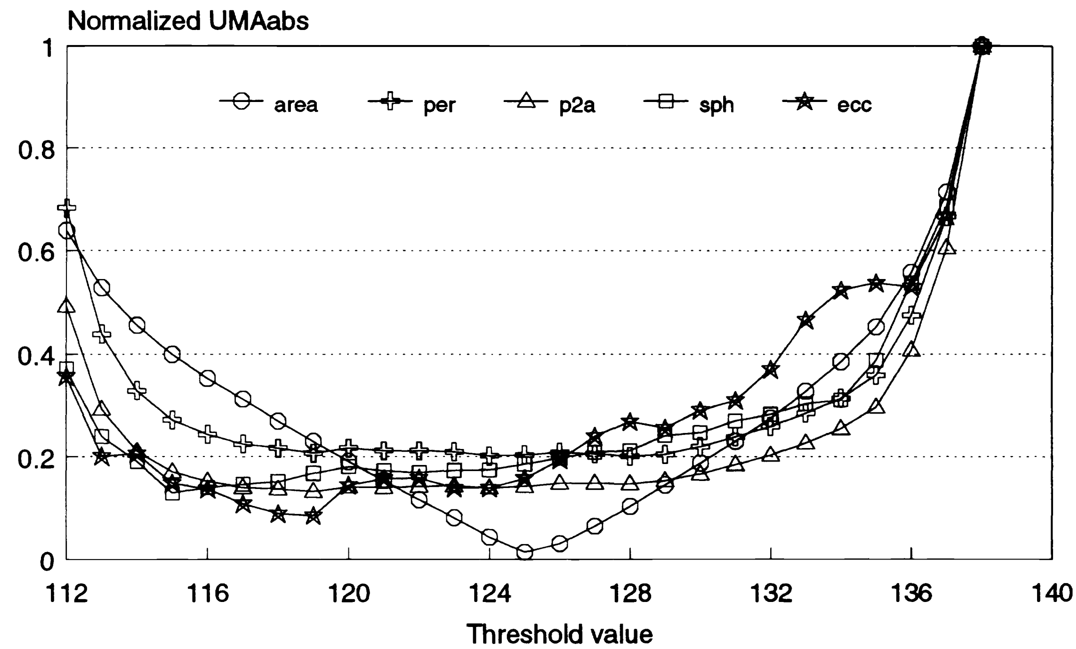


Fig.5 The results of five object features for the image with object #3 and SNR 16.

(3) The third experiment is to look at the behavior of the features in relation with different value of the SNR. The results for images with the same object but different SNR (along the vertical axis of Fig.4(A)) are collected. We present the curves of area and p2a separately in Fig.7 (A) and (B).

It is easy to be seen from Fig.7 that the influence of noise to the object features is quite different. The area curves are shifted to the left with the decrease of SNR, but they keep similar forms. It means its ability to test a small change of parameter values in the segmentation technique is relatively independent of the SNR. The p2a curves are monotonically going up with the decrease of SNR. In addition, the curve forms are also changed, from relative plat for the higher SNR case to somehow sharp for the lower SNR case. Both of them imply that the UMA of p2a is quite sensitive to the SNR of images. We have found that sphericity and eccentricity curves have similar behavior as p2a in this case.

V. CONCLUDING REMARKS

In this paper, we proposed to use UMA as criteria for assessing the performance of segmentation techniques. We also presented a general framework by which the UMA of various object features, using different algorithms and under diverse situations can be obtained and used for segmentation evaluation. The advantages are multifold. First is its generality, suitable for a wide range of techniques and real applications. Secondly, the measurements used in the evaluation are just what we want to measure in actual image analysis tasks. This goal-oriented approach is accurate and quantitative. Thirdly, there is no subjective bias by using UMA. In fact, the UMA is of much greater importance than, *e.g.*, error probability on pixel labeling, where accuracy is very context dependent, or some specially developed figures of merit, which are often subjective and often lead to cases where the results do not agree well with reality. Finally, our approach is based on a comparative measure with an exact reference. As we know "best" achievable results, absolute ranking of all algorithms becomes possible.

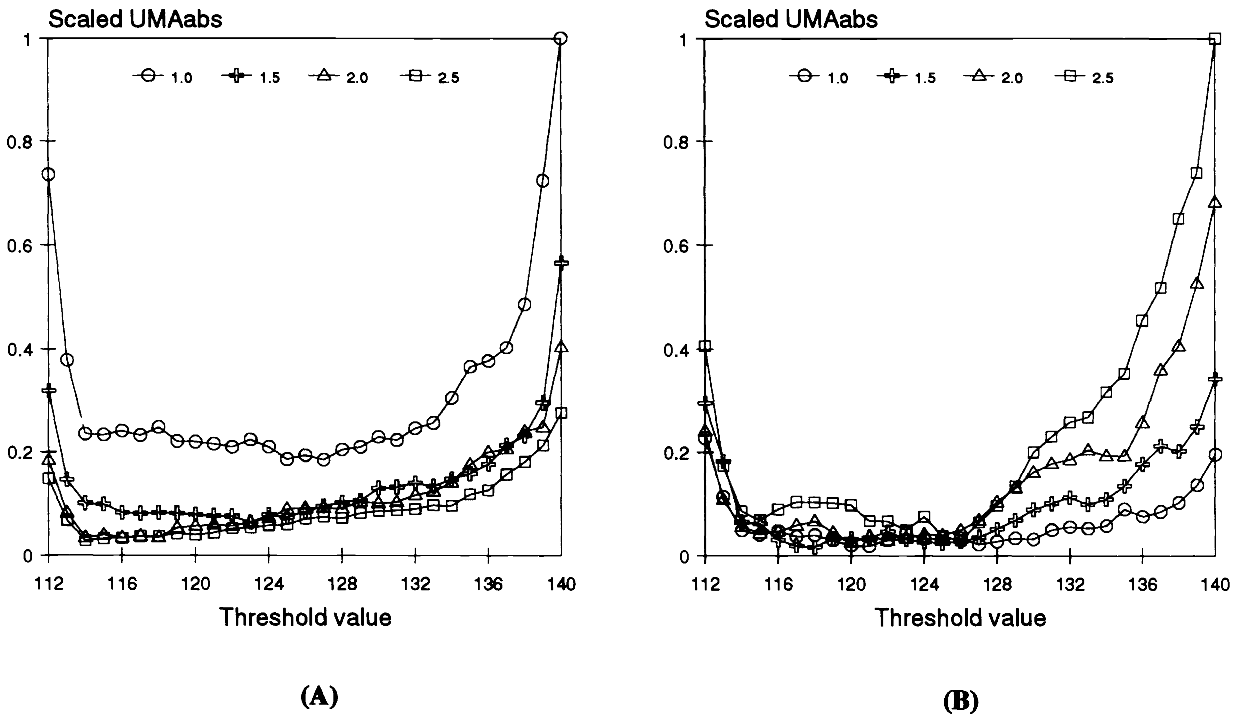


Fig.6 The results for (A) sphericity, (B) eccentricity with images having various objects and SNR 64.

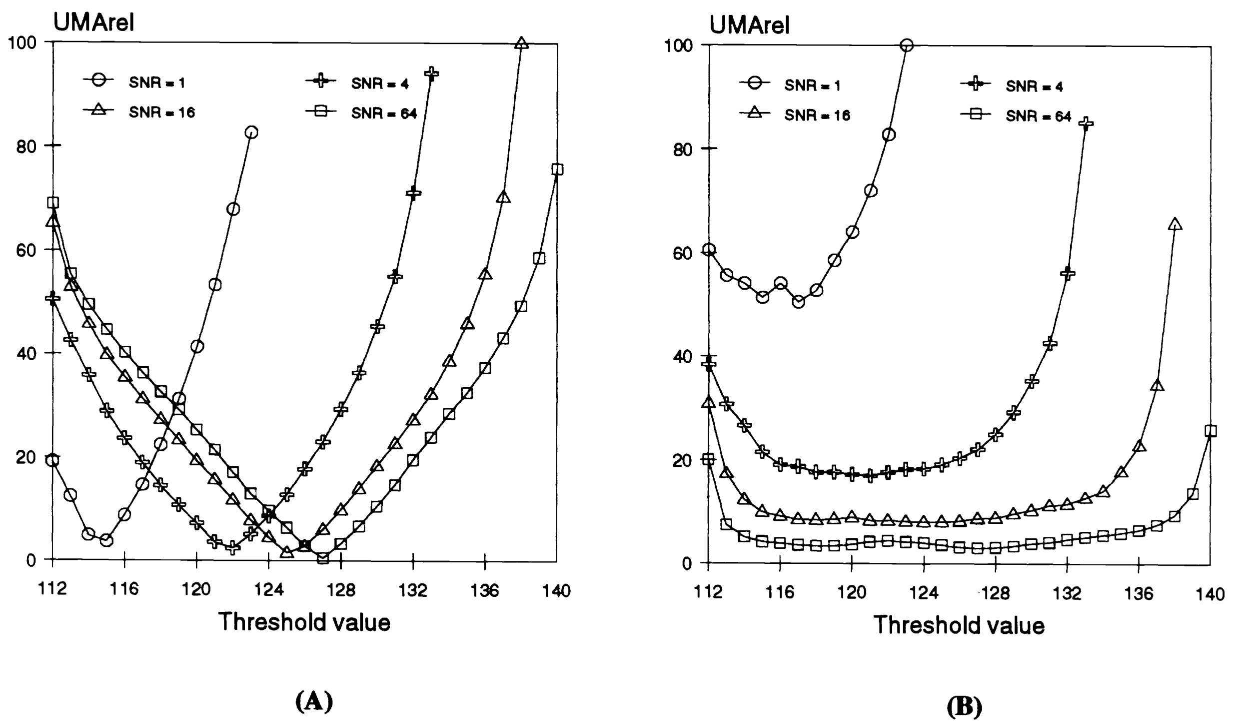


Fig.7 The results for (A) area, (B) p2a with images having diverse SNR and object #3.

In this paper, we also presented a study of several object features in the context of their ability to assess segmentation performance. What we learn from this study are that different object features have different evaluation ability, and that using different criteria, techniques would be ranked differently. This implies that the approach to segmentation evaluation presented here, as it can be adjusted to meet specific problem area, would be appropriate in different applications. From the other side, an algorithm which provides better segmented images (in the sense of obtaining more accurate measurement) for one object feature does not necessarily give better segmented images for other object features. Selecting appropriate techniques should be based on their performance in obtaining segmented images which provide good measurement accuracy in the context of desired object features.

There are at least three directions in which this work could be developed further. The first is to examine/study other object features which have current utility in image analysis, such as densitometry and texture features. So, wider application areas can be covered, and effective feature selection can be made. A second direction is to study how one can compromise between various object features in the segmentation evaluation task. For a specific application, it is valuable to achieve a unique score, so one can simply rank available algorithms. The third direction is to complete the basically experimental feature study by some analytic considerations.

ACKNOWLEDGEMENTS

This work was supported by The Netherlands' Project Team for Computer Science Research (SPIN), Three-Dimensional Image Analysis Project.

REFERENCES

1. K.S.Fu, J.K.Mui. "A Survey on Image Segmentation". *Pattern Recognition*, Vol.13, pp.3-16, 1981.
2. R.M.Haralick, L.G.Shapiro. "Image Segmentation Techniques". *Computer Vision Graphics Image Processing*, 29, pp.100-132, 1985.
3. P.K.Sahoo, S.Soltani, A.K.C.Wong, Y.C.Chen. "A Survey of Thresholding Techniques". *Computer Vision Graphics Image Processing*, 41, pp.233-260, 1988.
4. W.A.Yasnoff, J.K.Mui, J.W.Bacus. "Error Measures for Scene Segmentation". *Pattern Recognition*, Vol.9, pp.217-231, 1977.
5. J.S.Weszka, A.Rosenfeld. "Threshold Evaluation Techniques". *IEEE Transaction System, Man, and Cybernetic*, Vol.SMC-8, No.8, pp.622-629, 1978.
6. M.D.Levine, A.Nazif. "Dynamic Measurement of Computer Generated Image Segmentations". *IEEE Transaction Pattern Analysis Machine Intelligence*, Vol.PAMI-7, No.2, pp.155-164, 1985.
7. S.U.Lee, S.Y.Chung, R.H.Park. "A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation". *Computer Vision Graphics Image Processing*, 52, pp.171-190, 1990.
8. I.E.Abdou, W.Pratt. "Quantitative Design and Evaluation of Enhancement/Thresholding Edge Detectors". *Proceedings of IEEE*, Vol.67, No.5, pp.753-763, 1979.
9. F.Heyden. "Evaluation of Edge Detection Algorithms". *Proceeding of Third International Conference on Image Processing and its Applications*, Warwick, UK., pp.618-622, 1989.
10. Y.J.Zhang, J.J.Gerbrands. "Transition Region Determination Based Thresholding". *Pattern Recognition Letters*, Vol.12, pp.13-23, 1991.
11. Y.J.Zhang, J.J.Gerbrands. "On the Design of Test Images for Segmentation Evaluation". Accepted by *EUSIPCO-92*, Brussels, 1992.
12. Y.J.Zhang, J.J.Gerbrands. "Comparison of Thresholding Techniques Using Synthetic Images and Ultimate Measurement Accuracy". Accepted by *11 International Conference on Pattern Recognition*, The Hague, 1992.