A COMPARATIVE STUDY OF SEGMENTATION TECHNIQUES IN THE DEVELOPMENT OF A MEDICAL IMAGE ANALYSIS SYSTEM

A. Castro, S. Fernández and B. Arcay

Department of Information and Communication Technologies, Faculty of Computer Science, University of A Coruña, A Coruña, Spain

alfonso@udc.es, sfdelavega@udc.es, cibarcay@udc.es

Abstract: There has been a great deal of progress in medical image analysis as a consequence of the development of new image capture systems and power PAC. There is, moreover, a great deal of interest in developing automatic segmentation techniques given the benefits they imply; nonetheless, the area is still undergoing intensive research, and most of the systems developed to date have not been proved in a work environment.

Our objective is to develop a system that carries out an initial analysis and selects the best algorithm from among those available for the type of image in question, yet which allows the specialist to select an alternative algorithm and modify the parameters with a view to improving results.

Our system to date incorporates edge detection and fuzzy clustering algorithms; we used both because they have been widely studied, tested and used in medical image analysis. We tested these algorithms on a heterogeneous set of images: photographs of burn patients and x-rays of patients with hip implants, as we were of the opinion that a heterogeneous set of images would give a good approximation of the stability of the algorithms for broader sets of images.

Introduction

Recent studies show that in 90% of the diagnoses carried out by a physician, one or several modalities of medical imaging are implied in a different number of degrees and shapes [1]. The analysis and handling of this information however is complex, due to the specific characteristics of the images: multidimensionality, high inter-observer variability, multiple information sources, etc. Moreover, the area of medical imaging has known significant advances since late’s of 1980’s.

The result of this evolution is an increased interest for the development of techniques and tools that facilitate the analysis of the medical image by the medical staff: standards that guarantee the quality and Interchange of information (DICOM, Digital Imaging and Communications in Medicine) [2], databases for the teaching and comparison of cases, and techniques for the detection of interesting elements, the quantification and recommendation of a diagnostic and/or therapeutic output.

For many authors, the most important phase in the analysis of digital images is the segmentation [3], because the quality of the subsequent phases (representation and interpretation) largely depends on them. At present, this area is the object of intense research and solutions are not easily found.

The majority of the developed applications and systems focus mainly on the resolution of very specific problems and this under a large number of limiting circumstances; many other critical areas, such as the validation of the different methods, it’s not considered exhaustively [4]. The result of this attitude is a lack of generalisation that seriously complicates the integration of solutions into the daily clinical activities of the expert.

Our aim is to build a tool that allows the medical expert to analyse the images during his daily clinical routine and that recommends him the best available algorithm and the parameter values for the detection of relevant facts in a specific image.

We opted for the study of algorithms that belong to the techniques of edge detection and fuzzy clustering, because of the extensive experience in the development of segmentation algorithms for both techniques, and because of the large number of contrasted algorithms and applications in the field of medical image analysis based in those techniques.

The analysed images are images of burned patients and X-Rays of patients with a total hip replacement. We believe that by using a very heterogeneous set of medical images, it will help to determine the stability of the algorithms and to test whether they can be “adapted” to a large amount of highly diverging images. The objective is to avoid an combinatory explosion adding many algorithms depending about the type of image and motive.

Material and methods

Standard image processing techniques were used to study the images.

In the burned patient images (Figure 1) was observed that there was a great variation among the images, given that it is not possible to ensure the absolute uniformity of the elements that make up an image, such as patient position, lighting, background, objects, etc. The images were digitised to a size of 500x500 pixels and 24 colour bits per pixel.

ISSN: 1727-1983 © 2005 IFMBE

As for the x-ray images (Figure 2), inhomegeneities were observed - as would be expected for an irradiation technique - in the intensity levels for single elements (caused by different degrees of X-ray absorption by the patient tissues), as well as a low signal-noise ratio in many areas.

Another problem observed was superimposition between different elements, resulting from the position adopted by the patient at the time of taking the x-ray, a problem which makes it very difficult to correctly delimit the bone. There were also artefacts in areas of sharp changes in contrast, which also made the analysis of the images more difficult.

Figure 1: One of the photographs used in the study

To digitise the x-rays a scanner specially adapted to this kind of image was used. The image capture resolution was 100 pixels/inch, with 256 intensities of grey.

The clustering algorithms analysed were as follows: Fuzzy C-Means (FCM), a fuzzy generalisation of the C-Means algorithm; Fuzzy K-Nearest Neighbour (FKNN), a fuzzy version of the K Nearest Neighbour, and Modified Fuzzy C-Means (MFCM), a variation on FCM that divides the image into clusters on the basis of the image histogram, which avoids the problem of having to pre-determine the number of clusters and provide a sample for each.

For this study, the Canny, Heitger and Bezdek edge detection algorithms were analysed.

Different quantitative measures were used to analyse the results, given that the results of each algorithm have different characteristics.

The fuzzy clustering algorithms results were evaluated using RUMA (relative ultimate measurement accuracy) and overall discrepancy for the segmented pixels, whereas the edge detector results were evaluated using discrepancy based on probability of error and ROC (receiver operating characteristic) curves.

Figure 2: One of the X-rays used in the study

Fuzzy C-Means (FCM)

The FCM initially needs the number of clusters in which the image will be divided and a sample of each cluster [5]. The steps of this algorithm are:

1. Calculation of the membership of each element to each cluster:

$$u_k(i,j) = \left(\sum \frac{\left\| y(i,j) - v_k \right\|^{\frac{2}{m-1}}}{\left\| y(i,j) - v_j \right\|} \right)^{-1} \quad (1)$$

2. Calculation of the new centroids of the image:

$$v_k = \frac{\sum_{i,j} u_k(i,j)^m \, y(i,j)}{\sum_{i,j} u_k(i,j)^m}, k = 1, \ldots, C \quad (2)$$

3. If the error stays below a determined treshold, stop. In the contrary case, return to step 1.

Fuzzy K-Nearest Neighbour (FKNN)

The Fuzzy K-Nearest Neighbour [6] is, as its name indicates, a fuzzy variant of a hard segmentation algorithm. It needs to know the number of classes into which the set that must be classified will be divided.

The element that must be classified is associated to the class of the nearest sample among the K most similar ones. These K most similar samples are known as “neighbours”; if, for instance, the neighbours are classified from more to less similar, the destination class of the studied element will be the class of the neighbour that is first on the list.

ISSN: 1727-1983 © 2005 IFMBE

We use the following expression to calculate the membership factors of the pixel to the considered clusters:

$$u_i(x) = \frac{\sum_{j=1}^{K} u_{ij}\left(\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}}\right)}{\sum_{j=1}^{K}\left(\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}}\right)} \quad (3)$$

Modified Fuzzy C-Means (MFCM)

This algorithm is based on the work of Young Won Lim and Sang Uk Lee [7], who describe an algorithm for the segmentation of colour images through the study of the histograms of each colour bands. This algorithm also relies on the classification algorithm fuzzy c-means.

The MFCM consists of two parts:

1. A hard part that studies the histograms of an image in order to obtain the number of classes, and carries out a first global classification of the image; and
2. A fuzzy part that classifies the pixels that have more difficulties in determining the class to which they belong. The pixels of this area are called “fuzzy zone”.

Once obtained the initial clusters with its centroids, the algorithm uses the FCM membership function to classify the pixels. The fuzzy points are pixels between the initial clusters and pixels of clusters too little for its consideration.

Canny edge detector

This algorithm was developed by J. Canny [8]. The detector follows a series of steps:

1. Smoothen the original image with a bidimensional Gaussian function. The width of the function is specified by the user.
2. Calculate the derivation of the filtered image with respect to the two dimensions, in order to calculate the size and direction of the gradient.
3. Find the points of the edge, which correspond with a maximum. Non-maxima suppression: the objective is to eliminate non-maxima perpendicular to the edge direction, since we expect continuity of edge strength along an extended contour. Any gradient value that is not a local peak is set to zero.
4. Apply tresholding hysteresis. We eliminate those points that are below an inferior limit specified by the user. The points over the superior limit are considered to belong to the edge. The points between the two limits are accepted if they are close to a point with a high response.

Heitger edge detector

The focus of the Heitger edge detector [9] is different from that of Canny in that it tries to solve the weaknesses of algorithms that use anisotropic linear operators.

This algorithm uses a logical/linear operator, based on a set of filtres in quadrature phase (zero-mean Gabor filters), to detect the interesting elements of the image. The operator is based on the representation of the normal signal in a curve that depicts the edges/line dichotomy.

A suppression and enhancement phase is applied next in order to avoid the possible ambiguities which might arise from the use of this kind of operator. The responses of those image points for which the operator response is not characterised by an ideal edge or line are diminished or eliminated, while the corresponding responses are enhanced. The suppression is carried out from the first derivative of the response module, while the enhancement module is based on the second directional derivative of the response module.

Finally, a non-maxima suppression phase is applied, building a binary image from it by using a threshold value, which is a configurable parameter of the algorithm.

Bezdek fuzzy edge detector

The algorithm [10] is based on the analysis of the geometrical features that an edge must possess, and it develops feature detection functions that allow the detection of those features in the image. Finally, it analyses the result produced by the detectors by means of a fuzzy selection function of the candidate points, since it allows the integration of a certain learning capacity in the selection of pixels.

He uses a Sobel filter in horizontal and vertical direction as a function for feature detection in the current algorithm implementation. The function composition applied by him is based on Takagi-Sugeno’s fuzzy rules system. Finally, he constructs a binary image of edges on background using a threshold value that may be fixed as an algorithm parameter.

The quality of the algorithms was measured with various quantitative measures, because it is very difficult to reflect all the factors that have an influence on the result in one single measure. The masks were made by specialists. Also, the results show entirely different characteristics according to the applied algorithm, and as yet the validation measures are not independent from the provided result.

The success level of the clustering algorithms was measured by selecting RUMA [11] and global discrepancy based on the segmented pixels [12]. The success level of the edge detection algorithms was measured by selecting the edge detection probability [13] and ROC curves [14].

ISSN: 1727-1983 © 2005 IFMBE

RUMA (Relative Ultimate Measurement Accuracy)

The purpose of RUMA, proposed by Zhang, is to measure the quality of the segmentation in terms of the similarity of the measures carried out on the segmented image and on the real image. In most occasions, the purpose of segmentation is to obtain a measure on the object that is segmented: surface, perimeter, length, etc.; the smaller the difference between the values measured in the resulting image and the real image, the more efficient the segmentation algorithm (Eq. 4).

$$RUMA = \frac{|R_f - S_f|}{R_f} x100\% \quad (4)$$

ODS (Overall Discrepancy for Segmented Pixels)

This measure supposes that the error will be smaller if the error of each cluster classification is smaller, so it measures the error in the pixel classification of each cluster and weight it against the number of pixels of that cluster, in equation 5 we can observe this measure.

$$error = \sum_{j=1}^{n}\sum_{i=1}^{n}\frac{F_{ij}}{MASC_j}, i \neq j \quad (5)$$

EDP (Edge Detection Probability)

The performance of the edge detectors was firstly measured by calculating the edge detection probability [16]. We suppose an image that consists of an edge/background dichotomy: the bigger the probability of correctly classifying an edge pixel, the better the segmentation algorithm, as can be seen in equation:

$$D = \frac{N_b + N_h}{N} \quad (6)$$

N_b: number of pixels that are false positives in the result image
N_h: number of pixels that are false negatives in the result image
N: number of total image points

ROC (Receiver Operating Characteristic)

This method consists of the comparison by means of a chart between the false positives (points detected as edges which are not) and the true positives (accurate edge points). If sampling of the algorithm's parameter space is done at a sufficiently small interval, then a response curve may be created. In this curve we will need to look for the point with an optimal ratio: the relation between true positives and false positives is maximal.
A frequently used change consists of representing the percentage of undetected edge (false negatives) compared to the ratio of possible false positives (%undetected, %false). A cloud similar to the previous one but inverted is created.

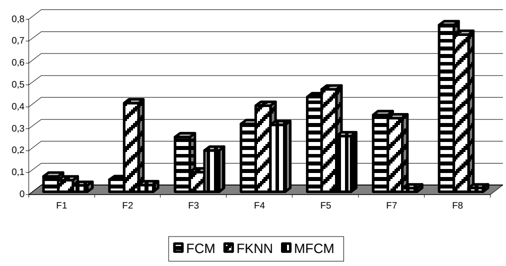
Since we are looking for each detector's best performance for the set of images, those points selected for creating the curve will be those with the lowest ratio of false pixels, and with the lowest value for undetected points.

In order to calculate the best ROC curve, we calculate the curve which fits best with the cloud of points defined by the algorithm. That curve which leaves less area underneath would be the best one for the latter case. An alternative to curve calculation is the use of the trapezoidal rule in order to calculate the area. Thus, if we consider a set of points *(TP_i,FP_i)*, assuming that *TP* does not decrease as *i* increases, then the rule would be the following:

$$A_{ROC} = \sum_{i=1}^{k-1}(FP_{i+1} - FP_i)x(TP_i + TP_{i+1})/2 \quad (7)$$

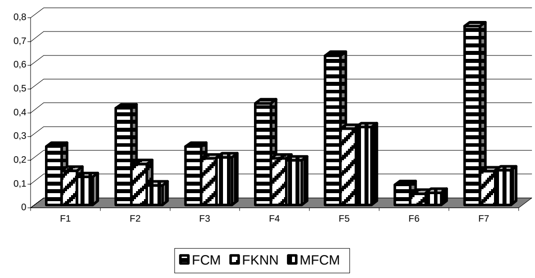
Results and Conclusions

In the calculation of RUMA we used the number of pixels obtained in each cluster in the segmented image as descriptor, while observing the discrepancy with the number of marked pixels for this area in the mask made by the expert. It is a rather precise approximation of the burned area and it is of interest to the medical expert. In Graphic 1 we can see the results for RUMA for some images of burned patients. The best results were obtained for MFCM algorithm.



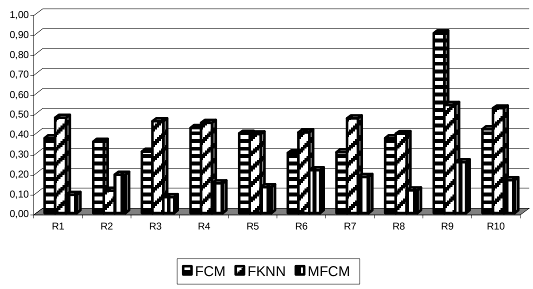
Graph 1: RUMA values for the some of the images of burned patients and fuzzy clustering algorithms

RUMA gives the success rate for one cluster, to resolve this problem we decide to use OSD. It allows us appreciate the global success rate of the algorithm in an easily interpretable way. In graphics 2 and 3 we can see the values for this measure for some of the burned patients and X-rays images for the different algorithms analysed.



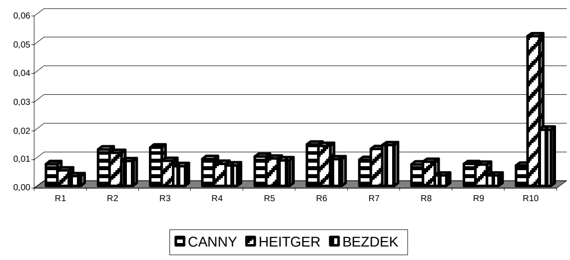
Grap 2: ODS values for the some of the images of burned patients and fuzzy clustering algorithms

ISSN: 1727-1983 © 2005 IFMBE

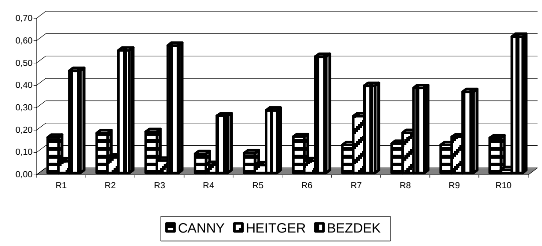


Graphic 3: ODS values for the some of the images of X-rays and fuzzy clustering algorithms

The next step was the evaluation of the edge detectors algorithms. The first measure that was used to evaluate the precision of the edge detectors was the calculation of EDP. However, this measure was unable to detect the cases that, due to a small number of false positives, gave good results, but lost important parts of interesting edges. In graphics 4 and 5 we can see the minimal and maximal values for this measure for three edge detectors.

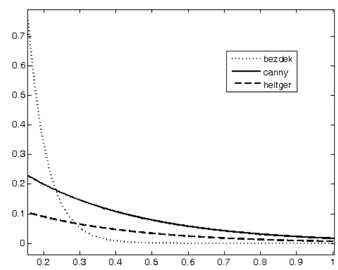


Graphic 4: The minimal values for EDP for the three edge detectors

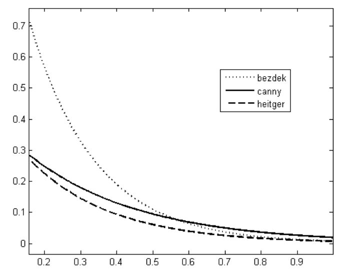


Graphic 5: The maximal values for EDP for the three edge detectors

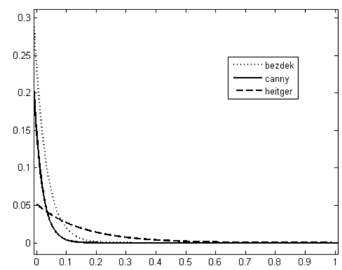
The most frequently used method to measure the performance of edge detection algorithms are the ROC curves. The advantage of this measure is that it allows us to observe more closely the performance of the algorithms in a set of images, not only in one image, which makes it more adequate for our purposes. Its main disadvantage is the complexity of its calculation, since it requires a large number of tests and its value is not calculated directly. In graphics 6, 7 and 8 we can see the ROC curves for some of the X-ray images.



Graphic 6: ROC curve for three edge detectors for R1 X.-ray



Graphic 7: ROC curve for three edge detectors for R8 X.-ray



Graphic 8: ROC curve for three edge detectors for R10 X.-ray

The tests show that the fuzzy clustering algorithms are able to segment correctly the images of burned patients, in spite of the variable conditions in which the photographs were obtained. The average global success level for the studied set of images is above 80% for the MFCM and FKNN algorithms (Graphic 2), with slightly superior results for the first. However, these algorithms are not able to produce high level results for the X-Ray images, because the inhomogeneity of the images makes it difficult to obtain quality centroids for the segmentation.

The edge detectors are able to delimit correctly the interesting edge in the X-rays and to detect most edges in the burned patient images with a success probability above 90% for both cases. It remains however difficult to find a set of parameters that can be applied to the different burned patient images and produce good results. We believe that this is due to the high variability of the environment and the acquisition condition of the pictures. The results for Heitger detector were nearly similar to Canny (, in some tests, they were better). However, we select Canny because its stability was better.

References

[1] PREDOSA C., CASANOVA R. (1997): 'Diagnostico por imagen', Mc Graw Hill, 2nd edition.
[2] DICOM, Internet site address: http://medical.nema.org.
[3] GONZALEZ R.C., WOODS R.E. (1996): 'Digital Image Processing', Addison-Wesley.
[4] COURTNEY P, THACKER N.A. (2001): 'Performance Characterisation in Computer Vision: The Role of Statistics in Testing and Design', Imaging and Vision Systems: Theory, Assessment and Applications, in Blanc-Talon J, Popescu D, (Eds.), NOVA Science Books.
[5] BEZDEK, J.C. (1981): 'Pattern recognition with fuzzy objective function algorithms', New York: Plenum.
[6] GIVENS JR. J.A., GRAY M.R., KELLER J.M. (1992): 'A fuzzy k-nearest neighbour algorithm', in: BEZDEK J.C., PAL S.K. (Eds.): 'Fuzzy models for pattern recognition: methods that search for structures in data', (IEEE Press, New York), pp. 258-263.
[7] LEE S.U., LIM Y.M. (1990): 'On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques', Pattern Recognition, 23 (9), pp. 935-952.
[8] CANNY J. (1986): 'A computational approach to edge detection', IEEE Trans. on Pattern Analysis and Machine Intelligence, 8 (6), pp. 679-698.
[9] HEITGER F. (1995): 'Detection using suppression and enhancement', Technical report n. 163, Image Science Lab, ETH-Zurich.
[10] BEZDEK J.C., CHANDRASEKHAR R., ATTIKOUZEL Y. (1998): 'A geometric approach to edge detection', IEEE Transactions on Fuzzy Systems, 6 (1), pp. 52-75.
[11] ZHANG Y.J., GERBRANDS J.J. (1992): 'Segmentation evaluation using ultimate measurement accuracy', SPIE, 1657, pp. 449-460.
[12] ALONSO A., ARCAY B., CASTRO A. (2000): 'Analysis and evaluation of hard and fuzzy clustering segmentation techniques in burned patients images', Image and Vision Computing, 18 (13), pp. 1045-1054.
[13] LEE S.U., CHUNG S.Y., PARK R.H. (1990): 'A comparative performance study of several global thresholding techniques for segmentation', Computer Vision, Graphics and Image Processing, 52, pp. 171-190.
[14] BOWYER K.W. (2000): 'Validation of Medical Image Analysis Techniques', in: SONKA M., FITZPATRICK J.M. (Eds.), 'Handbook of medical imaging', Volume 2: Medical image processing and analysis, SPIE, pp. 567-607.