A Method for Assessment of Segmentation Success Considering Uncertainty in the Edge Positions

Rubén Usamentiaga, Daniel F. García, Carlos López, and Diego González

Department of Computer Science, University of Oviedo, Campus de Viesques, 33204 Gijón, Asturias, Spain

A method for segmentation assessment is proposed. The technique is based on a comparison of the segmentation produced by an algorithm with an ideal segmentation. The procedure to obtain the ideal segmentation is described in detail. Uncertainty regarding the edge positions is accounted for in the discrepancy calculation of each edge using fuzzy reasoning. The uncertainty measurement consists of a generalization, using fuzzy membership functions, of the similarity metrics used by well-known assessment methods. Several alternatives for the fuzzy membership functions, based on statistical properties of the possible positions of each edge, are defined. The proposed uncertainty measurement can be easily applied to other well-known methods. Finally, the segmentation assessment method is used to determine the best segmentation algorithm for thermographic images, and also to tune the optimum parameters of each algorithm.

1. INTRODUCTION

Image segmentation is one of the most important components in an image analysis system. The objective of segmentation is to divide the image into meaningful regions. After the segmentation, features of each region are identified to be used for further analysis. Since the analysis of the image is based on the identified features, and the features are calculated from the segmented regions, the accuracy of the segmentation is crucial to the performance of the image analysis system.

Over the last few decades, many segmentation algorithms have been proposed [1]. However, the evaluation of the performance of these algorithms is usually poor, consisting of the presentation of a few segmented images.

In order to evaluate segmentation algorithms, several evaluation methods which can be used to determine the effectiveness of an algorithm, and which also allow the comparison of several algorithms, have been proposed.

In this work a new segmentation assessment method which does not have the well-known problems of other methods, and which also takes the uncertainty into account, is proposed. This method will be used to decide which algorithm is the best for the segmentation of thermographic images and also to find the optimum parameters of each algorithm.

2. PREVIOUS EVALUATION METHODS

Zhang [2] proposes a classification of existing assessment methods as "analytical," "empirical goodness," and "empirical discrepancy." Other authors, such as Yang et al.[3], use a different classification: "supervised" and "unsupervised," closer to the pattern matching terminology. The two classifications are equivalent, the supervised group corresponds to the empirical discrepancy group of Zhang, and unsupervised corresponds to the others.

Analytical methods attempt to characterize an algorithm in terms of principles, requirements, complexity, and so forth, with no reference to any concrete implementation of the algorithm or test data, such as time complexity or response to a theoretical data model.

Empirical goodness methods evaluate algorithms by computing a "goodness" metric on the segmented image without prior knowledge of the desired segmentation result. For example, Levine and Nazif [4] use intraregion gray-level uniformity as their goodness metric. Haralick and Shapiro [1] established other quality measures that could be classified in this group.

Empirical discrepancy methods calculate a discrepancy measurement between the result of the segmentation algorithm and the desired correct segmentation for the corresponding image (ideal segmented image). In the case of
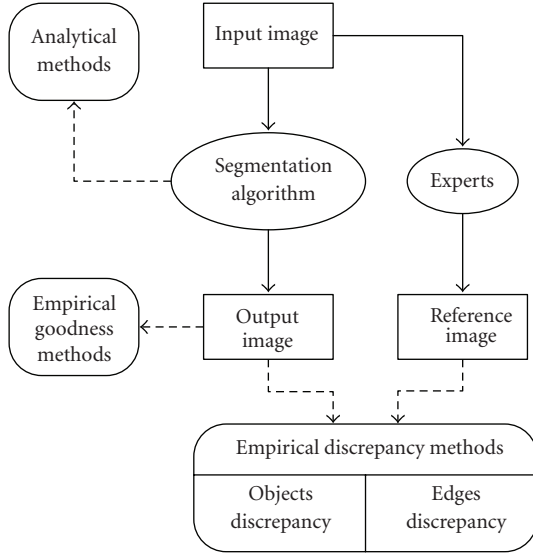
FIGURE 1: Segmentation and its evaluation.

synthetic images, the ideal segmentation can be obtained automatically from the image generation procedure, whereas in the case of real images, it must be produced manually by an experienced operator. The term "supervised," used also to characterize this group of methods, comes from the necessity of an ideal segmentation to provide the segmentation quality measure.

Analyzing Zhang's classification, most of the methods proposed in the literature belong to the Empirical discrepancy group. The reason is that, although analytical and empirical goodness methods are easier to apply since they do not need an ideal segmented image, they do not provide as much information about the performance of the segmentation as the empirical discrepancy methods. In fact, they are mainly used to obtain a preliminary metric, or when the ideal segmented image is not available.

Empirical discrepancy methods can be broken down into "empirical edges discrepancy" and "Empirical objects discrepancy" methods.

Empirical objects discrepancy methods use the properties of the segmented objects in the image to provide a measure of the quality of the segmentation process. An example of this group of methods called UMA (ultimate measurement accuracy) was proposed in [5], where features of segmented and ideal objects, such as area or perimeter, are used to measure the accuracy of the segmentation.

Empirical edges discrepancy methods use the position of the edges in the segmented image to measure the quality of the segmentation.

Figure 1 shows a general scheme of the segmentation and its evaluation methods using the analyzed classification.

The most common empirical edges discrepancy methods will be described in more detail, since the segmentation evaluation method proposed in this work belongs to this group. To describe this group of methods the notation in Table 1, based on the notation used in [6, 7], will be used.

Using this notation, the following relations can be established:

$$N_{SE} - N_{FP} + N_{FN} = N_{IE}, \tag{1}$$
$$N_{FP} + N_{FN} + N_{TP} + N_{TN} = N_P, \tag{2}$$
$$N_{FP} = N_{SE} - N_{TP}, \tag{3}$$
$$N_{FN} = N_{IE} - N_{TP}. \tag{4}$$

One of the approaches to measuring the quality of the segmentation is to consider the segmentation process as a pixel classification. Using this approach, a confusion matrix can be built considering two classes of pixels: edge pixels and nonedge pixels. Two error types can be calculated from this matrix which can be used as a measure of the performance of the algorithms.

The confusion matrix is shown in Table 2 and the error types are

$$\mathrm{Error}_{\mathrm{Edge}} = \frac{N_{FN}}{N_{TP} + N_{FN}} \times 100,$$
$$\mathrm{Error}_{\mathrm{Nonedge}} = \frac{N_{FP}}{N_{TN} + N_{FP}} \times 100. \tag{5}$$

Using the same approach, Lee et al. [8] proposed a different discrepancy measure termed "probability of error" (PE). For classification problems between object and background, the measure is defined as shown in (6), where $P(O)$ and $P(B)$ are a priori probabilities of objects and backgrounds, and $P(O/B)$ and $P(B/O)$ are the probabilities of error in classifying objects as background and vice versa:

$$\mathrm{PE} = P(O)P(B/O) + P(B)P(O/B). \tag{6}$$

Applying PE to a case where edges are considered objects and the remaining pixels are considered backgrounds, we obtain,

$$P(O) = \frac{N_{IE}}{N_P}, \tag{7}$$

$$P(B/O) = \frac{N_{FN}}{N_{IE}}, \tag{8}$$

$$P(B) = \frac{N_P - N_{IE}}{N_P}, \tag{9}$$

$$P(O/B) = \frac{N_{FP}}{N_P - N_{IE}}, \tag{10}$$

$$\mathrm{PE} = \frac{N_{IE}}{N_P} \frac{N_{FN}}{N_{IE}} + \frac{N_P - N_{IE}}{N_P} \frac{N_{FP}}{N_P - N_{IE}} = \frac{N_{FN} + N_{FP}}{N_P}. \tag{11}$$

Although methods based on pixel classification provide a satisfactory measurement of the quality of segmentation [2], they have a major drawback when applied to edges. The problem appears when $P(O)$ is much lower than $P(B)$, as the case is when considering edges as objects. In this case, the response of the quality measure is poor, with insufficient discrimination capability to distinguish small segmentation degradation.

TABLE 1: Notation used to describe the empirical edges discrepancy methods.

Term	Description
N_{FP}	Number of false positive detections: the number of pixels erroneously defined as edge pixels, that is, false alarms
N_{FN}	Number of false negative detections: the number of pixels erroneously defined as nonedge pixels, that is, missed detections
N_{TP}	Number of true positive detections: the number of pixels correctly defined as edge pixels, that is, hits
N_{TN}	Number of true negative detections: the number of pixels correctly defined as nonedge pixels, that is, correct rejections
N_{IE}	Number of pixels classified as edges in the ideal segmented image, that is, number of ideal edges
N_{SE}	Number of pixels classified as edges in the segmented image produced by an algorithm being evaluated, that is, number of found edges
N_P	Number of pixels in the image

TABLE 2: Confusion matrix for edge and nonedge pixels classes.

Class	Classification	
	Edge	Nonedge
Edge	N_{TP}	N_{FN}
Nonedge	N_{FP}	N_{TN}

Another widely used approach to measure the quality of segmentation is based on the distance from the missegmented pixels to their nearest ideal edge pixel. For example, Yasnoff et al. [9] proposed a distance metric, D, which can be calculated using (12), where $d(i)$ is the distance from the ith missegmented pixel to the nearest pixel that actually belongs to the misclassified class.

Yasnoff et al. proposed also the normalization of D (ND), to negate the influence of the image size, using (13).

$$D = \sum_{i=1}^{N_{FP}} d(i)^2, \qquad (12)$$

$$ND = \frac{\sqrt{D}}{N_P} \times 100. \qquad (13)$$

The measure of quality proposed by Yasnoff et al. has the disadvantage of taking into account false positive detections only, without considering false negative detections.

Another commonly used measure of quality is the figure of merit, proposed by Pratt [10]. This measure can be calculated using (14),[1] where M is calculated using (15), and p is a scaling constant (normally assigned to value 1):

$$FOM = \frac{1}{M} \sum_{i=1}^{N_{FP}} \frac{1}{1 + d(i)^2 p}, \qquad (14)$$

$$M = \text{MAX}(N_{IE}, N_{SE}). \qquad (15)$$

This measure is normalized in the range $[0, 1]$ and increases with the quality of the segmentation (a value of 1 represents perfect segmentation). However, supposing

$N_{FP} > N_{FN} > 0$, an increment of N_{FN} implies an increment of the measure, that is, the error is rewarded. Similar problems were detected [11].

An enhanced version of the previous measure was proposed by Strasters and Gerbrands [12] to deal with low error segmented images. The definition of this measure is as follows:

$$FOM_e = \begin{cases} \dfrac{1}{N_{FP}} \displaystyle\sum_{i=1}^{N_{FP}} \dfrac{1}{1 + d(i)^2 p} & \text{if } N_{FP} > 0, \\ 1 & \text{if } N_{FP} = 0. \end{cases} \qquad (16)$$

This measure, in the same way as the measure proposed by Yasnoff, only takes false positive detections into account.

Although the described measures are more commonly used, some authors have proposed different approaches. For example, in [6], an empirical procedure is proposed which is based on subjective human assessment. On the other hand, in [7] a statistical method to estimate the ideal segmentation automatically is proposed.

Despite all of these proposed methods, an agreement has not been reached among the image processing community about the proper evaluation method, probably because of the wide range of types of segmentation algorithms being analyzed. Thus, the procedure used to evaluate algorithms is usually chosen to be the one which best fits the characteristics of the segmentation algorithm.

3. PROPOSED EVALUATION METHOD

Before starting the design of any segmentation evaluation method, the following general properties are established as desirable.

(a) The set of ideal edges between regions for each image must be known, so that errors in segmentation can be detected and assessed one by one.

(b) The assessment method must provide a continuous magnitude, so the adjustment of the parameters of the segmentation algorithm can be carried out accurately.

(c) The values of the magnitude generated by the assessment procedure must be limited to a range, so they can be easily analyzed and compared.

(d) Both positive and negative errors must be taken into account.

---

[1] In the survey of segmentation evaluation methods carried out by Zhang [2], this equation is expressed incorrectly. The error is the use of M instead of N_{FP} as proposed by Pratt in [10].
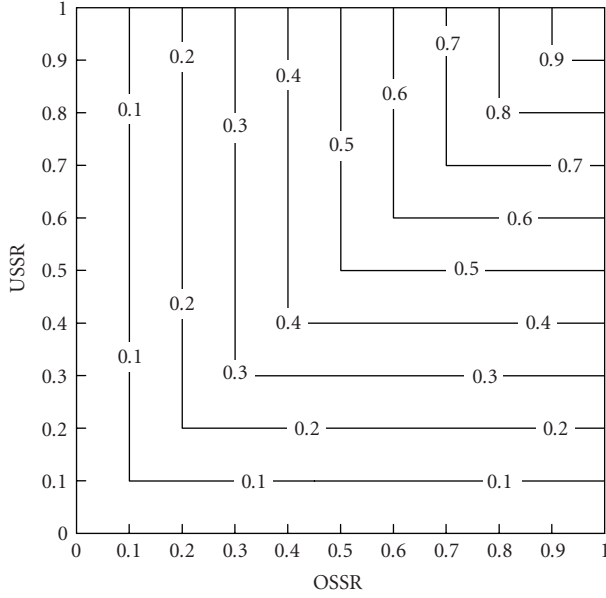
FIGURE 2: Combination of OSSR and USSR using the minimum operator.

(e) The assessment method must weigh up the error committed to detect each edge using the distance between the detected edge and the real edge. However, this must only happen when the position of the detected edge is within the influence area of the position of a real edge, and also when that real edge is closer to the detected edge than any other one. The metric established to measure the distance can be nonlinear.

(f) On some occasions and mainly due to the existing noise in the images, there is a degree of uncertainty in the determination of the edge position even by experienced operators; therefore, in this situation the assessment method needs to take the uncertainty of the ideal segmentation into account.

Analyzing the available methods described above, it can be concluded that none of them have all of these properties. For example, confusion matrix and probability of error have properties (a), (b), (c), and (d), but not (e) or (f); figure of merit has (a), (b), (c), and (e); and normalized distance and expanded figure of merit have (a), (b), (c), and (e). In this work, a new quality measure is proposed which incorporates all of these desirable properties.

3.1. Measure of quality

To simplify the problem initially, a situation without uncertainty will be considered, that is, when the match of a found edge and an ideal edge can be considered totally true or totally false.

In this situation two major types of error can arise: over-segmentation and under-segmentation, that is, false alarms and missed detections. Both of them can occur in the same image.

The success considering only over-segmentation error can be measured through the ratio OSSR (over-segmentation success ratio), shown in (17). The OSSR is 1 when all the found edges match ideal edges (all of them are hits), and decreases when the found edges do not match ideal edges (false alarms appear). When none of the found edges match an ideal edge, the OSSR is 0.

$$\text{OSSR} = \frac{N_{TP}}{N_{SE}}. \tag{17}$$

In the same way, the success considering only under-segmentation error can be measured through the ratio USSR (under-segmentation success ratio), shown in (18). The USSR is 1 when all the ideal edges are found (all of them are hits), and decreases when the ideal edges are not found (missed detections appear). When none of the ideal edges are found, the USSR is 0.

$$\text{USSR} = \frac{N_{TP}}{N_{IE}}. \tag{18}$$

Both the OSSR and USSR metrics define the success in the segmentation. Therefore, success in the segmentation can be obtained as a combination of the OSSR and USSR.

The combination of these values is carried out through an operator, T, which satisfies the following requirements.

(i) Boundary: $T(0,0) = 0$, $T(a,1) = T(1,a) = a$.
(ii) Monotonicity: $T(a,b) \leq T(c,d)$ if $a \leq c$ and $b \leq d$.
(iii) Commutativity: $T(a,b) = T(b,a)$.
(iv) Associativity: $T(a, T(b,c)) = T(T(a,b),c)$.

Although many different operators that fulfill previous requirements have been proposed [13, 14], the most commonly used operators are the minimum and the multiplication. Figures 2 and 3 show the graphic representation of both operators.

In this work, the multiplication operator will be used to make the combination of both ratios more restrictive. Finally the proposed measure termed success segmentation ratio (SSR) can be expressed as follows:

$$\text{SSR} = \text{OSSR} \times \text{USSR} = \frac{N_{TP}}{N_{SE}} \frac{N_{TP}}{N_{IE}} = \frac{(N_{TP})^2}{N_{SE} N_{IE}}. \tag{19}$$

4. TAKING UNCERTAINTY INTO ACCOUNT

In the previous section, the SSR assumes that N_{TP} is known with no uncertainty. In this case N_{TP} is an integer.

Next, the process to calculate N_{TP} under uncertainty will be described. In this case N_{TP} will be a real number.

4.1. Creation of ideal segmentation

When determining the effectiveness of a segmentation algorithm empirically, it is necessary to start from an ideal segmentation which defines the objective desired output of the segmentation process.

The way the ideal segmentation is created depends on the type of images used. Thus, when synthetic images are used,
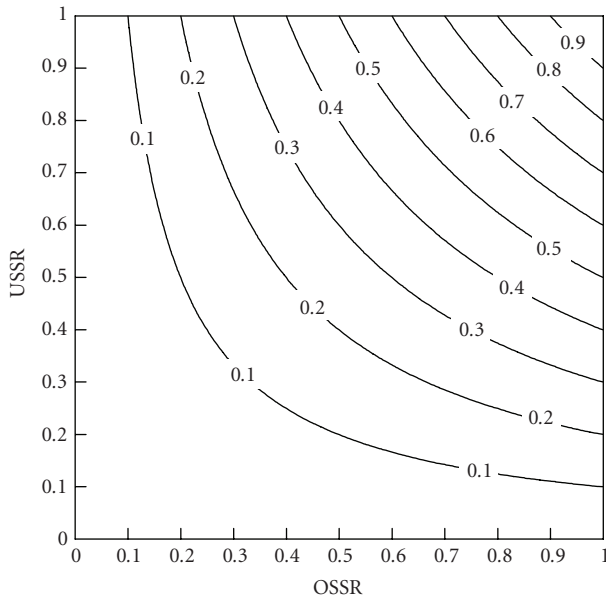
FIGURE 3: Combination of OSSR and USSR using the multiplication operator.

the creation of the ideal segmentation is a nonsubjective process. On the other hand, if real images are used, it is necessary to carry out a manual segmentation, where some kind of uncertainty could appear.

In this work, real images are used to obtain the ideal segmentation. Although the creation of the ideal segmentation of each real image must be manual, and thus, time-consuming, it avoids the problems derived from the validation of synthetic images. It is important to note that synthetic images should represent real images; therefore, some kind of validation should be carried out, which poses an additional problem.

The method proposed to obtain the ideal segmentation is described below.

(i) Select a subset of images that represents the set of images that are going to be segmented by the algorithm being evaluated.

(ii) Select a group of experienced operators to segment the images in the test set manually. Several experienced operators must be used to compensate for the subjectivity of finding the edges in the images. Each of the operators must provide an ideal segmentation for each image in the test set.

(iii) Define a single ideal segmentation for each image, where only those edges established by more than half of the experts will be considered.

(iv) The data for each edge in the ideal segmentation will consist of a set of positions established by the operators. The dispersion of the positions of each edge is the uncertainty introduced by the operators.

To make this process easier, a software tool is recommended to help the experts in the establishment of each edge. The

tool should include zoom capabilities and a visual editor of the edges of the image.

4.2. Generalization of discrepancy measures in an edge

Once the experienced operators have carried out the manual segmentation and the segmentations have been integrated to form the ideal segmentation for each image, comparisons between the segmentation found and the ideal segmentation can be made to obtain a similarity or discrepancy degree.

The similarity determination problem has two inputs: the list of edges produced by the segmentation algorithm and a matrix created by the experienced operators consisting of a list of lists, where the positions of the ideal edge are specified. For each ideal edge, a list of positions is available.

The similarity problem can be reduced initially to obtain the similarity between two edges, one belonging to the ideal segmentation and the other belonging to the segmentation produced by the algorithm.

As explained above, the position of each ideal edge is only specified as a list of possible positions, that is, the position of the edge is only known under uncertainty. Thus, the calculation of the discrepancy in an edge will be carried out under uncertainty, since this calculation is based on the ideal edge position. Therefore, the discrepancy in a found edge will point out that the edge is a "possible" match to an ideal edge, assigning a numerical match value, which suggests the confidence between true and false.

Classical logic has two possible values, true and false. Intuitively, it seems logical to think about many of the events that usually occur as neither totally true nor totally false; it is difficult to represent these events using a logic system that only uses two values. Using the idea of a multivalue logic, Zadeh [15] introduces the term fuzzy logic. Fuzzy logic provides the opportunity for modeling conditions that are inherently imprecisely defined. The classical set theory, where the membership μ of an element x to a set A will be 1 if $\mu(x) \in A$ and 0 if $\mu(x) \notin A$, is extended into fuzzy set theory, where the membership is defined as a function, $\mu_A(x)$, that takes values in the range $[0, 1]$.

Once the membership function is established, it is possible to determine the membership degree of the position of a found edge to the fuzzy set which describes the position of an ideal edge; in other words, the match value of the found edge and the ideal edge.

Fuzzy logic fits the desirable properties to calculate the discrepancy between two edges (found and ideal) perfectly. Therefore, in this work the use of fuzzy membership functions is proposed to determine the discrepancy between a found edge and an ideal edge under uncertainty.

Due to the versatility of the definition of the fuzzy membership functions, some of the available discrepancy methods can also be defined as fuzzy processes. Thus, a fuzzy membership function can be obtained for some of the available methods. For example, the method proposed by Lee et al.[8] uses a membership function that can be defined by (20), where Ei
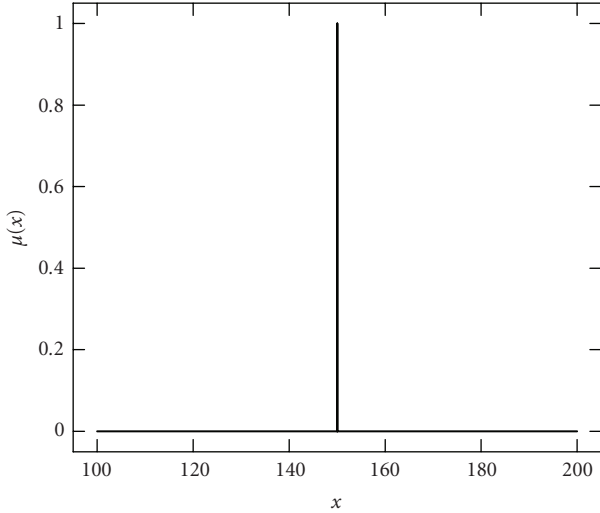
FIGURE 4: Fuzzy membership function equivalent to the similarity metric used by Lee et al.
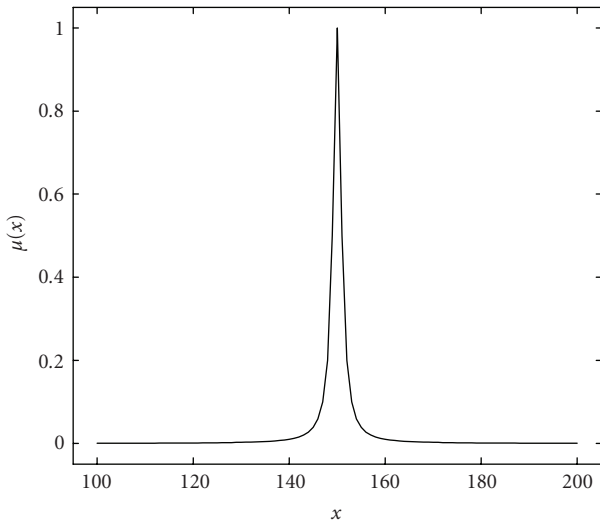

FIGURE 5: Fuzzy membership function equivalent to the similarity metric used by Pratt.

is the position of an ideal edge in the image:

$$\mu_{Ei}^{\text{Lee}}(x) = \begin{cases} 1 & \text{if } Ei = x, \\ 0 & \text{if } Ei \neq x. \end{cases} \tag{20}$$

Similarly, the membership function used by Pratt can be defined by.

$$\mu_{Ei}^{\text{Pratt}}(x) = \frac{1}{1 + (Ei - x)^2 \times p}. \tag{21}$$

Figures 4 and 5 show a graphic representation of the fuzzy membership functions equivalent to the similarity metrics used by Lee and Pratt, where Ei is 150.

In conclusion, the use of fuzzy membership functions to determine the discrepancy between two edges (found and

ideal) can be seen as a generalization of various existing discrepancy methods, mainly, methods in the empirical edges discrepancy group. To calculate the discrepancy between two edges under uncertainty, a fuzzy membership function based on the uncertainty of the position of the ideal edge will be defined.

The information provided by the experienced operators will be used to define the membership function. Through an analysis of the dispersion of the set of positions for each edge, the accuracy in the knowledge of this position can be defined, that is, the uncertainty. This uncertainty will be different for each edge.

Summarizing, the generalization of the calculation of the discrepancy in each edge can be seen from two points of view.

(1) The determination of the segmentation quality measure is based on the similarity to an ideal segmented image. The similarity measured can be generalized as a fuzzy process, which is based on the definition of a membership function.

(2) In the available methods, the same function is used to measure the discrepancy in each edge. This work proposes a different function for each ideal edge based on the uncertainty of its position.

Next, some fuzzy membership functions based on the uncertainty of the position of the ideal edge are defined.

4.3. Proposed alternatives for the discrepancy measurement of an edge

4.3.1. Discrepancy based on Pratt's method

A fuzzy membership function based on Pratt's method (PM) can be created by converting the scaling constant, p, originally defined empirically, into a function depending on the dispersion of the position of the ideal edge. Equation (22) shows the definition of the function, where $M(Ei)$ is the average position established by the experienced operators for edge Ei, and $p(Ei)$ is the scaling for edge Ei expressed as a function of the dispersion of the position of edge Ei:

$$\mu_{Ei}^{\text{PM}}(x) = \frac{1}{1 + (M(Ei) - x)^2 \times p(Ei)}. \tag{22}$$

Following the same technique Pratt used to limit the range of his measure, $p(Ei)$ can be defined as shown in (23), where $S(Ei)$ is the standard deviation of the positions established for edge Ei:

$$p(Ei) = \frac{1}{1 + S(Ei)^2}. \tag{23}$$

Figure 6 shows the graphic representation of (23) when $M(Ei)$ is 150 and $S(Ei)$ is 10.

4.3.2. Discrepancy based on a double confidence interval

Considering the set of positions established by the experts for each edge as a set of independent observations, it can be said
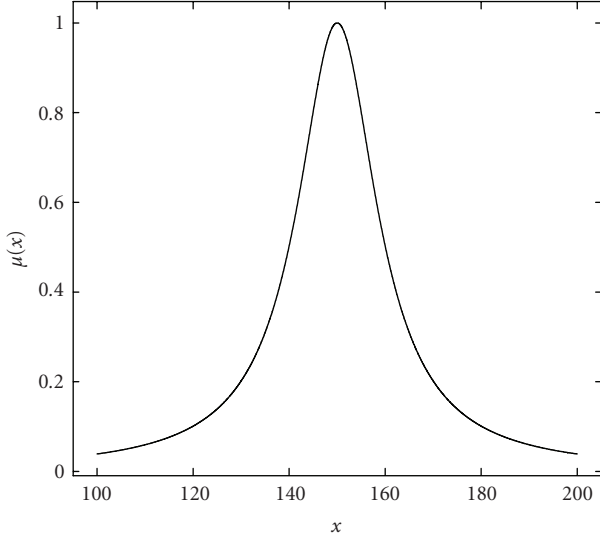
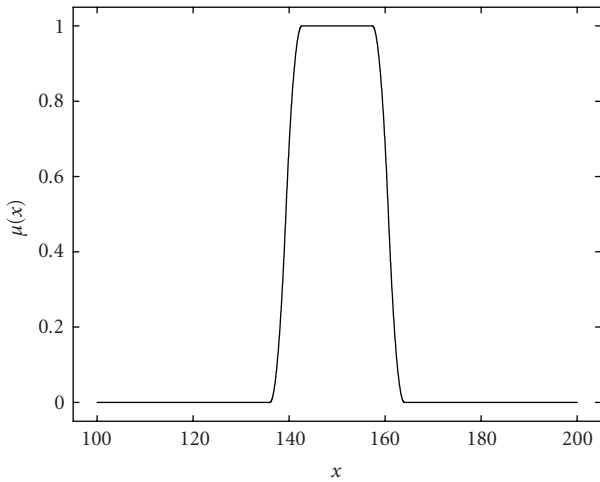FIGURE 6: Fuzzy membership function based on Pratt's method.

FIGURE 7: Fuzzy membership function based on a double confidence interval.

that the average of the position of the edges is described by a t-distribution for any number of operators, but if the number of operators is greater than 30, the normal distribution is a satisfactory approximation to the t-distribution and may be used instead.

To compare the distribution of the position of one edge provided by the operators with the position of the edge provided by the segmentation algorithm, a membership function can be defined based on two confidence intervals.

The description of each confidence interval is as follows.

(1) A confidence interval that points out that the edge found by the algorithm "truly" matches an edge in the ideal segmentation: this confidence interval is calculated using the significance level α_T.

(2) A confidence interval that points out that the edge found by the algorithm "possibly" matches an edge in

the ideal segmentation: this confidence interval is calculated using the significance level α_P.

Using both confidence intervals, a continuous function can be defined, consisting of a plateau of value 1, for input values inside the truly confidence interval, and two smooth slopes, one on each side, modeled as spline curves between the difference of the limits of both intervals.

The reason for using the spline curve, rather than the Gaussian or the sigmoidal, is that it is easier to implement, and it is easily limited in the range $[0, 1]$ for a set of input values. Also, the spline is preferred rather than the linear because it is smooth and does not have abrupt changes.

Equation (25) shows the definition of the membership function based on a double confidence interval. When the number of experienced operators is less than 30, the values, which determine where the limits of the intervals are, can be calculated using (26).

$$\mu_{Ei}{}^{\mathrm{DCI}}(x), \tag{24}$$

$$= \begin{cases} 0 & \text{if } x < C_1^{\alpha_P}, \\ 2\left(\dfrac{x - C_1^{\alpha_P}}{C_1^{\alpha_P} - C_1^{\alpha_T}}\right)^2 & \text{if } C_1^{\alpha_P} \leq x < \dfrac{C_1^{\alpha_T} + C_1^{\alpha_P}}{2}, \\ 1 - 2\left(\dfrac{C_1^{\alpha_T} - x}{C_1^{\alpha_P} - C_1^{\alpha_T}}\right)^2 & \text{if } \dfrac{C_1^{\alpha_T} + C_1^{\alpha_P}}{2} \leq x < C_1^{\alpha_T}, \\ 1 & \text{if } C_1^{\alpha_T} \leq x \leq C_2^{\alpha_T}, \\ 1 - 2\left(\dfrac{x - C_2^{\alpha_T}}{C_2^{\alpha_P} - C_2^{\alpha_T}}\right)^2 & \text{if } C_2^{\alpha_T} < x \leq \dfrac{C_2^{\alpha_T} + C_2^{\alpha_P}}{2}, \\ 2\left(\dfrac{C_2^{\alpha_P} - x}{C_2^{\alpha_P} - C_2^{\alpha_T}}\right)^2 & \text{if } \dfrac{C_2^{\alpha_T} + C_2^{\alpha_P}}{2} < x \leq C_2^{\alpha_P}, \\ 0 & \text{if } x > C_2^{\alpha_P}. \end{cases} \tag{25}$$

$$\begin{aligned} C_1^{\alpha_P} &= M(Ei) - t_{n-1;1-\alpha_P/2} \frac{S(Ei)}{\sqrt{n}}, \\ C_2^{\alpha_P} &= M(Ei) + t_{n-1;1-\alpha_P/2} \frac{S(Ei)}{\sqrt{n}}, \\ C_1^{\alpha_T} &= M(Ei) - t_{n-1;1-\alpha_T/2} \frac{S(Ei)}{\sqrt{n}}, \\ C_2^{\alpha_T} &= M(Ei) - t_{n-1;1-\alpha_T/2} \frac{S(Ei)}{\sqrt{n}}. \end{aligned} \tag{26}$$

Figure 7 shows the graphic representation of (25) when $M(Ei)$ is 150, $S(Ei)$ is 10, n is 7, α_P is 0.2 (80%), and α_T is 0.01 (99%).

4.3.3. Discrepancy based on a hypothesis test

The discrepancy problem can be established as a hypothesis test, where the null hypothesis is that x, the position in which the segmentation algorithm finds an edge, matches $M(Ei)$, $H_0 : x = M(Ei)$, that is, if the edge found by the algorithm matches an ideal edge. As an alternative hypothesis, x can be considered to be different from $M(Ei)$, $H_1 : x \neq M(Ei)$.
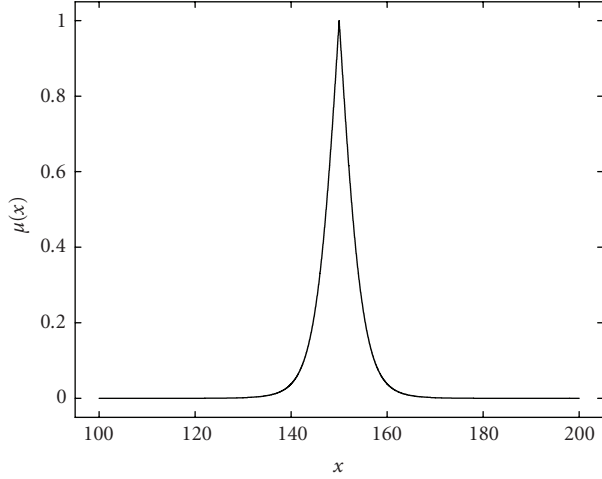
FIGURE 8: Fuzzy membership function based on the P-value of a hypothesis test.

The hypothesis test is the process which decides which of the two hypotheses is accepted and which is rejected. The decision is based on the evidence established by a sample which is used to calculate a statistic of the test T. T is the natural estimator associated to the parameter referenced in the hypothesis.

To decide if a hypothesis is accepted, a confidence interval is calculated using a significance level. If the value of x is within the interval, the hypothesis is accepted. Otherwise, it is rejected.

It is possible to accept a hypothesis using one significance level and reject it using another. The decision is binary: either it is accepted or rejected, that is, it is noncontinuous. However, the risk of accepting the hypothesis, which is a continuous value, is also used. The risk is measured using the P-value, which represents the minimum significance level which could be used to reject the null hypothesis.

The P-Value exists for every hypothesis test; it is a continuous value which measures the confidence in the acceptation of a hypothesis. All these properties describe a suitable membership function.

Equation (27) shows the definition of the membership function based on a hypothesis test, where P is the probability of the Student t-distribution (less than 30 experienced operators):

$$\mu_{Ei}^{\mathrm{HT}}(x) = P\left\{\frac{|M(Ei) - x|}{S(Ei)/\sqrt{n}}\right\}_{n-1}. \tag{27}$$

Figure 8 shows the graphic representation of (27) when $M(Ei)$ is 150, $S(Ei)$ is 10, and n is 7.

4.4. Uncertainty consideration in the metric of quality

In order to calculate the proposed measure of quality, it is necessary to calculate the value of N_{TP}. The calculation of this value will be carried out between the list edges found by the algorithm and the set of edges in the ideal segmentation.

```
Proc CreatePL (IdealEdgeList, FoundEdgeList): PairList
  List IEListNA = IdealEdgeList;
  List FEListNA = FoundEdgeList;
  PairList = Empty;

  While (IEListNA! = Empty) AND (FEListNA! = Empty)

   Min = MAX_DOUBLE;
   Foreach edge i in IEListNA
    Foreach edge j in FEListNA
     If (ABS(Pos(i)-Pos(j) < Min)
       IdealPair = i;
       FoundPair = j;
       Min = ABS(Pos(i)-Pos(j));
     End-If
    End-Foreach
   End-Foreach
   Add IdealPair and FoundPair to PairList;
   Eliminate IdealPair from IEListNA;
   Eliminate FoundPair from FEListNA;
  End-While
  Return PairList

End-Proc
```

ALGORITHM 1: Procedure for the creation of the pair of edges.

This search will produce a list of pairs of edges, found and ideal. Each of them will be used to calculate the discrepancy.

Algorithm 1 shows the detailed steps to create the list of pairs, where IEListNA means ideal edge list not assigned, and FEListNa means found edge list not assigned.

Once the list of pairs of edges is available, N_{TP} can be calculated easily. Typically, N_{TP} is calculated using (28), where PL is the list of pairs, Length(PL) is the number of pairs in the pair list, PL_$Es[k]$ is the found edge in the kth position of the pair list, and PL_$Ei[k]$ is the ideal edge in the kth position of the pair list. This equation counts the number of matching edges found:

$$N_{TP} = \sum_{k=0}^{\mathrm{Length(PL)}} \begin{cases} 1 & Si & \mathrm{PL}\_Ei[k] = \mathrm{PL}\_Es[k], \\ 0 & Si & \mathrm{PL}\_Ei[k] \neq \mathrm{PL}\_Es[k]. \end{cases} \tag{28}$$

Although the approach used in (28) is the most common, it should be used only if there is no uncertainty in the position of the edges, and the magnitude of the errors does not need to be considered. This approach can be generalized using a fuzzy membership function as in (29). Indeed, (28) is a particular case of (29) when the Lee fuzzy membership function is used:

$$N_{TP} = \sum_{k=0}^{\mathrm{Length}(PL)} \mu_{\mathrm{PL}\_Ei[k]}^{f}(\mathrm{PL}\_Es[k]). \tag{29}$$

The selection of the fuzzy membership function depends on the response required from the quality measure.
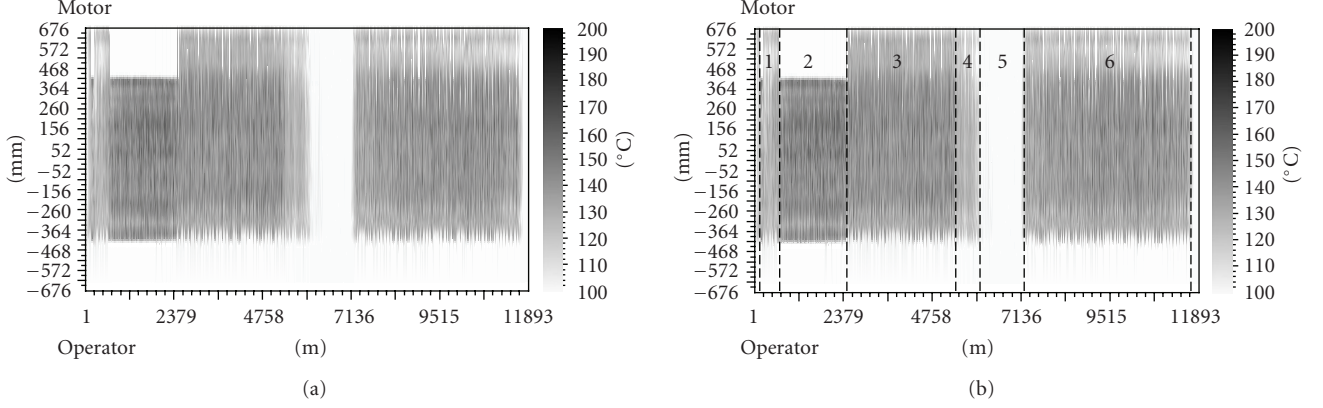
FIGURE 9: Thermographic image (a) and its desired segmentation (b) in patterns.

Substituting (29) into (19), we obtain

$$\text{SSR} = \frac{\left(\sum_{k=0}^{\text{Length}(PL)} \mu_{\text{PL\_}Ei[k]}^{f}\left(\text{PL\_}Es[k]\right)\right)^2}{N_{SE} N_{IE}}. \qquad (30)$$

This assessment method fits all the desirable properties established since it takes uncertainty into account, it is continuous, and it is limited to the range of $[0, 1]$; a value of 1 means perfect segmentation.

It is important to note that the uncertainty measurement proposed in this work can be applied to any other empirical edges discrepancy method, since it defines the uncertainty in the calculation of N_{TP}. For example, this uncertainty measure can be applied to the "probability of error" method; substituting (3) and (4) into (11), (31) is obtained; and substituting (29) into (31), (32) is obtained, which represents a parameterized probability of error:

$$\text{PE} = \frac{(N_{IE} - N_{TP}) + (N_{SE} - N_{TP})}{N_P}, \qquad (31)$$

$$\text{PE} = \frac{N_{IE} + N_{SE} - 2\sum_{k=0}^{\text{Length}(PL)} \mu_{\text{PL\_}Ei[k]}^{f}\left(\text{PL\_}Es[k]\right)}{N_P}. \qquad (32)$$

Although the uncertainty measurement proposed in this work is based on the different positions established by the experienced operators for each edge, an empirical approach has been widely used to define this function. For example, Pratt's evaluation method defines a function based on the quadratic distance. In the same way, different functions could be defined.

5. PRACTICAL APPLICATION: SEGMENTATION OF THERMOGRAPHIC IMAGES

The proposed measure of quality has been used to evaluate the segmentation carried out over thermographic images [16]. Next, the images, the algorithms and the SSR application are described.

5.1. Description of the images

Image acquisition is carried out using an infrared line scanner (IRLS), with which thermographic line scans are captured from hot steel strips while they are moving forward along a track.

The repetitive line scanning and the movement of the strip make the acquisition of a rectangular image possible. The image obtained consists of a stream of line-scans.

Typically, the resolution of the images resulting from the acquisition is 130 rows and 10 000 columns. Each pixel of the image represents the temperature in the range 100–200 °C.

5.2. Objective of the segmentation

The segmentation carried out over thermographic images tries to find regions of homogeneous temperature, that is, regions formed by a set of adjacent line scans which have a similar temperature pattern. This makes the result of the segmentation much more difficult to assess, due to the inherent subjectivity of the homogeneity definition.

Different regions in the thermographic image appear as a consequence of the changes of the manufacturing conditions of the strip over time. These changes produce a different thermographic line-scan pattern.

The segmentation procedure will group similar line scans producing, finally, a set of line scans temperature patterns.

Figure 9 shows an example of a thermographic image and its desired segmentation. As it can be seen, regions are always longitudinal segments of the image.

5.3. Description of the algorithms

In this work, two segmentation algorithms were proposed and tested. Both are adapted versions of well-known approaches: region-merging segmentation and edge-based segmentation. A description of both algorithms is included below. Further information can be found in [16].
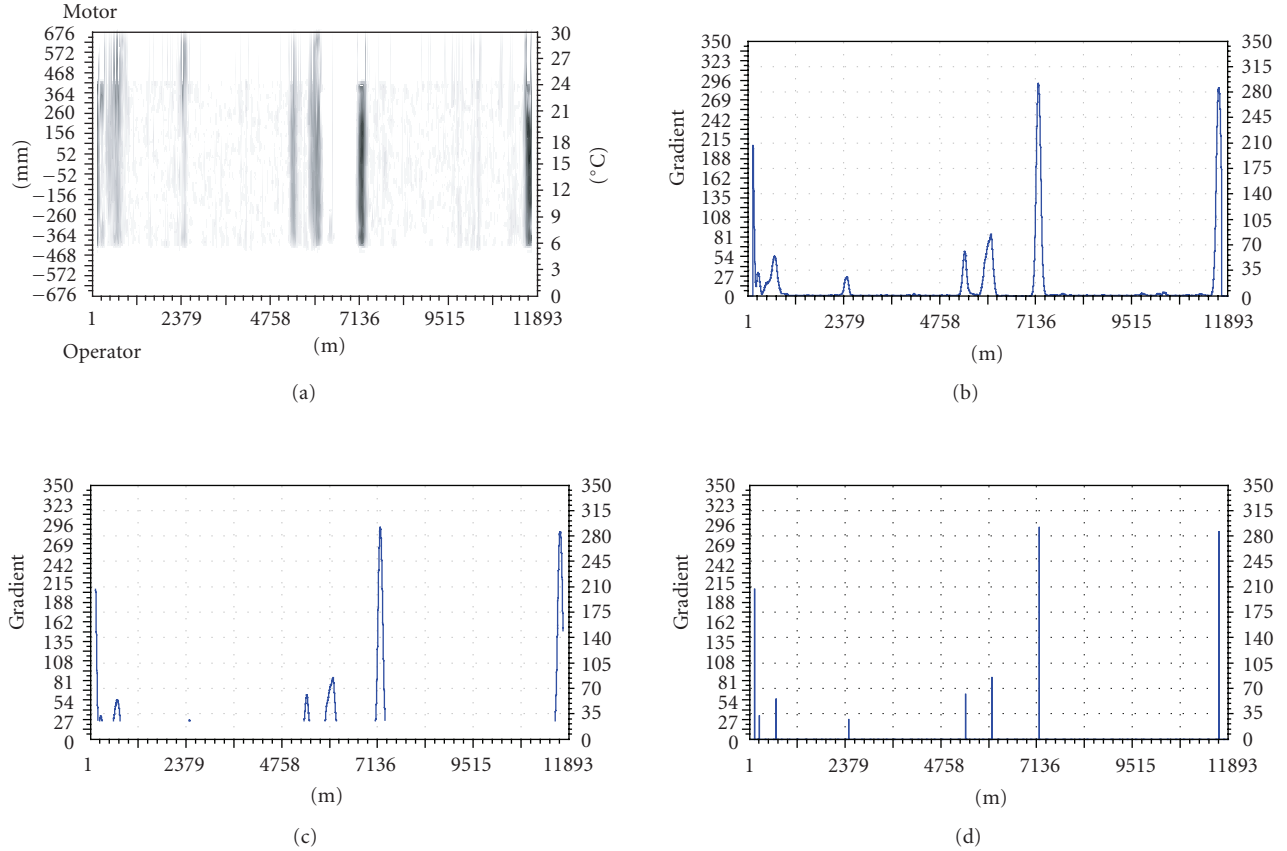
(a)

(b)

(c)

(d)

FIGURE 10: Steps in the segmentation using the edge detection approach: (a) thermographic gradient map, (b) quadratic projection, (c) threshold 25, quadratic projection, and (d) edges.

5.3.1. Region-merging segmentation

Region-merging segmentation methods search for adjacent regions within an image which meet some defined similarity criteria to merge them into a bigger one.

In this case, the image was initially divided into as many regions as line scans. Adjacent regions were then merged using an elaborated distance metric.

This algorithm was configured through four parameters: initialization size of a region, minimum region size, homogeneity threshold, and line scan confidence range.

5.3.2. Edge-based segmentation

Edge-based segmentation techniques rely on edges found in an image by edge-detection operators. These edges mark image discontinuities regarding some attributes of the image. Usually, the attribute is the luminance level; in this case, the temperature level was used.

Different gradient operators were tested to choose the one best suited for this kind of edge profiles, including boxcar (extended Prewit), LoG (Laplacian of Gaussian), and FDoG (first derivative of Gaussian). However, since the different operators produced a similar result, box-car was used

because its recursive implementation was faster than the others'. This operator can be described as

$$\begin{bmatrix} -1 & -1 & \cdots & -1 & -1 & 0 & +1 & +1 & \cdots & +1 & +1 \end{bmatrix}. \quad (33)$$

Once the edge operator was applied, a gradient for the image was obtained. The next step of the segmentation was the projection of the gradient to the longitudinal axis and the threshold.

This algorithm was configured through two parameters: threshold and operator length.

Figure 10 shows the steps carried out in the segmentation of the image shown in Figure 9 using the edge detection approach. Firstly, (a) the gradient is calculated, then, (b) the gradient is projected, and (c) thresholded, lastly, (d) edges are obtained as the highest value of each peak.

5.4. Applying the SSR metric

The performance of both segmentation algorithms was assessed using SSR, for which a set of images was selected. The selected test-image set included images with different patterns of temperature changes. The test set was manually segmented by a group of seven experts using a software tool to carry out the segmentation more easily.
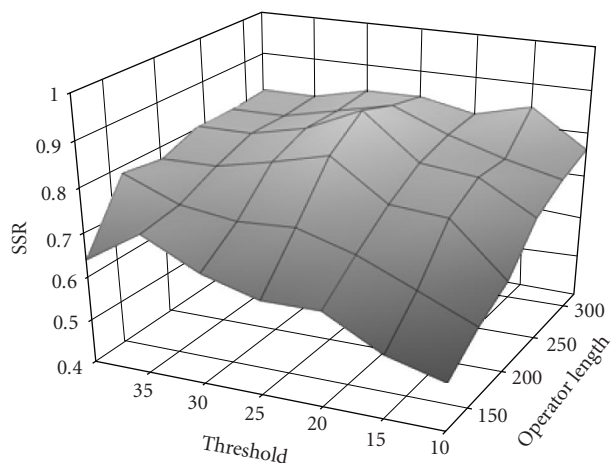
FIGURE 11: SSR of the edge-based algorithm for the thermographic images.

All the images in the selected test set were segmented by the two proposed algorithms with different values for their configuration parameters. The evaluation procedure was based on a complete factorial experimental design. The number of different combinations of parameters for the region-merging algorithm was 1 728 for each image. The number of different combinations of parameters for the edge-based algorithm was 144 for each image. This lower number is due to the lower number of parameters of the second algorithm.

The membership function based on a double confidence interval (DCI) was selected to measure the discrepancy on each edge. This selection was based on the profile of the edges that were to be detected. To define the confidence intervals, α_P was set to 0.2 (80%), and α_T was set to 0.01 (99%).

The segmentation of the test image set using the region-merging algorithm provided a geometric mean of SSR value of between 0 and 0.78, where only 5 combinations of the configuration parameters produced a geometric mean of SSR of over 0.7.

The segmentation of the test image set using the edge-based algorithm provided a value for the geometric mean of SSR of between 0 and 0.88, where most of the combinations of the configuration parameters produced a geometric mean of SSR of over 0.75. Figure 11 shows the geometric mean of SSR values produced by the edge-based segmentation algorithm for each parameters combination.

The edge-based segmentation algorithm performed better than the region-merging algorithm, since from the best 25 configurations of both algorithms, 23 correspond to the edge-based algorithm. Also, it is more robust since it reached a correct segmentation for nearly all of the strips.

The SSR measure of quality made it possible not only to determine the best algorithm, but also to determine the optimum parameter tuning of both algorithms.

6. CONCLUSIONS

A method for segmentation success assessment considering uncertainty in the edge positions has been proposed. It could be classified as an empirical edges discrepancy method. This method is based on the definition of a measure of quality, which combines the success in the over-segmentation (OSSR) and in the under-segmentation (USSR). It calculates a discrepancy between the segmentation produced by an algorithm and an ideal segmentation. The steps necessary to create the ideal segmentation using several experienced operators have been described in detail. Uncertainty is considered in the discrepancy calculation of each edge using fuzzy reasoning.

The procedure used to calculate the uncertainty is a generalization of the methods used by other authors. Indeed, the metrics used to calculate similarity can be interpreted as fuzzy membership functions, which can be used to calculate the membership degree of the position of a found edge to the fuzzy set which describes the position of an ideal edge.

Several alternatives are proposed to measure the discrepancy in an edge. They are based on a definition of a fuzzy membership function using a statistical characterization of the uncertainty produced by the different positions established by the experienced operators for each ideal edge.

Finally, the uncertainty measurement has been integrated in the proposed segmentation assessment method. This measurement can also be used in other existing methods as was demonstrated for the well-known method "probability of error."

The proposed method has been successfully applied to the assessment of thermographic image segmentation algorithms. It allowed both the determination of the best algorithm and the tuning of the optimal parameters.

REFERENCES

[1] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 1, pp. 100–132, 1985.

[2] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognition*, vol. 29, no. 8, pp. 1335–1346, 1996.

[3] L. Yang, F. Albregtsen, T. Lønnestad, and P. Grøttum, "A supervised approach to the evaluation of image segmentation methods," in *Proc. 6th International Conference on Computer Analysis of Images and Patterns (CAIP '95)*, pp. 759–765, Prague, Czech Republic, September 1995.

[4] M. D. Levine and A. M. Nazif, "Dynamic measurement of computer generated image segmentations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 2, pp. 155–164, 1985.

[5] Y. J. Zhang and J. J. Gerbrands, "Segmentation evaluation using ultimate measurement accuracy," in *Image Processing Algorithms and Techniques III*, vol. 1657 of *Proceedings of SPIE*, pp. 449–460, San Jose, Calif, USA, May 1992.

[6] R. Román-Roldán, J. F. Gómez Lopera, C. Atae-Allah, J. Martínez-Aroza, and P. L. Luque-Escamilla, "A measure of quality for evaluating methods of segmentation and edge detection," *Pattern Recognition*, vol. 34, no. 5, pp. 969–980, 2001.

[7] Y. Yitzhaky and E. Peli, "A method for objective edge detection evaluation and detector parameter selection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 8, pp. 1027–1033, 2003.

[8] S. U. Lee, S. Y. Chung, and R. H. Park, "A comparative performance study of several global thresholding techniques for segmentation," *Computer Vision, Graphics, and Image Processing*, vol. 52, no. 2, pp. 171–190, 1990.

[9] W. A. Yasnoff, W. A. Mui, and J. W. Bacus, "Error measures in scene segmentation," *Pattern Recognition*, vol. 9, no. 4, pp. 217–231, 1977.

[10] W. K. Pratt, *Digital Image Processing*, Wiley-Interscience, New York, NY, USA, 1977.

[11] F. van der Heyden, "Evaluation of edge detection algorithms," in *Proc. 3rd International Conference on Image Processing and its Applications*, pp. 618–622, Warwick, UK, July 1989.

[12] K. C. Strasters and J. J. Gerbrands, "Three-dimensional image segmentation using a split, merge and group approach," *Pattern Recognition Letters*, vol. 12, no. 5, pp. 307–325, 1991.

[13] R. R. Yager, "On a general class of fuzzy connectives," *Fuzzy Sets and Systems*, vol. 4, no. 3, pp. 235–242, 1980.

[14] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, NY, USA, 1980.

[15] L. A. Zadeh, "Fuzzy sets and systems," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[16] R. Usamentiaga, D. F. García, C. López, and J. A. González, "Algorithms for real-time acquisition and segmentation of a stream of thermographic line scans in industrial environments," *Journal of Imaging Science and Technology*, vol. 49, no. 2, pp. 138–153, 2005.

Rubén Usamentiaga was born in Santander, Spain, on 30 December 1974. He received his M.S. and Ph.D. degrees in computer science from Oviedo University in 1999 and 2005, respectively. He is currently an Associate Professor at the Department of Computer Science and Engineering at Oviedo University. In the recent years, he has been also working on several projects related to computer vision and industrial systems. His research interests include real-time imaging systems and thermographic observers for industrial processes.

Daniel F. García is a Professor at the Department of Computer Science and Engineering at Oviedo University. In 1988, he received his Ph.D. degree in industrial engineering from Oviedo University. Since 1994, he has been responsible for the computer engineering area at the University of Oviedo. His current research interest is in the area of the development of real-time and embedded systems applied to quality assurance and production inspection in industry. For the last ten years, Dr. García has been conducting research projects in the area of information technologies applied to industry at national and European levels. He is a Member of ACM and the IEEE Computer Society.

Carlos López was born in Oviedo, Spain, on 16 May 1979. He received his M.S. in computer science from Oviedo University in 2002. He is currently an Associate Professor at the Department of Computer Science and Engineering at Oviedo University. In the recent years, he has been working on several projects related to computer vision and industrial systems. His research interests include real-time imaging systems and range measurement techniques.

Diego González was born in Gijón, Spain, on 24 October 1976. In 2003, he received his M.S. degree in Computer Science from Oviedo University. He is currently an Associate Professor at the Department of computer science and Engineering at Oviedo University. In the last years, he has been working in a project related to computer vision and industrial systems. His research interests include real-time imaging systems and data mining in industrial processes.