A Crowdsourcing Web Platform - Hip Joint Segmentation by Non-Expert Contributors

Alberto Chávez-Aragón, Won-Sook Lee, Aseem Vyas
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
achavez@uottawa.ca, wslee@uottawa.ca, avyas025@uottawa.ca

***Abstract*—In this paper a crowdsourcing web platform for the interactive segmentation of hip joint structures is introduced. The system collects information on how non-expert volunteers segment anatomical components from MR Images, thereby forming a knowledge base on the solution of this type of problems. The information collected permits to determine tuning parameters for automatic and semi-automatic segmentation approaches, and it provides data for training automatic segmentation algorithms. The findings on the human-computer interaction process can be applied in the design of user interfaces for manual and semi-automatic interactive segmentation tools.**

Keywords—segmentation; MRI; hip joint; medical; imaging; crowdsourcing

I. Introduction

Femoro-Acetabular impingement (FAI) is the leading cause of joint degradation of the hip in the youth [1]. The FAI condition consists in a pathologic contact during hip joint motion between abnormal skeletal prominences of the femur head neck and the acetabulum, causing significant cartilage damage and hip pain. Patients with this condition have their hip range of motion, typically flexion and internal rotation, reduced. In order to correct the FAI condition a corrective surgery is often required [2]. Doctors need to accurately determine the impingement (collision) level to formulate the best surgical plan for each patient [3]. For this purpose Magnetic Resonance Imaging (MRI) is used. MRI produces two-dimensional cross-sectional images of the pelvic region as it is shown in Fig. 1. By combining tomographic images it is possible to generate 3D reconstructions which permit examining bones, tissue, vessels and carbon based flesh among other structures. However, the extraction (segmentation) of useful anatomical structures from MRI and CT scans is still a hard task in medical imaging due to the nature of the data. Image segmentation, which consists in the identification of which part of the image represents the desired structure, supports others tasks such as visualization, measurement and 3D reconstruction of human organs. Image segmentation procedures can be performed fully-automatically, semi-automatically and manually. Nevertheless, the best results are obtained when competent people are involved in the task. Computers cannot equal the human brain in solving this problem yet. Computer science researchers usually emphasize the computational part neglecting the study of the human abilities for the solution of such a problem. Our goal is to better understand how non-expert people can contribute to solving a problem which has remained pending for solution.

In this paper we introduce a Crowdsourcing web platform for collecting manual segmentations of hip joint structures from volunteer solvers. The information collected permits to create a knowledge base for the solution of this and similar problems. And it can be used for finding tuning parameters for automatic and semi-automatic approaches, as well as for training automatic segmentation algorithms. By analyzing the user interaction with the platform we expect to better understand the human-computer interaction process in the context of medical segmentation. Those findings can be applied in the design of user interfaces for manual and semi-automatic interactive segmentation tools.

II. Literature Review

A. Image Segmentation

Automatic techniques for image segmentation typically need the intervention of a human user to initialize the method, evaluate the accuracy or correct the results manually. Automatic methods sometimes fail to segment medical images properly [4]. There are some reasons for this; medical images are likely to contain noise and incomplete data. Besides, most automatic algorithms base their operation in quantifying jumps in the intensity values of the images' pixels. A big jump in a certain area of the image is interpreted by the segmentation algorithms as a border between two regions. However, restrictions imposed by the image acquisition apparatus, frequently makes difficult to determine the right threshold value considered as a big jump. On the other side of the spectrum, manual segmentation of anatomical structures in medical images provides more exact and reliable results. For a trained eye, it is relatively easy to identify different regions in an image and accurately delineate the contours of them. Nevertheless, image segmentation is more often an intermediate stage for measurement, 3D reconstruction and visualization of anatomical structures which requires a reasonable volume of segmented images, resulting in a timeconsuming and tedious chore for the person doing the manual segmentation [5].

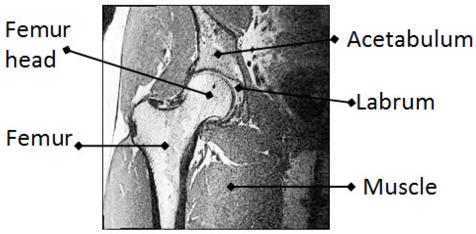


Fig. 1. Hip joint MR Image

Semi-automatic separation of regions in medical images combines the two methods described above. A user chooses a small area belonging to the structure to be segmented. Then, an automatic algorithm grows the small region to occupy all pixels that both share similarities with the user's sample and are connected to it [6] [7]. The result is a polygon enclosing the segmented region. The user is able to refine the result by either changing the shape of the silhouette or by providing the algorithm with feedback so that the method can adjust its parameters and improve the results. This iterative process with feedback loops between the user and the program is run until the desired result is achieved.

B. *Human Computation*

Image Segmentation is a typical problem in which computers have shown poor performance and it is better solved by the intervention of humans. Among problems with similar degree of difficulty the following stand out: optical character recognition, email classification, image labelling, speech recognition, language translation, handwriting recognition and conversational interactions. In recent years, a group of scholars have been experimenting with a different approach to solve this type of problems. Rather than providing computers with formal descriptions (algorithms) for the solutions of such problems, they use computers as a tool for broadcasting a specific problem to a group of human solvers and for collecting and interpreting their solutions, creating this way a symbiotic interaction among humans and computers. This new field was named human computation [8]. Human Computation has its roots in The Open Mind Initiative [9] [10] which provides a framework to collect data from Internet users, and then they are used to train machine learning algorithms. The framework makes available algorithms and data for free. Another very popular crowdsourcing platform is The Amazon Mechanical Turk. This platform permits to coordinate workers that carry out a broad range of tasks that are very difficult for modern computers to match [11]. Currently, the most representative and global level human-based computation application is reCAPTCHA [12] which is an anti-bot service that helps to digitalize books. This service derives from CAPTCHA [13]. CAPTCHA stands for "Completely Automated Public Turing Test to Tell Computers and Humans Apart". CATCHA is a user dialog system that ensures that an entity interacting with a web site is a person. The test is useful for sites like yahoo and Google to prevent automated programs from obtaining free emails accounts or other benefits. It accomplishes its goal by presenting to the user an image containing a series of distorted characters and then asking the user to read and type back the word. People can do this task easily but computers cannot. ReCAPTCHA operates in the same way as CATCHA, by asking the user to decipher distorted characters for telling human and bots apart [14]. But unlike CATCHA, the distorted characters are scanned words from old books that optical character recognition (OCR) algorithms failed to recognize. As a result, this system helps to digitize millions of words with an accuracy that exceeds the 99% as it is claimed by the authors in [8]. Another system that takes advantage of human skills to solve problems that currently are beyond the ability of computers is Duolingo [15]. Duolingo is a free language-learning website and crowd sourced text translation platform. The service helps people to learn languages such as Spanish, English, German, French, Portuguese and Italian. As users progress through the lessons, they simultaneously help to translate websites and other documents. The principal researcher behind this project claims that the platform would be able to translate the English Wikipedia into Spanish in five weeks with 100,000 active users and in 80 hours with a million active users.

III. CROWDSOURCING SEGMENTATION PLATFORM

In this paper, MERCI, a crowdsourcing web platform for the manual segmentation of hip joint structures is introduced. MERCI stands for "MEdical imaging processoR based on Collective Intelligence". The segmentation platform is a web application that permits broadcast image segmentation problems to a group of voluntary solvers. The graphic interface lets users to define the boundaries of anatomical structures by selecting a group of contour points (vertices) from an image. Additionally, it provides means to amend the position of the vertices by dragging and dropping them for better representation of the desired silhouette. MERCI was implemented under the client-server model. The client module was implemented using HTML5 and JavaScript to ensure it functions in the major web browsers and on different platforms, while the server side was implemented in PHP programing language. The general architecture of the system is depicted in Fig. 2.

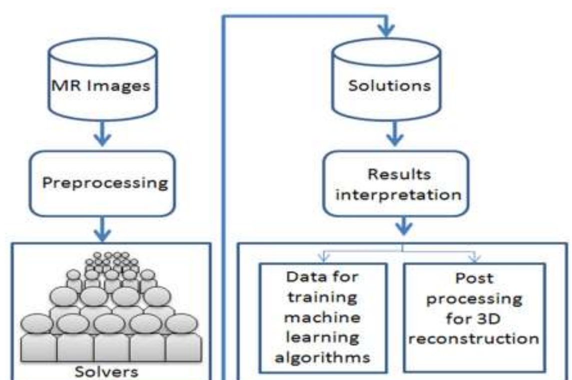


Fig. 2. General Architecture of the System.

The database contains MR Images of acetabulofemoral joint regions. The preprocessing module performs a global contrast enhancement to make easier for the users to distinguish the region of interest. For the experiments reported in this paper, users were asked to segment the femur bone from the acetabulum and the surrounding tissue. The images to be segmented were presented to the user in an arbitrary order. As a way of help, the system presents along with the MR Image a sketch of the bone's shape so that the user can localize the organ in the image as it is shown in Fig.3. It is worth to mention that sketches are chosen from a set of predefined figures corresponding with different organs. And the selection of the sketch is performed by an automatic object detection algorithm that determines which images from MRI studies contain a femur bone. The system offers the user to choose between segmenting the current image and loading a different one. When users complete the segmentation task, the contributions are stored for later analysis. Then, the system provides a new image, the user can stop this process when he/she so wishes. The results interpretation module converts the information into the desired format according to the requirements.

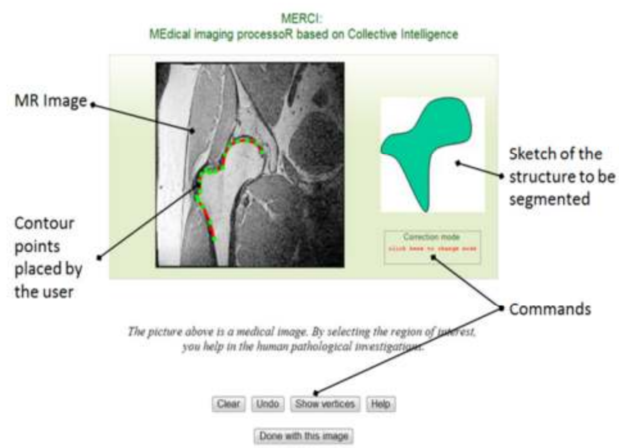
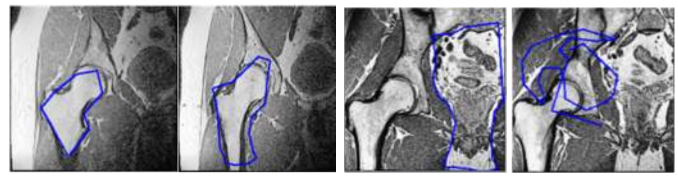


Fig.3. Crowdsourcing web segmentation platform for broadcasting image segmentation problems to a group of voluntary solvers.

IV. Data Analysis

MERCI was released to a closed community of users with no experience in medical imaging from the School of Electrical Engineering and Computer Science at the University of Ottawa. Its database contains *45* images of the hip joint region. In the first two weeks, the system collected *274* segmented images. A total of *11* users during that period of time segmented on average *24.9* images each. After the data collection phase, the database of segmented images was analyzed to verify that the images met the minimum requirements in terms of number of contour points, accuracy and meaningful segmentation. Regarding number of contour points, the check is passed when the number of vertices of a particular contribution exceeds a threshold value set for that image. The threshold values, which were determined experimentally, are equal to the *5%* of the number of vertices calculated by an automatic segmentation algorithm for each image. With respect to meaningful segmentation and accuracy, the verification consist in a) rejecting all the contributions containing open paths and b) rejecting those polygons which are sufficiently dissimilar from other users' segmentations for the same image. The similarity is calculated using the Euclidean distance among feature vectors that contain the geometric center of the polygon and coordinates of a bounding box enclosing the contour points. This method to discard bad segmentations permits the feasibility of the system when large volumes of data are involved. Fig. 4 shows images that failed to satisfy the inclusion criteria. After cleaning up the data *231* images remained which represent the *84.3*% of the total. That means that each image, on average, was segmented *5.1* times. Fig. 5 shows the actual frequency in which each image was segmented. As can be noted from the chart after removing those images that do not make any contribution in terms of content the distribution of frequencies remains almost the same.



a b c d

Fig. 4. Some users' contributions were eliminated due to the following reasons: a) Insufficient number of contour points, b) and c) Imprecise segmentation, d) Nonsensical segmentation.

Users defined the contour of the femur by using a different number on contour points in the interval [*13, 114*]. With a median of *35.15* and a standard deviation of *14.62* as it is shown in Fig. 6.

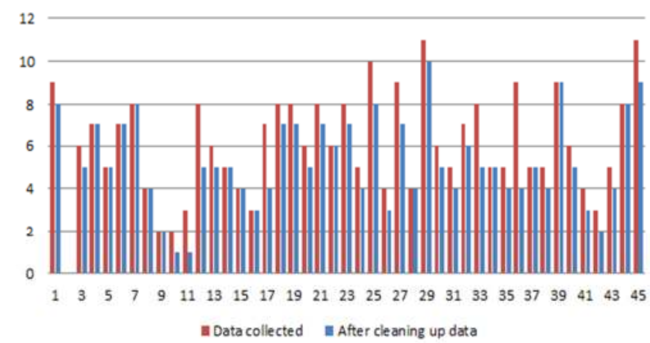


Fig. 5. Segmentation frequency of each image from the database before and after the cleaning up process.

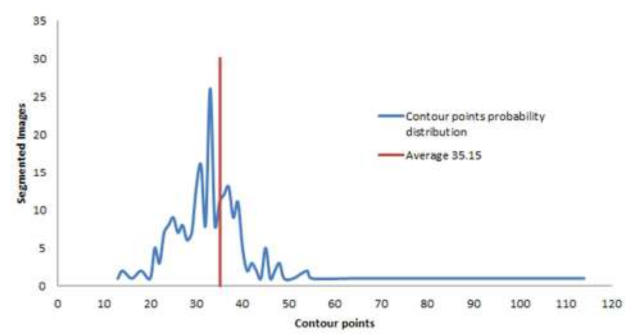


Fig. 6. Probability distribution for the number of contour points used by the volunteers to segment the femur bone.

Next, we performed a qualitative comparison of femur segmentations provided by non-expert contributors and by a fully automatic segmentation algorithm tailored for the femur segmentation problem. Our segmentation algorithm separates the femur area from the background as follows. First, a small area of the image belonging to the femur is taken as a sample of the type of texture we are seeking. This job is performed by the Hough transform algorithm for circles which takes advantage of the fact that the femoral head is circular in shape. Then, the normalized histogram of the sample is calculated. The histogram can be seen as a probability function that provides the likelihood that a pixel in the image belongs to the femur region. By replacing the intensity value for each pixel with its corresponding probability value (back projection) we obtain a probability map as it is shown in Fig. 7a where darker areas are more likely to belong to the femoral region. After that, a threshold is applied over the probability map resulting in a black and white image, where the density of white points is greater in the desired region. Then, morphological operators are applied to conect high density regions and to eliminate low density ones (see Fig. 7b). After that, a labelling algorithm isolates the femur region as Fig. 7c depicts. Finally, the outline of the object is calculated. Some of the functions used in the implementation of the algorithm described above are from the openCV library [16].

a b c

Fig. 7 a) probability map for a texture sample collected from the femoral head area, b) binary image after applying morfological operators, c) segmented image.

It is worth mentioning that despite the simplicity of the segmentation method, it provides acceptable results for the mayority of the images in the database. Regarding manual segmentation, efforts were made to ensure that the segmentations had a enough contour points to be compared with the results from the automatic segmentation method. Thus, for images that were manually segmented with a small number of vertices, we combined the contour points provided by more that one solver. The later, results in a set of polygons with a wide range of sizes. Due to the degree of variability of the number of contour points of both: the manual segmentation and the automatic one, we implemented a strategy to equalize the number of vertices preserving their visual characteristics so that we can perform a fair comparison. For evaluation purposes, we chose as a number of vertices the lesser of the two values. The strategy to reduce the number of vertices while the visual features are preserved is as follows. Given a polygon, a significant measure K for each vertex was calculated using equation (1).

$$K(S_1, S_2) = \frac{\beta(S_1,S_2)\, l(S_1)\, l(S_2)}{l(S_1) + l(S_2)} \qquad (1)$$

Where $\beta(S_1, S_2)$ is the turn angle at the common vertex of the segments S_1, S_2 and l is the length function normalized with respect to the perimeter of the silhouette. The lower the value of K, the less the contribution to the contour is provided by the shared vertex of segments S_1 and S_2. So, we simplified the curve by removing the least important vertices according to K. Once a vertex is removed, its neighboring vertices are connected. This process is repeated until obtaining the desired outline simplification. Fig. 8 shows the application of the curve evolution strategy for a polygon that orignailly contained *379* vertices. It is possible to see that the visual features remain the same during the evolution process.

Fig. 8 A silhouette shown in various evolutional stages. From left to right the polygons contain 379, 50 and 15 vertices respectively.

Fig. 9, on the right, shows the result of a femur bone segmented by the automatic segmentation technique and on the right the combination of the results from voluntary solvers. Both segmentations contain the same number of contour points after the curve evolution strategy was applied for both images.

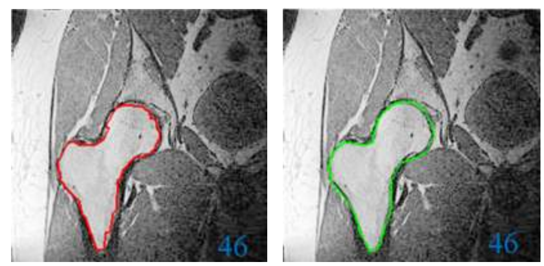
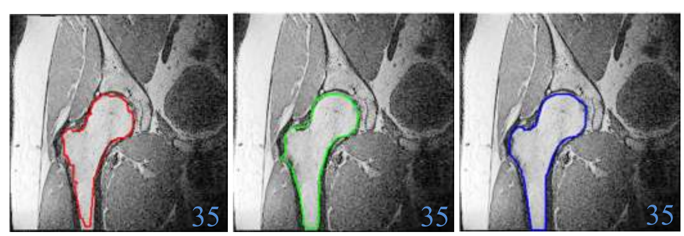


Fig. 9 Comparison of two polygons of 46 vertices, on the left automatic segmentation on the right a colaborative manual segmentation.

An additional experiment is shown in Fig. 10, where three curves, obtained from different techniques, were simplified so that they contain 35 contour points each. Fig. 10a shows the curve calculated by the automatic segmentation algorithm described above. Fig. 10b is the corresponding colaborative segmentation collected by the proposed web platform, and Fig. 10c shows the result of applying grabcut, a semi-automatic segmentation technique.



a b c

Fig. 10 Comparison of three polygons of 35 vertices, a) automatic segmentation, b) collaborative segmentation collected by the proposed web platform , c) semi-automatic segmentation produced by the grabcut segmentation technique.

The openCV grabcut function produces more accurate results that the fully automatic algorithm, but the downside is that it requires the interventention of the user to label some points belonging to the foreground and background along with a bounding box enclosing the object of interest. The algorithm permits to refine the segmentation in an iterative manner using the user's feedback. The image shown in Fig. 10c was obtained after five iterations. On the basis of the foregoing, we can say that the collaborative segmentation method produces more accurate results that the fully automatic technique and comparable with those obtained by using a much more sophisticated semi-automatic segmentation method such as grabcut. Lastly, an analysis is presented below on how the users tackled the segmentation problem in terms of the predilection of certain regions of the images to start the job and the direction in which the polygons were constructed. It is known that sharp corners help to keep the visual features of an outline. The significant measure *K* for vertices at tight bends is higher than other vertices. After an analysis on which vertices were set first, we realized that users frequently chose as first and last vertices those with the high values of *K* as Fig. 11 shows.

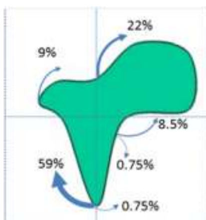


Fig. 11. Users select as a main contour point those vertices with the highest values of *K*.

The diagram above illustrates that *59%* of time users preferred to star segmenting the bone by choosing as a first vertex the most significant contour point in terms of visual characteristics (maximum values of *K)* or another point closed to it. *90.75%* of the time users defined the contour moving clockwise. *59.75%* of the segmentations were done from the button up. If we consider that the fist vertex as the main contour point, this experiment suggests that users select as main point one which is close to the vertices with high values of visual features.

V. Conclusions and Future Work

In this paper, a collaborative web platform for the segmentation of MR Images was proposed. The platform aims to collect examples of segmentations for the hip joint region and to shed light on how non-expert users utilize a platform like this to segment anatomical components from MRI. The collected information forms a knowledge base for the solution of medical imaging segmentation problems. The analysis of the data can be useful for determining tuning parameters for automatic and semi-automatic segmentation approaches. Additionally, the series of experiments provides useful information that should be taken into account when designing semi-automatic interactive segmentation tools. Currently, the platform continues to collect data on segmentation of femur for further analysis. In the near future, the system will start collecting information on the task of segmenting anatomical structures such as acetabulum, labrum and muscles.

Acknowledgments

The authors acknowledge the support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR) to this research via the CHRP initiative project CHRPJ/ 398837-2011, as well as the collaboration of the students from the School of EECS at the University of Ottawa who participated in this study.

References

[1] R. Ganz, J. Parvizi, M. Beck, M. Leunig, i. H. Notzl and K. Siebenrock, "Femoroacetabular impingement: A cause for osteoarthritis of the hip," *Clinical Orthopaedics and Related Research,* no. 417, pp. 112-120, 2003.

[2] S. Wisniewski and G. B., "Femoroacetabular impingement: An overlooked cause of hip pain," *American Journal of Physical Medicine and Rehabilitation,* no. 85, pp. 546-549, 2006.

[3] D. Cai, W. Lee, C. Joslin and P. Beaulé, "Rapid Impingement Detection System with Uniform Sampling for Ball-and-Socket Joint," *Recent Advances in the 3D Physiological Human,* pp. 179-192, 2009.

[4] S. D. Olabarriaga and a. W. Smeulders, "Interaction in the segmentation of medical images: a survey," *Medical image analysis,* vol. 142, no. 2, p. 127, 2001.

[5] A. Z. Reza, S. Yoshinobu, S. Toshihiko, N. Takashi, S. Nobuhiko and Y. Kazuo, "Automated Segmentaiton of Acetabulum and Femoral Head From 3-D CT Images," *IEEE Transactions on Information Technology in Biomedicine,* vol. 7, no. 4, pp. 329-343, 2003.

[6] L. Álvarez, L. Baumela, P. Henríquez and M.-N. P., "Morphological Snakes," in *Computer Vision and Pattern Recognition*, San Francisco, 2010.

[7] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE transactions on medical imaging,* vol. 22, no. 2, pp. 137-154, 2003.

[8] E. Law and v. A. Luis, Human Computation, Morgan & Claypool Publishers, 2011, p. 121.

[9] D. G. Stock, "The Open Mind Initiative," in *IEEE Intelligent Systems and Their Applications*, 1999.

[10] D. G. Stork and L. C. P., "Open Mind Animals: Ensuring the quality of data openly contributed over the World Wide Web," in *AAAI Workshop on Learning with Imbalanced Data Sets*, 2000.

[11] Amazon, "Amazon Mechanical Turk," 2005. [Online]. Available: https://www.mturk.com/mturk/welcome. [Accessed 11 December 2012].

[12] "reCAPTCHA," 2012. [Online]. Available: http://www.google.com/recaptcha. [Accessed 10 December 2012].

[13] L. vonAhn, M. Blum and J. Langford, "Telling humans and computers apart automatically," *Commun. ACM,* vol. 47, no. 2, pp. 56-60, February 2004.

[14] L. vonAhn, B. Maurer, C. McMillen, D. Abraham and M. Blum, "reCAPTCHA: Human-Based Character Recognition via Web Security Measures," *Science,* pp. 1465-1468, 14 August 2008.

[15] "Duolingo," 2012. [Online]. Available: http://duolingo.com. [Accessed 02 December 2012].

[16] OpenCV, "Open Source Computer Vision Library," 2012. [Online]. Available: http://opencv.org/. [Accessed September 2012].