

The Performance of Various Edge Detector Algorithms in the Analysis of Total Hip Replacement X-Rays

Alfonso Castro, Carlos Dafonte, and Bernardino Arcay

Dept. of Information and Communications Technologies,
Faculty of Computer Sciences,
University of A Coruña,
Spain
{alfonso,dafonte,cibarcay}@udc.es

Abstract. Most traumatology services use radiological images to control the state and possible displacements of total hip replacement implants. Prostheses are typically and traditionally detected by means of edge detectors, a widely used technique in medical image analysis. This article analyses how different edge detectors identify the prosthesis in X-Rays by measuring the performance of each detection algorithm; it also determines the clinical usefulness of the algorithms with the help of clinical experts.

1 Introduction

Traumatology services spend a considerable amount of time on the follow-up of patients with prostheses or screws in order to check the state of a particular orthopedic device. These follow-ups are usually based on a series of X-rays and comply with standardised protocols [1]. Even so, the medical expert controls the displacement or loosening of the prosthesis with certain measurements that are often subjective and fail to detect small movements that may be significant.

At present, several research groups are developing systems that automatically segment the bone and prostheses and as such provide the expert with quantitative measurements that help him to evaluate the patient's condition. Downing [2] developed a new and accurate method to automatically segment, classify and measure the femoral components of cemented total hip replacement in radiographs; Ozanian [3] described the development of a system that assists the expert in trajectory planning for computer-assisted internal fixation of hip fractures; and Behiels [4] evaluated various image features and different search strategies to apply Active Shape Models (ASM) to bone object boundaries in digitalized radiographs and carry out quantitative measures.

The aim of our group is to build a system that is based on the analysis of X-Rays and automatically segments the bone and the orthopedic device, providing the expert with a series of measurements that improve his diagnosis. The system consists of a module that recommends the most convenient algorithm to evaluate the images, and tries to automate its use as much as possible; the applied algorithm and the values of the different parameters can then be modified by the user in order to enhance the detection [5][6][7].

We selected the following algorithms on the basis of their use in the field of medical images analysis: Canny [8], Heitger [9], and the Bezdek fuzzy edge detector [10]. And they have provided results in a range of edge detector studies such as the works of Bowyer [11][12], which are among the best in computerized vision literature. Bowyer's first work presented the results of various algorithms to several observers for evaluation; the second work compared different edge detectors against a heterogeneous set of images by means of ROC curves.

The Canny algorithm is at present considered the reference algorithm in edge detection and has been used in a wide range of image analysis applications. The Bezdek fuzzy edge detector is being used in the development of a system for the detection of tumors in mammographies. We have not found any applications in the field of medical images analysis for the Heitger algorithm, but we nevertheless decided to include it because it provided the best results in Bowyer's edge detectors study [12], which is commonly considered one of the most extensive studies on this technique.

The tests are based on a heterogeneous set of images that presents the most commonly found characteristics: marks made by the expert, sudden contrast changes, etc. The results were contrasted with ground truth masks that present the perfect segmentation carried out by several experts.

In recent years researchers have used techniques based on models for the analysis of X-Ray prosthesis images, i.e. the ASM technique [13]. This technique however requires a considerably "clear" training set (well delimited and with clearly marked areas), it can generate critical outliers in difficult classification areas, and the results are not easily modifiable by the operator. Even though edge detectors do not provide a completely automatic segmentation, we believe that they are better adapted to the characteristics of our system, since their parameters can easily be modified and the results can be corrected with erase and mark tools.

2 Selection of the Test Set

The images used in this study were selected with traditional image processing techniques [14] and with the histogram as a basic tool (see Figure 1). The number of available images amounts to 40.

The available bank of images was analyzed to select a subset of X-Rays that reflected as precisely as possible the different characteristics of the population (artifacts caused by the prosthesis, superpositions, fuzzy edges between the areas, saturation, etc.). We made a visual analysis and histogram exam of the images to decide whether to include or reject the image.

Firstly, as was to be expected in a case of irradiation techniques such as X-Rays, we observed inhomogeneities in the intensity levels of one and the same element, due to the uneven absorption of the radiation by the patient's tissues.

We also noticed that in most cases the histogram is bimodal: one of the modes corresponds to the background in the black area, whereas the other represents the prosthesis with the highest luminosity levels. The transition between the modes is continuous, with a detectable value for the different grey values, which implies that

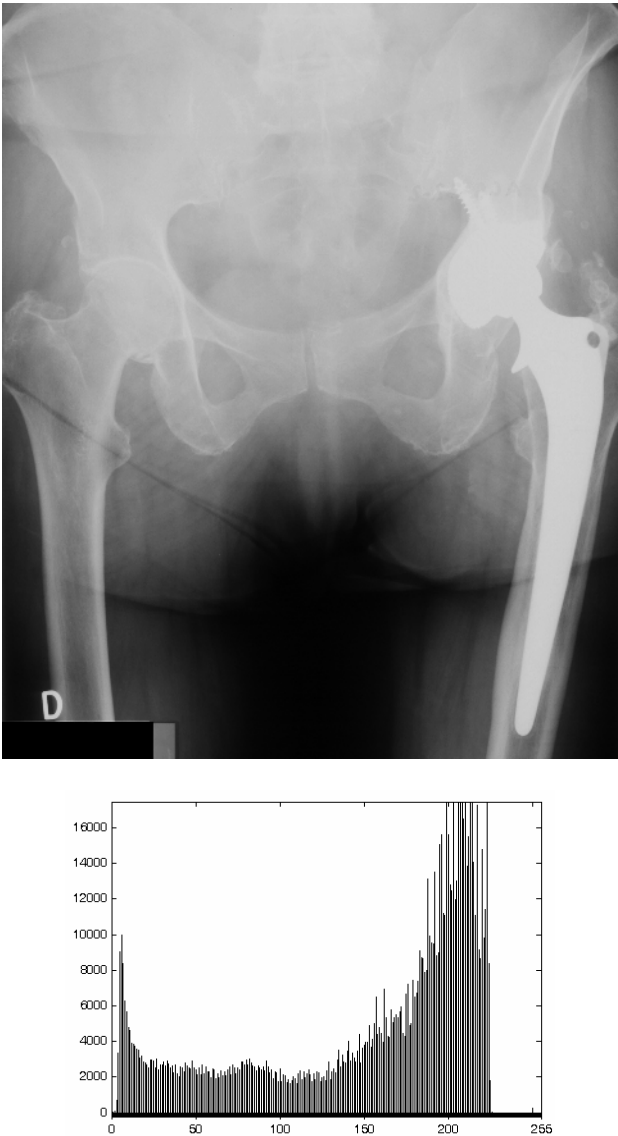


Fig. 1. Image and histogram used in this study

the edges between the elements are fuzzy. The image confirms this. We also observed that the shape of the histogram is similar for all the X-Rays and that the differences are due to the number of pixels between the modes, which varies according to the presence of more or less tissue in the X-Ray.

The images were digitalized with an especially designed scanner; the resolution for the image acquisition was 100 pixels/inch and the number of grey levels 256, which is the highest available value.

Since we observed that most images were similar with regard to the motives that appear and to the intensity levels and their distribution, we finally selected a set of 10 images. The choice of this number was based on the large amount of results that will be obtained from the edge detectors and the fact that the results will be evaluated by experts. An excessively high number of images would make tedious and complicate this evaluation process.

3 Edge Detectors

The algorithms were selected according to two basic criterions: they had to be used in the development of analysis systems for medical images, and they had to apply different techniques for the detection of points that are considered edges.

3.1 Canny Edge Detector

This algorithm was developed by J. Canny [8] as a response to the following requirements:

- Good detection: maximise the probability of marking an edge where there actually is one, and minimise the probability of marking an edge where there is none.
- Good location: the points of the edge that are indicated by the operator should be as close as possible to the edge.
- Unique response: there should be a unique response for each edge.

This algorithm was developed by J. Canny. 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 must be suppressed: we want 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 thresholding 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.

3.2 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 isotropic linear operators.

This algorithm uses a logical/lineal operator, based on a set of filters in the quadratic phase (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.

In order to eliminate possible ambiguities caused by the use of this operator, we apply a phase of suppression and enhancement; the responses of those image points for which the operator does not present an ideal edge or line are suppressed, whereas the responses that do meet the requirements are improved. The suppression is based on the first derivative of the response module, the enhancement on the second directional derivative of the response module.

Finally, we apply a non-maxima suppression phase on which we build a binary image by using a threshold value. The latter is a configurable parameter of the algorithm.

3.3 Bezdek Fuzzy Edge Detector

The edge detector developed by J. Bezdek [9] consists of four phases:

- An enhancement function that is charged with filtering the image in order to facilitate the analysis by the feature detector.
- A feature detection function that applies various filters to the image, in order to detect the points that may correspond to an edge.
- A composition function that combines the results of the different filters, selecting the points that are part of the image edge.
- A thresholding function, whose purpose is to provide a binar background-edge image by using a threshold value.

The algorithm is based on the analysis of the geometrical characteristics that an edge is supposed to have, the development of feature detection functions that allow us to detect these properties in the image, and finally the analysis of the detectors' result by means of a fuzzy function that selects the candidate points. The latter allow us to introduce a certain learning capacity in the selection of the pixels.

The current implementation of the algorithm uses a Sobel filter, in horizontal and vertical direction, as a function for feature detection. The applied function composition is based on the fuzzy rules system of Takagi-Sugeno [15], with a control parameter that checks the fuzziness of the system; this system builds a binary image of edges on a background through a thresholding value that can be fixed as an algorithm parameter.

4 Results

The study was based on the methodology proposed by Bowyer, in which he starts from wide intervals for each parameter and progressively refines the interval by means of ever smaller increases for the values of the parameters that provide the best results.

The parameters that were modified for each algorithm are the following:

- Canny: sigma of the gaussian, inferior and superior hysteresis threshold,
- Heitger: sigma of the gaussian, enhancement factor and thresholding value.
- Bezdek: control parameter of the fuzzy system and thresholding value.

The results were measured with two criteria: the probability of edge detection and the ROC curves. It is indeed very difficult to reflect with only one criterium all the factors that affect the result and correct the deficiencies of each measurement.

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 1:

$$D = \frac{N_b + N_h}{N} \quad (1)$$

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

Figure 2, 3 and 4 shows the best results of this measure for all the X-Ray images for Canny, Heitger and Bezdek.

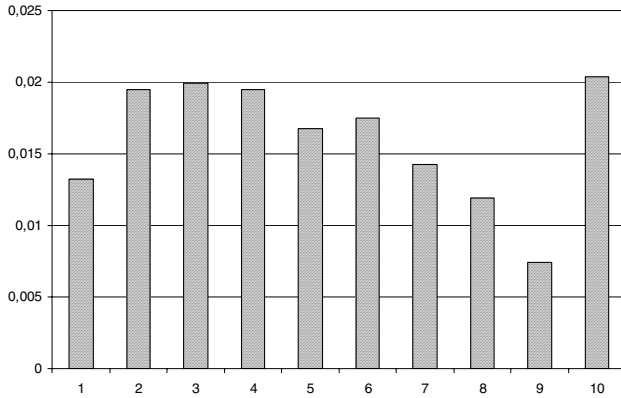


Fig. 2. Discrepancy measurement values for Canny applied to the test set

The most frequently used method to measure the performance of edge detection algorithms are the ROC curves [17]. This method consists in comparing the false positives (points that are erroneously detected as edges) and the true positives (real edge points) in a graphic (Figure 5). If the sampling of the algorithm's parameters space takes place at a sufficiently small interval, we can create a response curve and find the point at which the ratio is optimal, i.e. when the relation between the true and false positives is maximal.

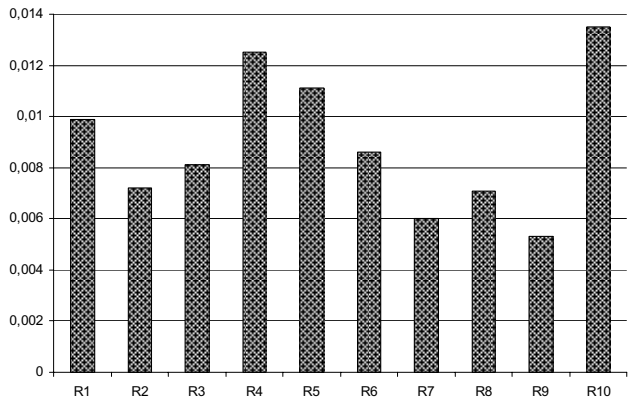


Fig. 3. Discrepancy measurement values for Heitger applied to the test set

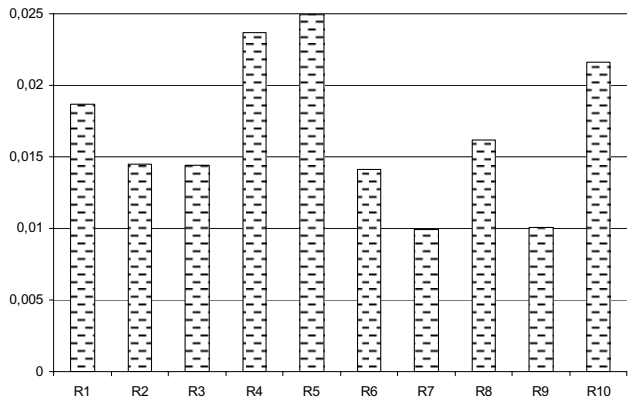


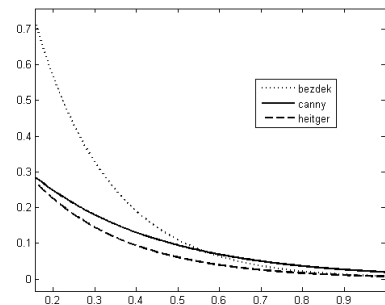
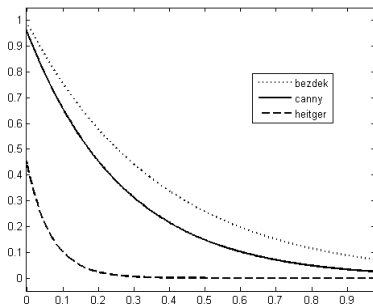
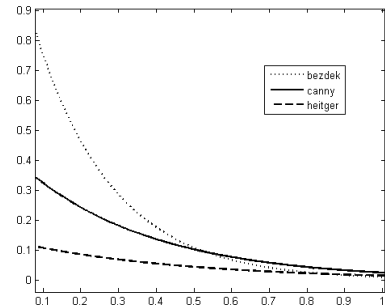
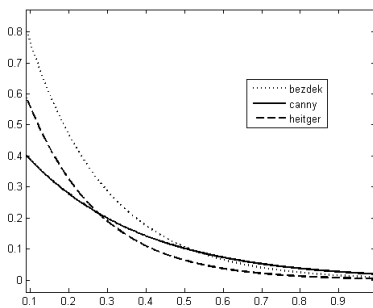
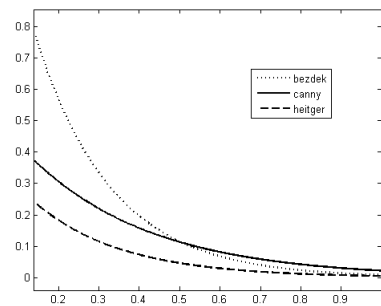
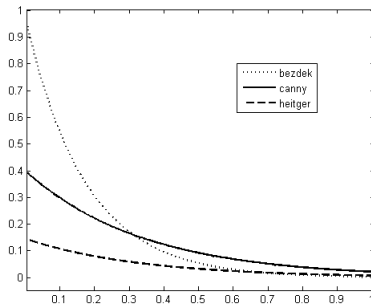
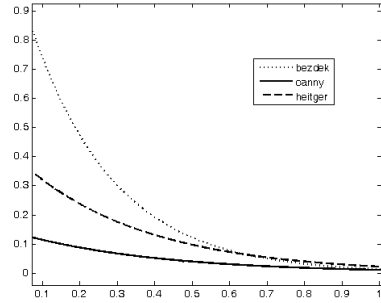
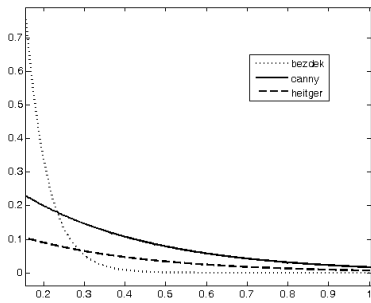
Fig. 4. Discrepancy measurement values for Bezdek applied to the test set

We evaluated the results according to Bowyer’s methodology, contrasting the false negatives against the possible false positives (% not detected, % false). The best algorithm is that whose curve has the smallest area; it is usually calculated with the trapezoidal rule.

4.1 Evaluation by Medical Experts

In this phase, the specialists were asked to evaluate the clarity of the elements of the diagnostic image; they were shown the original image with superposition of the detected edges. The final purpose was to evaluate the usefulness of the resulting images during the diagnosis.

We developed a web application that allows the experts to qualify the images in a fast and comfortable manner. Figure 7, 8, and 9 show the experts’ punctuations for each algorithm in each category and for each image.



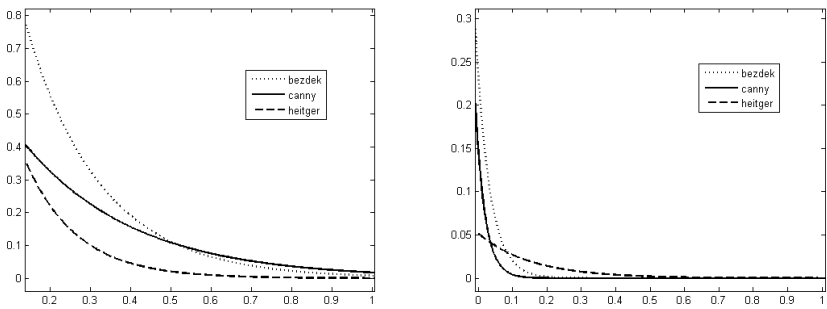


Fig. 5. Curves that represent the different edge detection performances for the X-ray images set. The graphs correspond with the R1-R10 images of left to right and top to down.

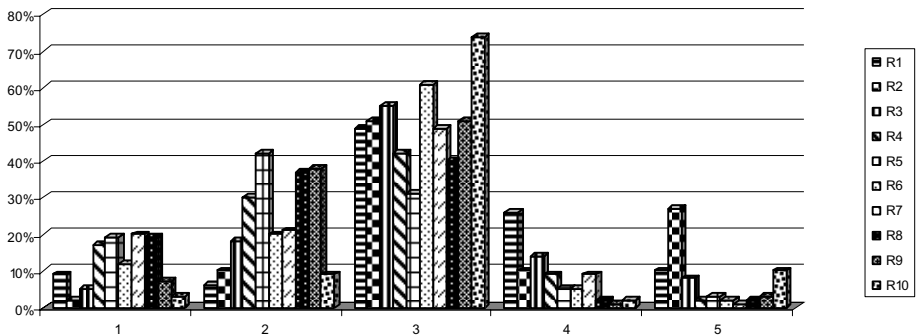


Fig. 6. Qualitative evaluation for Canny

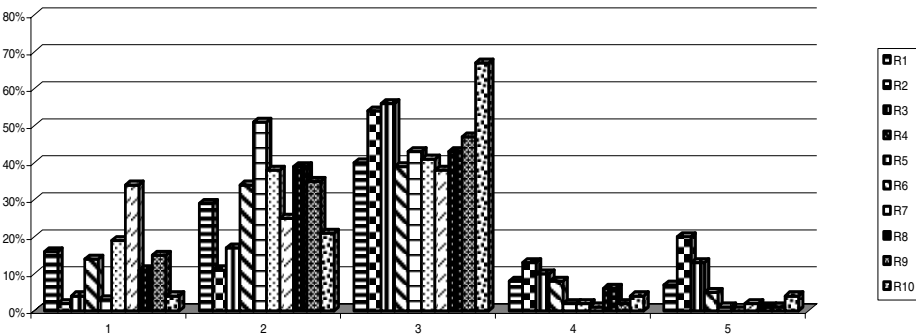


Fig. 7. Qualitative evaluation for Heitger

The evaluation was carried out by four specialists in the field and one resident physician. They worked independently and with anonymous data; there was no contact whatsoever between the evaluators in order to guarantee absolute objectivity.

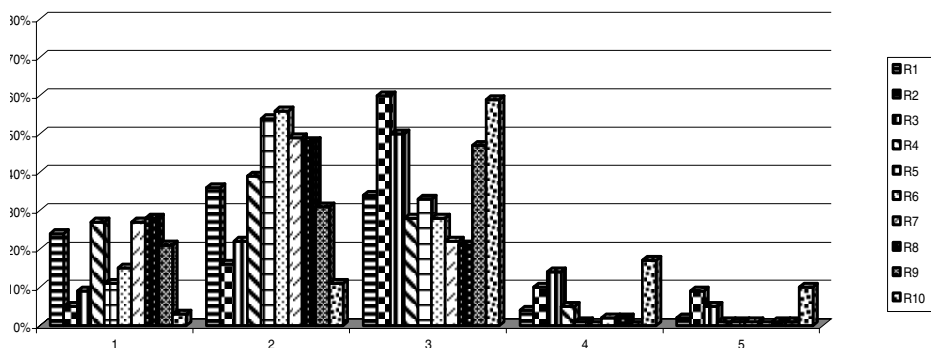


Fig. 8. Qualitative evaluation for Bezdek

Each result was rated with a value between 1 (lowest value) and 5 (highest value). We opted for this interval because the expert can easily associate it with semantic categories (very bad, bad, regular, good, very good).

In order to be considered clinically useful for the specialists, an algorithm must obtain for at least 60% of the results a punctuation of 3 or more. This criterium is based on the assumption that if the punctuation of an image is above 50 %, it is clearly positive. Due to the characteristics of the algorithms and the complexity of the images, the expert will consider a significant part of the results to be of low quality. So if an algorithm obtains an approval percentage that is above 50 %, we can assume that the images provided by the edge detector are really of use to the expert.

5 Conclusions

The study shows that the best results are provided by the Canny and Heitger edge detectors, and that the Bezdek algorithm provides considerably worse results both for the quantitative measures and for the experts. We believe that the main reason for this outcome is the fact that the Bezdek detector uses a Sobel mask in horizontal and vertical for the detection of the candidate points, which provokes the appearance of double edges in the resulting image.

Furthermore, we observe that the Canny algorithm obtains its best results with a low sigma value. High values cause considerable loss of information because of the low signal/noise relationship (mainly in the area where the bone and the iron coincide). The maximal threshold should not have a low value, because too many points are then considered edges, whereas a high minimal threshold makes too many points disappear.

The Heitger algorithm shows the same effect for sigma: since here the most critic parameter is the threshold, small variations in its value create noticeable differences in the result.

The critical parameter for the Bezdek fuzzy edge detector is the control parameter of the fuzzy rules. This means that small variations in its value considerably affect the quality of results.

The quantitative measures by the Heitger and Canny algorithms are very similar but vary according to the analysed image. Also, even though in the analysis of the ROC curves in the global set the Heitger algorithm is more effective, the Canny algorithm provides more stability (interval of the parameters values in which the algorithm gives a result that can be considered good). The experts give the same positive evaluation for both algorithms and their clinical usefulness, but manifest a slight preference for the Canny algorithm.

Although these algorithms do not provide a totally automatic segmentation of the image, we believe that they simplify the edge identification task for the medical expert. We have therefore incorporated both algorithms to the system and are currently trying to determine which algorithm provides the best results on the basis of the image's characteristics.

References

1. Johnson R. C., et al: Clinical and Radiographic Evaluation of Loose Total Hip Reconstruction, a Standard System Terminology for Reporting Results. *Journal of Bone and Joint Surgery*. 72A (1990) 161-168.
2. Downing M.R., Undrill P.E., Ashcroft P., Hutchison J.D., Hukins D.W.L.: Interpretation of Femoral Component in Total Hip Replacement Radiographs. *Image Understanding and Analysis'97*. Oxford, UK (1997) 149-152.
3. Ozanian T.O., Phillips R.: Image analysis for computer-assisted internal fixation of hip fractures. *Medical Image Analysis*. 4 (2000) 137-159.
4. Behiels G., Maes F., Vandermeulen D., Suetens P.: Evaluation of image features and search strategies for segmentation of bone structures in radiographs using Active Shape Models. *Medical Image Analysis*. 6 (2002) 47-62.
5. Pereira J., Castro A., Ronda D., Arcay B., Pazos A.: Development of a System for Access to and Exploitation of Medical Images. *Proceedings of Fifteenth IEEE Symposium on Computer-Based Medical Systems*. Maribor, Eslovenia (2002) 309-314.6. Anonymous.
6. Alonso A., Arcay B., Castro A.: Analysis and Evaluation of Hard and Fuzzy Clustering Segmentation Techniques in Burned Patients Images. *Image and Vision Computing*. 18 (2000) 1045-1054.
7. Pereira J., Castro A., Castro A., Arcay B., Pazos A.: Construction of a System for the Access, Storage and Exploitation of Data and Medical Images Generated in Radiology Information Systems (RIS). *Medical Informatics and Internet in Medicine*. 27 (3) (2002) 203-218.
8. Canny J.: A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 8 (6) (1986) 679-698.
9. Heitger, F.: Detection using Suppression and Enhancement. Technical report n. 163. Image Science Lab, ETH-Zurich (1995).
10. Bezdek J. C., Chandrasekhar R., Attikouzel Y.: A geometric approach to edge detection. *IEEE Transactions on Fuzzy Systems*. 6 (1) (1998) 52-75.
11. Heath M.D., Sarkar S., Sanocki T., Bowyer K.W. A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(12) (1997) 1339-1359.
12. Bowyer K.W., Phillips P.J.: Empirical Evaluation Techniques in Computer Vision. IEEE Computer Press (1998).

13. Kotcheff A.C.W., Redhead A., Taylor C.J., Porter M.L., Hunkins D.W.L.: Shape Model Analysis of THR Radiographs. IEEE Proceedings of ICPR'96. (1996) 391-395.
14. Forsyth D.A., Ponce J.: Computer Vision A Modern Approach. Prentice Hall. (2002).
15. Takagi T., Sugeno M. Fuzzy identification of systems and its application to modeling and control. IEEE Trans. Sys., Man and Cybern 15(1) (1985) 116-132.
16. Lee S.U., Chung S.Y., Park R.H.: A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation. Computer Vision, Graphics and Image Processing. 52 (1990) 171-190.
17. Dougherty S., Bowyer K.W.: Objective Evaluation of Edge Detectors Using a Formally Defined Framework. Empirical Evaluation Techniques in Computer Vision. IEEE Computer Press. (1998) 211-234.