Medical Image Segmentation Based on Mutual Information Maximization

Jaume Rigau, Miquel Feixas, Mateu Sbert, Anton Bardera, and Imma Boada

Institut d'Informatica i Aplicacions, Universitat de Girona, Spain
{jaume.rigau,miquel.feixas,mateu.sbert,anton.bardera,imma.boada}@udg.es

Abstract. In this paper we propose a two-step mutual information-based algorithm for medical image segmentation. In the first step, the image is structured into homogeneous regions, by maximizing the mutual information gain of the channel going from the histogram bins to the regions of the partitioned image. In the second step, the intensity bins of the histogram are clustered by minimizing the mutual information loss of the reversed channel. Thus, the compression of the channel variables is guided by the preservation of the information on the other. An important application of this algorithm is to preprocess the images for multimodal image registration. In particular, for a low number of histogram bins, an outstanding robustness in the registration process is obtained by using as input the previously segmented images.

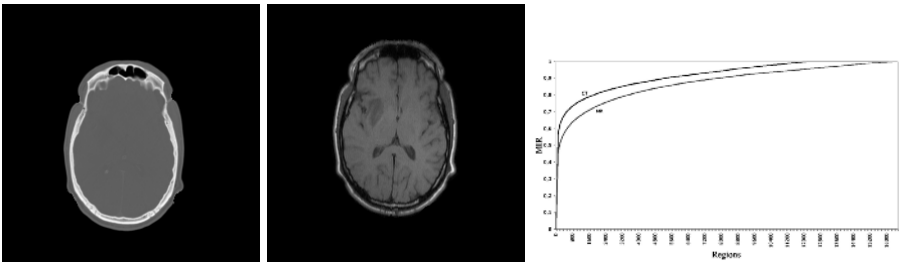
1 Introduction

In image processing, grouping parts of an image into units that are homogeneous with respect to one or more features results in a segmented image. Thus, we expect that segmentation subdivides an image into constituent regions or objects, a significant step towards image understanding. The segmentation problem is very important in clinical practice, mainly for diagnosis and therapy planning.

In this paper we introduce a new algorithm for medical image segmentation based on mutual information (MI) optimization of the information channel between the histogram bins and the regions of the partitioned image. The first step of the algorithm partitions an image into relatively homogeneous regions using a binary space partition (BSP). The second step clusters the histogram bins from the previously partitioned image. This algorithm provides us with a global segmentation method without any human interaction. Our approach has similar characteristics to the agglomerative information bottleneck method [6] applied to document clustering.

An important application of the previous algorithm is to use the segmented images in the registration process. This allows for an extremely robust and very fast registration. Multimodal image registration is a fundamental task in medical image processing since it is a necessary step towards the integration of information from different images of the same or different subjects. Results obtained from different image modalities show good behavior of our segmentation algorithm and resulting segmented images perform well in medical image registration using mutual information-based measures.

C. Barillot, D.R. Haynor, and P. Hellier (Eds.): MICCAI 2004, LNCS 3216, pp. 135–142, 2004.



(*a*) MR (*b*) CT (*c*) MIR_p

Fig. 1. Test images: (*a*) MR and (*b*) CT. The two plots in (*c*) show the MIR_p with respect to the number of regions for (*a*) and (*b*).

2 Information Theoretic Tools

Some of the most basic information theoretic concepts [3] are presented here. The *Shannon entropy* $H(X)$ of a discrete random variable X with values in the set $\mathcal{X} = \{x_1, \ldots, x_n\}$ is defined as $H(X) = -\sum_{i=1}^{n} p_i \log p_i$, where $n = |\mathcal{X}|$ and $p_i = Pr[X = x_i]$. Shannon entropy expresses the average information or uncertainty of a random variable. If the logarithms are taken in base 2, entropy is expressed in bits. If we consider another random variable Y with probability distribution q, corresponding to values in the set $\mathcal{Y} = \{y_1, \ldots, y_m\}$, the *conditional entropy* is defined as $H(X|Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} p_{ij} \log p_{i|j}$ and the *joint entropy* is defined as $H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log p_{ij}$, where $m = |\mathcal{Y}|$, $p_{ij} = Pr[X = x_i, Y = y_j]$ is the joint probability, and $p_{i|j} = Pr[X = x_i|Y = y_j]$ is the conditional probability. Conditional entropy can be thought of in terms of an *information channel* whose input is the random variable X and whose output is the random variable Y. $H(X|Y)$ corresponds to the uncertainty in the information channel input X from the point of view of receiver Y, and vice versa for $H(Y|X)$. In general, $H(X|Y) \neq H(Y|X)$ and $H(X) \geq H(X|Y) \geq 0$.

The *mutual information* between X and Y is defined as

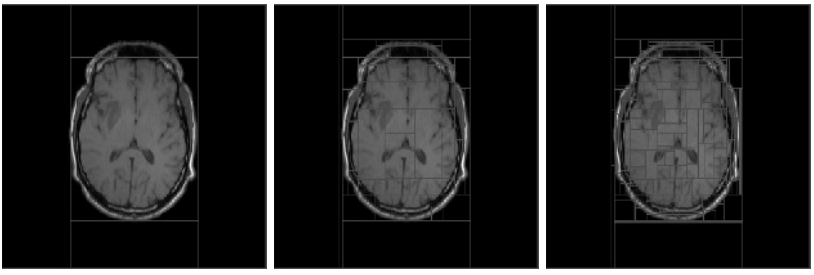
$$I(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i q_j}. \tag{1}$$

It can also be expressed as $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$ and is a measure of the *shared information* between X and Y.

A fundamental result of information theory is the *data processing inequality* which can be expressed in the following way: if $X \to Y \to Z$ is a Markov chain, i.e., $p(x,y,z) = p(x)p(y|x)p(z|y)$, then

$$I(X,Y) \geq I(X,Z). \tag{2}$$

This result demonstrates that no processing of Y, deterministic or random, can increase the information that Y contains about X.



(*a*) MIR_p=0.2 (*b*) MIR_p=0.4 (*c*) MIR_p=0.5

Fig. 2. Partition of the MR image of Fig.1.a with three different MIR_p.

3 Two-Step Segmentation Algorithm

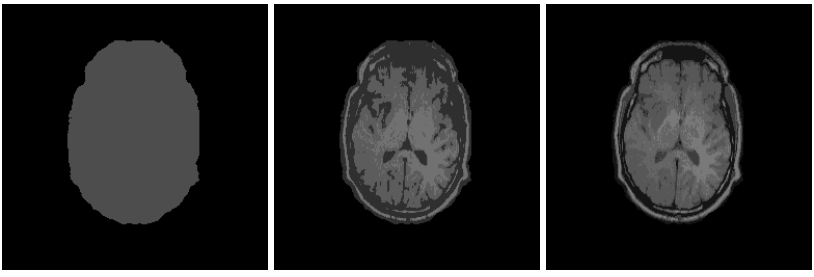
In this section, we present a general purpose global two-step segmentation algorithm that can be applied to different medical image modalities.

Given an image with N pixels and an intensity histogram with n_i pixels in bin i, we define a discrete information channel $X \rightarrow Y$ where X represents the bins of the histogram, with marginal probability distribution $\{p_i\} = \{\frac{n_i}{N}\}$, and Y the pixel-to-pixel image partition, with uniform distribution $\{q_j\} = \{\frac{1}{N}\}$. The conditional probability distribution $\{p_{j|i}\}$ of this channel is given by the transition probability from bin i of the histogram to pixel j of the image, and vice versa for $\{p_{i|j}\}$. This channel fulfills that $I(X,Y) = H(X)$ since, given a pixel, there is no uncertainty about the corresponding bin of the histogram. From the data processing inequality (2), any clustering or quantization over X or Y, respectively represented by $\widehat{X}$ and $\widehat{Y}$, will reduce $I(X,Y)$. Thus, $I(X,Y) \geq I(X,\widehat{Y})$ and $I(X,Y) \geq I(\widehat{X},Y)$.

3.1 Image Partitioning

The first step of the algorithm is a greedy top-down procedure which partitions an image in quasi-homogeneous regions. Our splitting strategy takes the full image as the unique initial partition and progressively subdivides it with vertical or horizontal lines (BSP) chosen according to the maximum MI gain for each partitioning step. Note that other strategies, as a quad-tree, could be used, obtaining a varied polygonal subdivision. This partitioning process is represented over the channel $X \longrightarrow \widehat{Y}$. Note that this channel varies at each partition step because the number of regions is increased and, consequently, the marginal probabilities of $\widehat{Y}$ and the conditional probabilities of $\widehat{Y}$ over X also change. Similar algorithms were introduced in the context of pattern recognition [5], learning [4], and DNA segmentation [1].

The partitioning algorithm can be represented by a binary tree [5] where each node corresponds to an image region. At each partitioning step, the tree acquires information from the original image such that each internal node i



(*a*) C=2 (*b*) C=4 (*c*) C=6

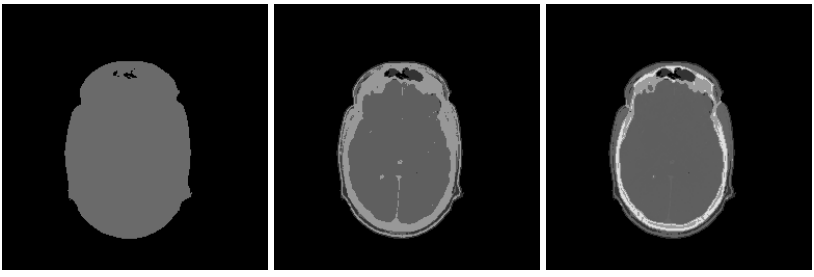
Fig. 3. MR image segmentations obtained respectively from the partitioned images of Fig.2, indicating the number of colors C chosen in each case. The efficiency coefficients α are, respectively, 0.25, 0.22, and 0.22.

contains the mutual information I_i gained with its corresponding splitting. The total $I(X,\widehat{Y})$ captured by the tree can be obtained adding up the MI available at the internal nodes of the tree weighted by the relative area $q_i = \frac{N_i}{N}$ of the region i, i.e., the relative number of pixels corresponding to each node. Thus, the total MI acquired in the process is given by $I(X,\widehat{Y}) = \sum_{i=1}^{T} \frac{N_i}{N} I_i$, where T is the number of internal nodes. It is important to stress that this process of extracting information enables us to decide locally which is the best partition. This partitioning algorithm can be stopped using different criteria: the ratio $MIR_p = \frac{I(X,\widehat{Y})}{I(X,Y)}$ of mutual information gain, a predefined number of regions R, or the error probability [3,5]. The partitioning procedure can also be visualized from equation $H(X) = I(X,\widehat{Y}) + H(X|\widehat{Y})$, where the acquisition of information increases $I(X,\widehat{Y})$ and decreases $H(X|\widehat{Y})$, producing a reduction of uncertainty due to the progressive homogenization of the resulting regions. Observe that the maximum MI that can be achieved is $H(X)$.

Figure 1 shows the two test images used in our experiments. The two plots in Fig. 1.c indicate the behavior of MIR_p with respect to the number of partitions. Both plots show the concavity of the MI function. It can be clearly appreciated that a big gain of MI is obtained with a low number of partitions. Thus, for instance, 50% of MI is obtained with less than 0.5% of the maximum number of partitions. Observe that in the CT image less partitions are needed to extract the same MIR_p than in the MR image. Figure 2 presents the results of partitioning the MR test image. We show the partitioned images corresponding to three different MIR_p. Observe that the first partitioned image (Fig. 2.a) only separates the brain structure from the background.

3.2 Histogram Quantization

The second step of the algorithm is a greedy bottom-up segmentation procedure which takes as input the previously obtained partition and results in a histogram clustering based on the minimization of the loss of MI.

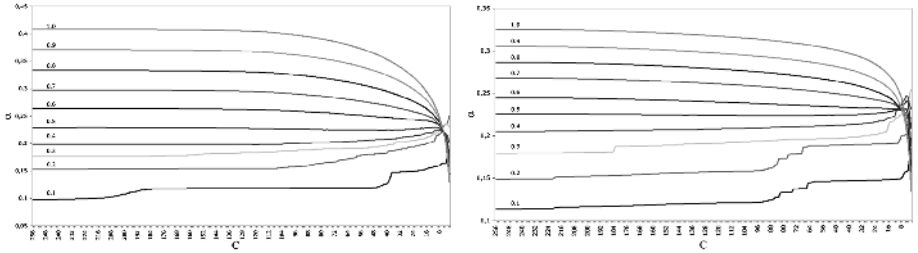


(a) C=2 (b) C=4 (c) C=8

Fig. 4. CT image segmentations obtained from a partition of Fig.1.b (with MIR_p=0.5), indicating the number of colors. The coefficients α are, respectively, 0.21, 0.24, and 0.23.

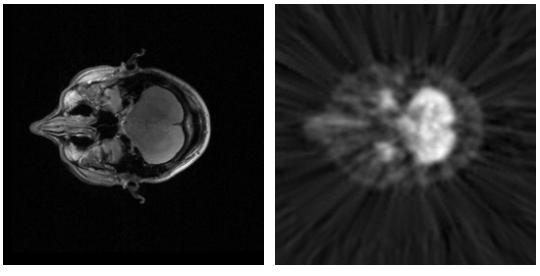
The basic idea underlying our quantization process is to capture the maximum information of the image with the minimum number of colors (histogram bins). The clustering of the histogram is obtained efficiently by merging two neighbor bins so that the loss of MI is minimum. The stopping criterion is given by the ratio $MIR_q = \frac{I(\widehat{X},\widehat{Y})}{I(X,\widehat{Y})}$, a predefined number of bins C, or the error probability P_e [3,5]. Our clustering process is represented over the channel $\widehat{Y} \longrightarrow \widehat{X}$. Note that this channel changes at each clustering step because the number of bins is reduced. At the end of the quantization process, the following inequality is fulfilled: $I(X,Y) \geq I(X,\widehat{Y}) \geq I(\widehat{X},\widehat{Y})$.

In our experiments, the quality of the segmentation is measured by the coefficient of efficiency $\alpha = \frac{I(\widehat{X},\widehat{Y})}{H(\widehat{X},\widehat{Y})}$ [2]. In Fig. 3 we show three MR segmented images obtained from the corresponding partitions of Fig.2 with the indicated number of colors. Also, Fig. 4 shows three CT segmented images (2, 4, and 8 colors) built from a partition with MIR_p=0.5. Graphs plotting α versus the number of bins are presented in Fig. 5, showing a similar behavior in the two cases. These graphs correspond to a sequence of partitioned images of the MR and CT images of Fig. 1. It is important to observe from Fig. 5 that the images with a high number of partitions (with MIR_p from 0.6 to 1) have a decreasing efficiency, while the images with a low number of partitions (from 0.1 to 0.4) present a maximum of efficiency for a low number of bins. Observe also that MR and CT plots in Fig. 5 show a relatively stable α value for a partitioned image with MIR_p=0.5. In all performed experiments, an interesting pattern has been found. For each image, all MIR curves cross at a common point (see Fig. 5). This can be interpreted as an intrinsic property of the image. The number of bins at the crossing point might correspond inversely to the segmentation complexity of the image. For instance, in the plots (a) and (b) of Fig. 5, we have values 6 and 9 respectively, which reflects the higher complexity of the MR image.



(*a*) (*b*)

Fig. 5. Plots of α corresponding to a sequence of partitioned images (with MIR_p from 0.1 to 1.0) for the (*a*) MR (Fig. 1.a) and (*b*) CT (Fig.1.b) images.



(*a*) MR (*b*) PET

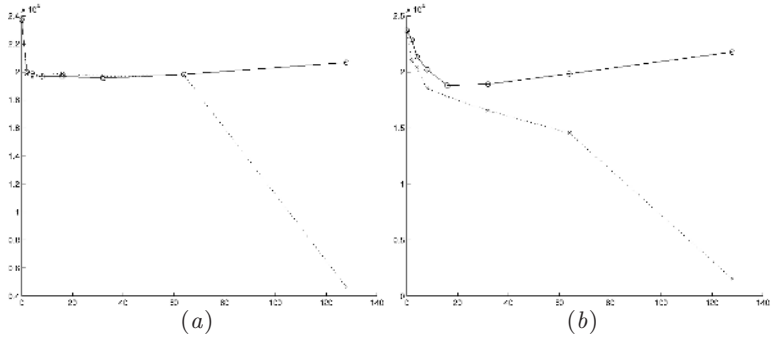
Fig. 6. Test images: (*a*) MR and (*b*) PET.

4 Application to Image Registration

In this section, segmented images obtained with the MI-based segmentation algorithm are applied to medical rigid registration. The registration measure criterion used is the normalized MI (NMI) introduced by Studholme et al.[7]:

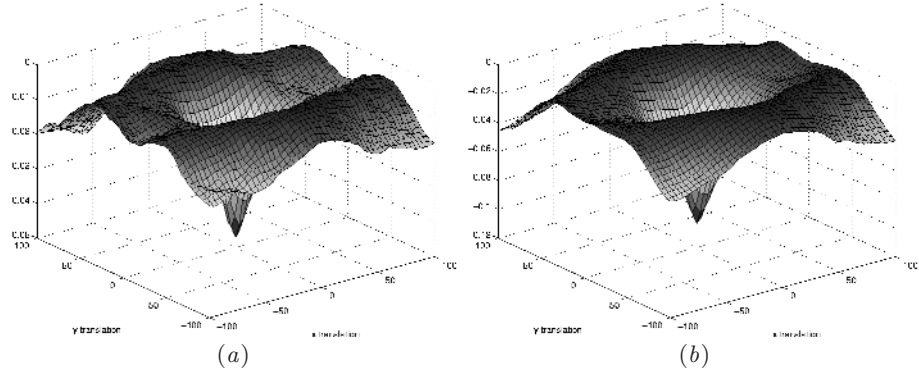
$$NMI(X,Y) = \frac{H(X)+H(Y)}{H(X,Y)} = 1 + \frac{I(X,Y)}{H(X,Y)}. \tag{3}$$

Experiments on MR-CT and MR-PET images are presented, showing a high robustness of NMI for a low number of histogram bins. The robustness of NMI has been evaluated in terms of the partial image overlap. This has been done using the parameter AFA (Area of Function Attraction) introduced by Čapek et al.[8] This parameter evaluates the range of convergence of a registration measure to its global maximum, counting the number of pixels, i.e. x-y translations in image space, from which the global maximum is reached by applying a maximum gradient method. The higher the AFA, the wider the attraction basin of the measure is.



(*a*) (*b*)

Fig. 7. AFA values obtained from the registration of (*a*) MR-CT (Fig. 1) and (*b*) MR-PET (Fig. 6) images using uniform (dotted line) and MI-based (solid line) quantizations. The x-axis represents the size of the bins.



(*a*) (*b*)

Fig. 8. Attraction basins of NMI obtained with 4 bins for the MR-PET (Fig. 6) registration with (*a*) uniform and (*b*) MI-based quantizations.

The robustness of NMI is analyzed with respect to the uniform quantization of the histogram and the quantization given by the MI-based segmentation algorithm. In Fig. 7, we show the AFA values obtained from the registration of the MR-CT pair of Fig. 1 and the MR-PET of Fig. 6, where the bin size is given by the successive powers of 2, going from 1 to 128. In these experiments, we have quantized only the MR image, partitioned with MIR_p=0.5, although similar results are obtained by quantizing both images. Figure 8 shows the attraction basins of NMI obtained with only 4 bins (bin size = 64) for the MR-PET images with a uniform quantization (Fig. 8.a) and a MI-based quantization (Fig. 8.b). Note the smoothness of the basin obtained with the MI-based segmentation. From these results, we can observe that the use of the segmented images provides us with a high robustness. For a high number of bins, the uniform quanti-

zation has a positive effect since it basically reduces the noise of the images. On the other hand, a negative effect is obtained when a very coarse quantization is applied without an information preservation criterion. In contrast, while the MI-based segmentation does not improve the uniform quantization for a high number of bins, it presents a remarkable improvement for a very coarse quantization. This is basically due to the fact that the segmentation is guided by an information optimization criterion.

5 Conclusions

We have presented in this paper a general purpose two-step mutual information-based algorithm for medical image segmentation, based on the information channel between the image intensity histogram and the regions of the partitioned image. We have applied this algorithm to preprocess the images for multimodal image registration. Our experiments show that, using as input the segmented images, an outstanding robustness is achieved in the registration process for a low number of histogram bins. In the future we will compare our segmentation algorithm against other state-of-the-art methods, and analyze two open problems: the optimal partition and the optimal number of clusters.

References

1. Pedro Bernaola, José L. Oliver, and Ramón Román. Decomposition of DNA sequence complexity. *Physical Review Letters*, 83(16):3336–3339, October 1999.
2. Torsten Butz, Olivier Cuisenaire, and Jean-Philippe Thiran. Multi-modal medical image registration: from information theory to optimization objective. In *Proceeding of 14th International Conference on Digital Signal Processing (DSP'02)*, July 2002.
3. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.
4. Sanjeev R. Kulkarni, Gábor Lugosi, and Santosh S. Venkatesh. Learning pattern classification – a survey. *IEEE Transactions on Information Theory*, 44(6):2178–2206, 1998.
5. Ishwar K. Sethi and G.P.R. Sarvarayudu. Hierarchical classifier design using mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(4):441–445, July 1982.
6. Noam Slonim and Naftali Tishby. Agglomerative information bottleneck. In *Proceedings of NIPS-12 (Neural Information Processing Systems)*, pages 617–623. MIT Press, 2000.
7. Colin Studholme. *Measures of 3D Medical Image Alignment*. PhD thesis, University of London, London, UK, August 1997.
8. Martin Čapek, Lukas Mrož, and Rainer Wegenkittl. Robust and fast medical registration of 3D-multi-modality data sets. In *Proceedings of the International Federation for Medical and Biological Engineering*, volume 1, pages 515–518, June 2001.