Mutual information based binarization of multiple images of an object: An application in medical imaging

1Yaniv Gal 2Andrew Mehnert 1Stephen Rose
1Stuart Crozier
February 4, 2013

1Centre for Medical Diagnostic Technologies in Queensland, The University of Queensland, QLD, Australia, 4072.
2Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden.

Corresponding author: Yaniv Gal, email: ygal@itee.uq.edu.au

Abstract

A new method for image thresholding of two or more images that are acquired in different modalities or acquisition protocols is proposed. The method is based on measures from information theory and has no underlying free parameters nor does it require training or calibration. The method is based on finding an optimal set of global thresholds, one for each image, by maximizing the mutual information above the thresholds while minimizing the mutual information below the thresholds. Although some assumptions on the nature of images are made, no assumptions are made by the method on the intensity distributions or on the shape of the
image histograms. The effectiveness of the method is demonstrated both on synthetic images and medical images from clinical practice. It is then compared against three other thresholding methods.

1 Introduction

The goal of image segmentation is to differentiate between objects and background [13]. More specifically it involves partitioning the support of an image into subsets each of which corresponds to an object or to the background. When both the background and objects have distinct ranges of gray-levels then segmentation can be achieved using gray-level thresholding (i.e. image binarization). This essentially involves partitioning the gray-level histogram, either globally or locally, such that each partition corresponds to an object or the background. In the simplest case the histogram is bimodal with one peak corresponding to the background and the other to the objects. A suitable threshold value then lies somewhere between the two peaks. The result is a binary image (also called a binary mask) where object pixels are assigned one binary state (e.g. 1) and background pixels are assigned the other.

Numerous algorithms have been devised for automatically locating the threshold value. A survey of bi-level thresholding methods presented in [11] concluded that no single thresholding method can perform well on all images, even for a single application type. A more recent review of medical image segmentation techniques [7] concluded that every segmentation algorithm “has its suitable application field”. Nevertheless, the majority of these algorithms require different parameter tuning for each application and sometimes for different sets of images (e.g. acquisition protocols) of the same application.

There are many situations where multiple variables are available for each pixel. Some examples are multispectral images (remote sensing), co-registered
medical images from different modalities (e.g. Computerized Tomography (CT), Positron Emission Tomography (PET), Ultrasound, Magnetic Resonance Imaging (MRI)), and multiple focal planes of a given field-of-view, acquired on a light microscope. Several multivariable thresholding methods have been devised [5, 2, 1]. These usually seek a threshold or a set of thresholds that will maximize the amount of entropy or mutual information above and below the thresholds. The underlying assumption is that the foreground, and possibly the background has some type of similarity in different images. However, in some environments, such as multi-modal medical imaging, this assumption is not valid. Furthermore, the background in images acquired from different modalities often has completely different properties, including noise models and acquisition artefacts.

Gray-level thresholding is essentially based on a single attribute: gray-scale intensity. This fact sometimes make this family of methods a relatively “blunt tool” for image segmentation, as it usually assumes that the object occupies a certain range of intensities while the background occupies a different (non-overlapping) range of intensities. However, the simplicity of the method usually allows the algorithm to make fewer assumptions regarding the content of the image than more sophisticated segmentation algorithms and in principle be more robust to the type of image it operates on.

Nevertheless, automatic and robust binarization is still one of the hardest tasks in image processing [6]. Automatic binarization methods usually make assumptions about the distribution of intensities in the image [9, 1, 11, 3, 4, 17, 14] or require parameter tuning [12]. In real images, the information that can be extracted from a single image’s histogram is often not sufficient for satisfactory binarization. This has motivated the development of binarization methods that rely on information from more than one image, such as the two-dimensional en-
tropy based binarization [1]. These assume that the two dimensional histogram of an image can be divided into two partitions that maximize the amount of information from intensities above and below the thresholds. However, this assumption might falter when the object and the background are of similar intensities (e.g. smooth transition between object and background) or when the variety of intensities in object pixels is large. An objective method for image binarization that makes minimal or no assumptions on the distribution of intensities in the image(s) is thus needed. Such an algorithm can pave the way for further computerized automated analysis or computerized visualization of three dimensional images.

In this paper a new method for automatic binarization of two or more images from different modalities or different acquisition protocols is presented. The goal of the proposed method is not to segment a specific region of interest in the image (which is clearly application dependent). Rather, it seeks to perform a 'blind' separation of object from background by exploiting the mutual properties of the different images. An analogy for this approach is an untrained human reader who needs to delineate an unfamiliar object of interest. Understanding the extent of an unfamiliar object from one modality can at times be a difficult or impossible task for an untrained observer, due to the lack of contextual (prior) information. However, when information from different modalities is given, it can be 'learned' what is an object and what is the background more easily. This is done by looking for consistent intensity behaviour between the different images.

The proposed method can be viewed as an expansion of the Mutual Information binarization method that was originally proposed by Conaire et. al. [5]. It uses the mutual information both above and below the threshold (i.e. both object and background) to determine the best threshold value, while the
Conaire method only accounts for the mutual information above the thresholds. The proposed method can incorporate information from more than two images, assuming that the background in the images have different properties. It does not make any assumptions about the shape or distribution of intensities in the images, rather it only assumes high joint probabilities of object pixel intensities and low joint probabilities of background pixel intensities in the images. However, because of these underlying assumptions, the method is most effective for sets of images that have different background characteristics, such as images originating from different modalities or different acquisition protocols in medical imaging. The effectiveness of the proposed method is visually evaluated against three other binarization methods. The results of the evaluation are demonstrated on both synthetic images and medical images from clinical practice of different types, protocols and modalities without the need for any modification or tuning.

2 The proposed method

Given a pair of, spatially registered, grayscale images of the same object, acquired using different imaging methods (e.g. modalities), we would like to produce a binary image that has a value of 1 where the pixel is considered ”object” and a value of 0 where the pixel is considered ”background”. In medical imaging it is common to acquire images of the same organ, using several modalities, such as nuclear medicine (PET/SPECT), MRI, computerized tomography (CT) and Ultrasound. Hence, the underlying assumption is that the respective intensity properties of the background is expected to be different. This may be due to different noise models, different acquisition artefacts (e.g. partial volume artefacts) or the nature of the different imaging method (e.g. physical properties, imaging tracer, imaging protocol). In this work, the proposed method is demon-
strated on image pairs taken from PET acquisition and contrast enhanced (CE) MRI and on images taken from different types of MRI acquisitions (i.e. T1 and T2 weighted images).

2.1 Underlying assumptions

The proposed algorithm exploits the spatial mutual information that can be acquired from the two different images. It is thus assumed that the intensity pattern in the background of the images is different. This may happen for a number of reasons including differences in: realization of noise, noise models, acquisition related artefacts, physical properties and more. In medical imaging specifically, these differences in background are a common phenomenon because an organ suspected of disease is often imaged using several different modalities. It is also assumed that the intensities of object pixels spatially correspond between the two images. This means that regions that have homogeneous intensity in one image will be homogeneous in the other image and vice versa. This assumption however, does not restrict the object from having different intensity levels or boundaries (i.e. gradients) that look different in the different images.

2.2 Description for the two image case

Let $A$ and $B$ be two different images of the same object, each of size $N$ pixels. The goal of the algorithm is to choose two thresholds, $t_A$ and $t_B$, such that the spatial correspondence of the intensities above both thresholds between the images will be maximized while the spatial correspondence of the intensities below both thresholds between the images will be minimized.

Mutual information is a similarity measure that is derived from information theory [15, 16] and has been widely used for image registration [8, 10]. One of the advantages of this measure is that it does not assume that intensities
of the same object in different images have to be similar, or even to correlate. Rather, the joint probability histogram should have a high level of information in terms of measured entropy [10]. This assumption makes the mutual information measurement attractive in terms of robustness to changes in intensity levels and gradient magnitude between different images. The spatial correspondence of intensities between the images is measured using the mutual information [6], as defined for two discrete random variables:

\[
I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right)
\]  

where \( p(x, y) \) is the joint probability distribution function and \( p(x) \) and \( p(y) \) are the marginal probabilities. In the case of images, we normalise each image histogram of intensities to be a discrete probability function.

Given a pair of images \( A, B \), of identical size and a pair of corresponding thresholds \( t_A, t_B \), we define the thresholded version of image \( A \) to be:

\[
\hat{A}_{t_A, t_B} = \{ i \in A \mid A_i \geq t_A \land B_i \geq t_B \}
\]  

And the residual of the thresholded image \( A \) to be:

\[
\check{A}_{t_A, t_B} = \{ i \in A \mid A_i < t_A \land B_i < t_B \}
\]  

In a similar way we define the thresholded version of \( B \) and the residual of the thresholded version of \( B \):

\[
\hat{B}_{t_A, t_B} = \{ j \in B \mid A_i \geq t_A \land B_i \geq t_B \}
\]  

\[
\check{B}_{t_A, t_B} = \{ j \in B \mid A_i < t_A \land B_i < t_B \}
\]
Given a pair of thresholds, $t_A$ and $t_B$ we define the masked mutual information between $A$ and $B$ to be:

$$\hat{I}_{t_A,t_B} = \hat{I}(A_{t_A}; B_{t_B}) = I(\hat{A}_{t_A}; \hat{B}_{t_B})$$  \hspace{1cm} (6)$$

and the residual mutual information to be

$$\check{I}_{t_A,t_B} = \check{I}(A_{t_A}; B_{t_B}) = I(\check{A}_{t_A}; \check{B}_{t_B})$$  \hspace{1cm} (7)$$

The pair of thresholds that will yield the best separation between the object and the background, in the proposed method, is:

$$(T_A, T_B) = \text{argmax}_{t_A,t_B}(\hat{I}_{t_A,t_B} - \check{I}_{t_A,t_B})$$  \hspace{1cm} (8)$$

Based on this method, the set of object pixels in each of the images $A$ and $B$, respectively, will then be: $\hat{A}_{T_A,T_B}$ and $\check{B}_{T_A,T_B}$.

The method seeks to create a binary mask where the pixels inside the mask have a high level of mutual information, assuming that the intensity of object pixels corresponds between the two images. However, due to the different nature of the images, it is assumed that the background in the different images do not correspond and thus have a low level of mutual information. Given that some types of medical images, the background may contain large amounts of zero (or minimum value) intensity pixels, these may randomly correspond between the images. In order to avoid this situation, we ignore the possible solution $(T_A, T_B) = (\text{min}\{A\}, \text{min}\{B\})$ of equation 8, which is the solution that takes the minimum intensity from both images to be the thresholds.
2.3 Generalization to more than two images

The method can be easily generalized to three images. Using more than two images allows the method to refine the results of the object segmentation by exploiting information from an additional image. The method can be generalized to three images as follows: Given three images $A$, $B$, $C$, and three thresholds $t_A$, $t_B$, $t_C$, we redefine the thresholded version of image $A$ to be:

$$
\hat{A}_{t_A, t_B, t_C} = \{ i \in A \mid A_i \geq t_A \land B_i \geq t_B \land C_i \geq t_C \}
$$

(9)

The residual of the thresholded image $A$ will then be:

$$
\tilde{A}_{t_A, t_B, t_C} = \{ i \in A \mid A_i < t_A \land B_i < t_B \land C_i < t_C \}
$$

(10)

In a similar way, we define the thresholded and residual versions of images $B$ and $C$. The optimal triplets of thresholds for separating the object from the background in the images will be defined as an expansion of equation 8 to include possible correspondences:

$$
(T_A, T_B, T_C) = \arg\max_{t_A, t_B, t_C} \left[ (\hat{I}_{t_A, t_B} + \hat{I}_{t_A, t_C} + \hat{I}_{t_C, t_B}) - (\tilde{I}_{t_A, t_B} + \tilde{I}_{t_A, t_C} + \tilde{I}_{t_C, t_B}) \right]
$$

(11)

Not all possible correspondences have to be taken into account, although, this method ensures that all the information in the system will be exploited. The same approach can be used in order to incorporate information from any number of images. Nevertheless, the computational complexity of the method is exponential to the number of images. Thus, when more than three images are used, running time might quickly become impractical. In order to reduce computational complexity, it is possible to select one image as a “master” image,
while the other images are measured against it in terms of mutual information. The selection of a best master image in this case may not always be a trivial task, as this should be the image that is the “least similar” to other images. The selection of a master image can be done by measuring the mutual information between all possible pairs of images and selecting the image with the lowest mean mutual information to be the master image. Alternatively, instead of selecting a master image, a non-linear optimization algorithm can be used in order to find the best threshold rather than an exhaustive search (e.g. Levenberg-Marquardt or Nelder-Mead Simplex). In both cases, however, an optimal solution is not guaranteed as the objective function of the optimisation problem might not be convex. From our observations, however, the objective function tends to be approximately convex and thus non-linear optimisation seems to be a reasonable approach to take.

3 Experimental results

In order to test the method, two experiments were performed, to demonstrate the performance of the proposed method on synthetic and real images from clinical practice, in comparison to other thresholding methods. Three thresholding methods were chosen for the comparison: Otsu thresholding [9], 2D entropy based binarization [1] and the method of Conaire et al. [5]. The Otsu method was chosen because it is widely used, whereas the Conaire and the 2D entropy methods provide comparable thresholding methods that exploit information from two images. Given that the Otsu method can only handle one image at a time it was applied to each image separately and the final mask of the object was chosen to be the overlap between the two resulting masks.
3.1 Synthetic images

The goal of this experiment was to test the proposed method on synthetic images and compare its performance on these images to the two other methods. Two different images were generated for this experiment. The images consist of a non-uniform rectangular object and a non-uniform background (Figure 1). The results of the three different methods are illustrated in Figure 2. The histograms of the images show that perfect separation between object and background is possible. Nevertheless, all methods but the proposed one failed to find a threshold that yields this perfect separation. It seems that the “spiky” shape of the histogram contributed to this result because most of the algorithms make assumptions about the distribution (Otsu) of intensities or on the amount of information that adjacent intensity levels provide (Entropy). The Conaire method seems not to perform well because of the complicated structure in the foreground and the background weakens the relevancy of mutual information above the threshold. Given that the proposed method looks to also minimise the amount of mutual information below the threshold, it minimises overlap between the background of the different images.

3.2 Medical images

For evaluating the performance of the algorithm on medical images a series of experiments were performed on five sets of images of primary brain or breast cancer, and one set of breast cancer images, from clinical practice. Each set of images was acquired from a different patient. The brain images were acquired using PET and different acquisition protocols of MRI which included CE perfusion and susceptibility weighted imaging (SWI). The proposed method was compared to the Conaire method and to the 2D entropy based binarization on the pairs of images. Quantitative comparison was performed by manually
thresholding each one of the images and then taking the overlap of the masks as ground truth. Each threshold was selected to be the maximum intensity level such that all the object voxel intensities were equal or higher than the selected threshold (i.e. high sensitivity). The result of each method was then compared to the ground truth mask using the DICE coefficient score (defined by \( DICE(A, B) := 2 \cdot |A \cap B| / (|A| + |B|) \) for the sets \( A \) and \( B \)). The results are summarised in Table 1. Note that although the ground truth is a good approximation it is not perfect because various regions in the body do not show high contrast in many medical images as they are not of interest to the clinicians (e.g. air in the lungs, Dura mater and subdural space). This fact makes it impractical to accurately segment the object of interest by hand. The results were also evaluated qualitatively and are presented in Figure 4 (brain) and Figure 5 (breast). Note that the Otsu method failed to segment the imaged object from the background in all cases and provided pairs of threshold values that generated an empty mask. The results suggest that the proposed method outperforms the 2D entropy binarization. A comparison between the new method and the Conaire method shows that Conaire method overestimates the magnitude of the thresholds in a couple of cases (brain patients 1 and 4 in Figure 4) which causes parts of the head, which contain brain tissue, to be masked out. Such low sensitivity is usually highly undesirable in clinical practice. In the breast MR images (Figure 5) the Conaire method underestimates the magnitude of thresholds which causes the air in the lungs and around the body to be considered as part of the object. In one case, however, (Figure 4, bottom row) the Conaire method clearly outperforms the proposed method which underestimates the threshold levels. The underestimation causes the proposed method to produce low-specificity results by including regions of background pixels. This example represents the worst result generated by the proposed method. The
Table 1: DICE coefficient score between the masks of the different methods and the ground truth mask.

<table>
<thead>
<tr>
<th></th>
<th>Proposed</th>
<th>Conaire</th>
<th>2D Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain 1</td>
<td>0.947</td>
<td>0.927</td>
<td>0.082</td>
</tr>
<tr>
<td>Brain 2</td>
<td>0.965</td>
<td>0.916</td>
<td>0.029</td>
</tr>
<tr>
<td>Brain 3</td>
<td>0.977</td>
<td>0.973</td>
<td>0.110</td>
</tr>
<tr>
<td>Brain 4</td>
<td>0.867</td>
<td>0.955</td>
<td>0.074</td>
</tr>
<tr>
<td>Brain 5</td>
<td>0.964</td>
<td>0.975</td>
<td>0.063</td>
</tr>
<tr>
<td>Breast</td>
<td>0.823</td>
<td>0.806</td>
<td>0.544</td>
</tr>
</tbody>
</table>

low specificity of the proposed method, however, can be improved by incorporating additional image(s), as shown in equations 9 and 10. The usefulness of the proposed method on three images was also tested using the SWI images. The DICE coefficient scores were improved and are given in Table 3. The results for patients 1 and 4 are presented in Figure 6 (note the significant improvement for patient 4, both quantitatively and qualitatively). The results show that the binary mask can be further refined by using additional information in the system.

The effect of noise on the method was also tested in comparison to the Conaire method. Ten different levels of Gaussian additive noise were added to each image and the performance of each method (in terms of DICE coefficient score) was measured. Gaussian additive noise was used, with a standard deviation of 0.1 to 1.0 (in steps of 0.1) of the mean signal in the image. The noise level was calculated for each image independently. The mean results for each noise level are presented in Table 2. The results suggest that the proposed method is less sensitive to high levels of noise than the Conaire method.

4 Discussion and Conclusions

A new method for automatic thresholding based on two or more images has been proposed. The method finds a set of thresholds for differentiating between
Table 2: Performance comparison between the proposed and Conaire method as a function of noise level. The noise column indicates the ratio of noise to mean signal in each image. The DICE coefficient scores for the proposed and the Conaire methods are given for each level of noise.

<table>
<thead>
<tr>
<th>Noise</th>
<th>Proposed</th>
<th>Conaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.819</td>
<td>0.811</td>
</tr>
<tr>
<td>0.2</td>
<td>0.798</td>
<td>0.760</td>
</tr>
<tr>
<td>0.3</td>
<td>0.770</td>
<td>0.754</td>
</tr>
<tr>
<td>0.4</td>
<td>0.754</td>
<td>0.707</td>
</tr>
<tr>
<td>0.5</td>
<td>0.725</td>
<td>0.677</td>
</tr>
<tr>
<td>0.6</td>
<td>0.723</td>
<td>0.682</td>
</tr>
<tr>
<td>0.7</td>
<td>0.732</td>
<td>0.664</td>
</tr>
<tr>
<td>0.8</td>
<td>0.694</td>
<td>0.635</td>
</tr>
<tr>
<td>0.9</td>
<td>0.664</td>
<td>0.644</td>
</tr>
<tr>
<td>1.0</td>
<td>0.677</td>
<td>0.645</td>
</tr>
</tbody>
</table>

Table 3: DICE score between the ground truth mask and the proposed method with three images (CE-MRI, PET and SWI).

<table>
<thead>
<tr>
<th>Patient</th>
<th>DICE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain 1</td>
<td>0.956</td>
</tr>
<tr>
<td>Brain 2</td>
<td>0.972</td>
</tr>
<tr>
<td>Brain 3</td>
<td>0.980</td>
</tr>
<tr>
<td>Brain 4</td>
<td>0.958</td>
</tr>
<tr>
<td>Brain 5</td>
<td>0.971</td>
</tr>
</tbody>
</table>
object and background from a set of two or more images of the same object, acquired from different modalities or protocols. The underlying assumption of the method is that different images of the object have high mutual information while the background characteristics differ between the modalities. This assumption also introduces an implicit limitation of the proposed method, because at least two different images of the object must be acquired, using different sources of information (e.g. different modalities, sensors or acquisition protocols). The proposed method was tested on both synthetic and medical images from clinical practice and compared against three other thresholding methods: the Conaire method, the popular Otsu thresholding method and 2D entropy based binarization. The results of the experiments on the synthetic images suggests that none but the proposed method managed to find a pair of thresholds that will perfectly differentiate between the object in the image and the background. A possible explanation for this result may be the irregular histogram shape and spatial patterns in the image. This result suggests that the proposed method is less sensitive to such irregularities as it does not make assumptions about the distribution of intensities in the images. The result of the experiment with clinical images suggest that the 2D entropy binarization tends to pick thresholds that are too high and it is usually biased by high intensity tissues in the image. The Conaire method seemed to pick thresholds that are too high for two of the tested brain datasets and too low thresholds for the breast dataset. The proposed method did select thresholds that are too low in one of the brain datasets, resulting in over-segmentation. However, by incorporating more images in the new method, its specificity can be further improved. Moreover, the underlying assumptions in the proposed method can be generalized to families of images from the same modality or acquisition protocol and thus provide consistent results that are not dependent on the distribution of intensities in the image.
Thus, in the presence of two or more images, providing different information about same object, the proposed method can provide an objective, parameter free, thresholding approach.

5 Acknowledgment

This research was supported under the Australian National Health and Medical Research Council (NHMRC) funding scheme (project grant 631567). Also, we would like to thank Dr Michael Fay from Radiation Oncology, The Royal Brisbane and Women’s Hospital for his help with patient recruitment, to Dr Paul Thomas, Queensland PET Service for assisting with PET imaging, to Dr Nicholas Dowson from CSIRO for assisting with PET data preparation and to Ms Kimberly Nunes from the Centre for Medical Diagnostic Technologies for performing proof reading.

References


Figure 1: The two synthetic images used with their corresponding histograms.
Figure 2: The results of the three binarization methods on the pair of synthetic images: (a) the proposed method; (b) Conaire MI thresholding; (c) 2D entropy; (d) Otsu
Figure 3: Sample slices from the clinical brain images used for evaluating the method: (a) CE MRI; (b) PET images
Figure 4: The results of the proposed method and the 2D entropy based binarization on a pair of brain images acquired by PET and MRI (the resulting mask is in green): (a) ground truth; (b) the resulting masks of the proposed method; (c) the resulting masks of the Conaire method; (d) the binary mask from 2D entropy binarization
Figure 5: The results of the proposed method and 2D entropy based binarization on a pair of breast MR T1 and T2 weighted images (the resulting mask is in green): (a) the original T1 image; (b) the original T2 image; (c) ground truth; (d) binary mask resulting from the proposed method; (e) binary mask resulting from the Conaire method; (f) the binary mask resulting from the 2D entropy binarization.
Figure 6: The results of the proposed method when applied to two images and three different images (CE-MRI, PET and SWI-MRI), from brain patients 1 (top) and 4 (bottom): (a) the original SWI image; (b) ground truth mask; (c) original result of the proposed method (i.e. CE-MRI and PET only), and; (b) Results of the proposed method when the SWI image is added (i.e. on 3 images)