Contributed paper

OPTO-ELECTRONICS REVIEW **11**(3), 253–259 (2003)

Information theory based medical images processing

K. KUCZYŃSKI* and P. MIKOŁAJCZAK

Laboratory of Information Technology, University of Maria Curie-Skłodowska,
5 Curie-Skłodowskiej Sq., 20-031 Lublin, Poland

Increasing application of non-invasive medical techniques (like stereotactic radiosurgery) generates a high demand for modern image processing algorithms. Image registration and segmentation are the two essential examples of this. The algorithms need to be reasonably fast, reliable, accurate, and highly automated. Information theory provides a means to create such systems. In this paper we present thresholding segmentation using image entropy and a registration technique based on maximization of mutual information. Then we show some experimental results using real-world computed tomography (CT) and medical resonance imaging (MRI) data.

Keywords: image segmentation, image registration, entropy, mutual information.

1. Introduction

There is a rich variety of available medical imaging techniques (CT, MRI, PET, SPECT, USG, …). Their quality is improving continuously. Growing popularity of non-invasive therapeutic techniques, like stereotactic radiosurgery, based mainly on image diagnostic data causes a high demand for image processing algorithms. These should be highly automated, fast and reliable in order to make a profound use of the acquired data.

Despite their excellence, various imaging techniques provide different and often complementary kinds of information about physical properties of tissues, so all of them are required. In case of brain imaging it is common to acquire both CT and MRI scans. CT provides precise anatomical information, especially about bones. Voxel intensities are proportional to the radiation absorption of the underlying tissues, which is particularly valuable in radiation therapy planning. MRI is less accurate, but soft tissue (including tumour) delineation is significantly enhanced. However, interpretation of the visualized parameters (spin-lattice relaxation time T_1, spin-spin relaxation time T_2, spin density ρ; Table 1, Ref. 1) is not as straightforward as in CT.

The task of integrating two datasets is not trivial due to non-similar patient's spatial orientation, different resolutions and voxel intensity profiles. The goal of a registration (matching, alignment) process is, given two images of the same object, to find a spatial transformation T that relates them. The next step is a fusion of the registered datasets.

Another crucial technique is image segmentation, which is a process of partitioning an image into regions homogenous with respect to a given criterion. It is often the first step in image processing, visualization and analysis. The review and classification of segmentation methods can be found in Ref. 2.

Table 1. Range of T_1, T_2, and ρ values at 1.5 T magnetic field for tissues found in a magnetic resonance image of the human head (after Ref. 1).

Tissue	T_1 [s]	T_2 [ms]	ρ^*
CSF	0.8–20	110–2000	70–230
White	0.76–1.08	61–100	70–90
Grey	1.09–2.15	61–109	85–125
Meninges	0.5–2.2	50–165	5–44
Muscle	0.95–1.82	20–67	45–90
Adipose	0.2–0.75	53–94	50–100

*Based on ρ =111 for 12 mM aqueous $NiCl_2$.

The tasks of registration and segmentation are closely related. Segmented images are easier to register. On the other hand, having two registered images, better segmentation can be obtained. Here, we present an approach to registration and segmentation based on elements of information theory.

2. General theory

Application of information theory methods in image processing is possible, assuming that we can treat images as random variables. In this chapter, a theoretical background of the implemented registration and segmentation procedures is presented [3,4].

*e-mail: karol.kuczynski@umcs.lublin.pl

The most frequently used measure of information is the Shannon-Wiener entropy measure. The entropy H of a discrete random variable X with the values in the set $\{x_1,x_2,\ldots..x_n\}$ is defined as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i, \tag{1}$$

where $p_i = Pr[X = x_i]$.

The entropy definition of a single random variable can be extended to a pair of random variables. The joint entropy $H(X,Y)$ of a pair of discrete random variables with a joint distribution p_{ij} is

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_{ij}. \tag{2}$$

The conditional entropy $H(Y|X)$ is defined as

$$\begin{aligned} H(Y|X) &= \sum_{i=1}^{n} p_i H(Y|X = x_i) \\ &= -\sum_{i=1}^{n} p_i \sum_{j=1}^{m} p_{j|i} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_{j|i} \end{aligned} \tag{3}$$

where $p_{j|i} = Pr[Y = y_j|X = x_i]$.

Mutual information between two discrete random variables X and Y is defined as

$$I(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i p_j}. \tag{4}$$

It can be shown, that

$$\begin{aligned} I(X,Y) &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) \\ &= I(Y,X). \end{aligned} \tag{5}$$

The mutual information represents the amount of information that one random variable gives about the other random variable, or in other words a measure of the reduction in the entropy of one variable, given the other variable. Normalised mutual information [5] is defined as

$$NI(X,Y) = \frac{H(X) + H(Y)}{H(X,Y)}. \tag{6}$$

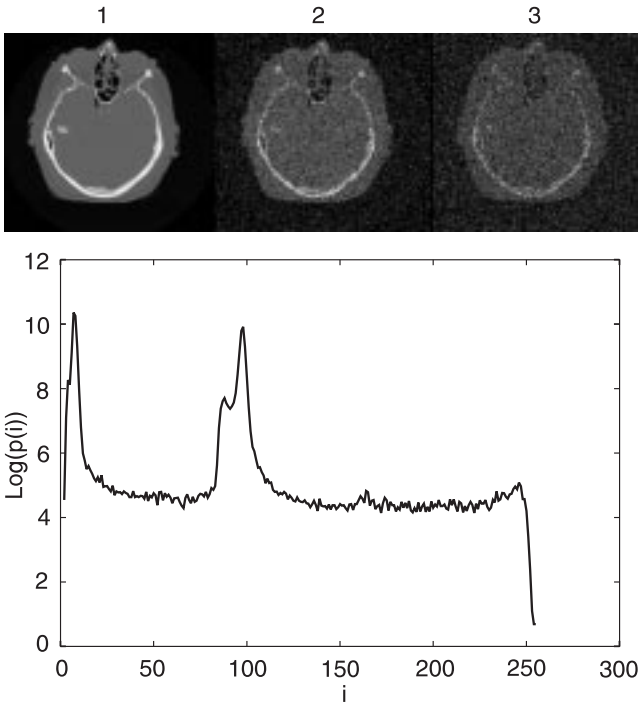
The image entropy, Eq. (1), is usually estimated using a histogram [6]

$$p_i \equiv \frac{g_i}{g_{total}}, \tag{7}$$

where g_i is the number of pixels with the intensity i and g_{total} is the total number of pixels while n in Eq. (1), is the number of grey-levels. However, this approach has a significant drawback: the pixels are assumed to be independent and the spatial information is ignored. Random rearrangement does not affect the entropy, which is counter-intuitive.

Table 2 shows three images. Their histograms are identical so, the histogram entropies H_1, H_2, and H_3 are equal, too.

Table 2. Image entropy calculation – examples.



$H_1 = 3.17$	$H_2 = 3.17$	$H_3 = 3.17$
$H_{s1} = 13.6$	$H_{s2} = 17.6$	$H_{s3} = 19.4$

Another approach, often referred as a "monkey model", assumes an image to be made up of a fixed number of photons (unit grey-levels) G randomly allocated in n cells (pixels, Fig. 1) [7]. Here the entropy calculation [Eq. (1)] is used over again, but

$$p_i = \frac{g_i}{G} \tag{8}$$

where g_i is the grey-level of the pixel i.

In this case, the spatial information is still ignored, like in the previous model. However, if we use a modified form of Eq. (1) [8]

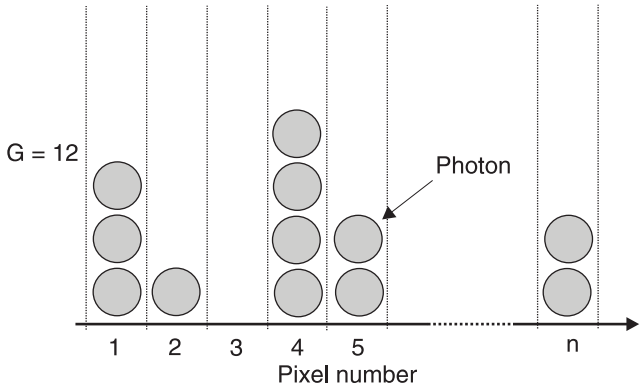


Fig. 1. "Monkey model" of a 1-dimensional image (after Ref. 7).

$$H_s = -\sum_{i=1}^{m} p_i \log \frac{p_i}{m_i}, \tag{9}$$

where m is the model based measure defined over the same domain as p, it can be used to incorporate dependence between pixels, for example [7]

$$m_i = 1 + \sigma_i^2, \ \sigma_i^2 = \sum_{i \in N_3} \frac{(g_i - \mu_{N_3})^2}{9}. \tag{10}$$

σ_i^2 is a grey-level variance over the 3×3 neighbourhood N_3 of the pixel i and μ_{N_3} *is the mean grey-level in* N_3. *It is applied in order to emphasize the image's features, relevant for the understanding of the image by a human (contours, homogenous areas, etc.). The more variable image, the greater the variance so, it appears to be one sensible choices*.

The last row of Table 2 presents entropy calculation results using the above model. The more random the image, the higher the entropy ($H_{s1} < H_{s2} < H_{s3}$).

3. Mutual information based image registration

Registration methods can be classified with respect to many criteria [9]. The most general classification distinguishes feature-based and voxel-based techniques. In the first case, some corresponding points (either external artificial markers or anatomical structures localised by an expert) need to be recognized prior to the registration, then the spatial distances between them are to be minimized. In voxel-based techniques a similarity measure between two images is computed with the intensities of all (or most of) voxels in the images. Neither segmentation nor special pre-processing is required. The last approach is becoming more and more popular, however it requires more processing time.

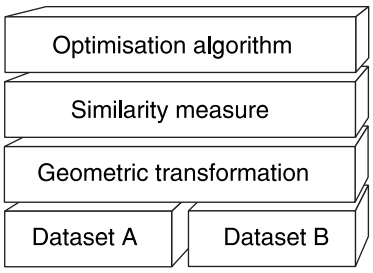


Fig. 2. Optimisation framework.

Regardless of the method used, the registration framework is always similar (Fig. 2, Ref. 10). In order to register two images, a geometrical transformation needs to be implemented. There is a wide class of local and global transformations (rigid, affine, projective, curved) that can be used [9]. An affine 3D transformation can be easily described using a single constant 4×4 matrix. The rigid body transformation implemented in our program is defined as

$$T = \begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(R_x) & -\sin(R_x) & 0 \\ 0 & \sin(R_x) & \cos(R_x) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(R_y) & 0 & \sin(R_y) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(R_y) & 0 & \cos(R_y) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(R_z) & -\sin(R_z) & 0 & 0 \\ \sin(R_z) & \cos(R_z) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{11}$$

where T_x, T_y, T_z, R_x, R_y, and R_z are the transformation parameters.

The joint entropy could be used as a similarity measure between two images. It is convenient to employ a 2-dimensional greyscale histogram (scatter-plot) for this purpose. Figure 3 shows two pairs of images and their histograms.

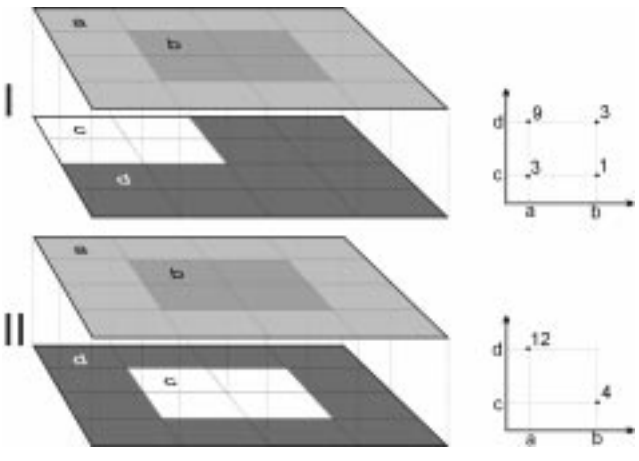


Fig. 3. Two pairs of images and their 2-dimensional histograms.

The joint entropy for these images is calculated as

$$H_I(M,N) = \\ = -\left(\frac{3}{16}\log\frac{3}{16} + \frac{1}{16}\log\frac{1}{16} + \frac{9}{16}\log\frac{9}{16} + \frac{3}{16}\log\frac{3}{16}\right) \approx 0.488, \quad (12)$$

$$H_{II}(M,N) = -\left(\frac{12}{16}\log\frac{12}{16} + \frac{4}{16}\log\frac{4}{16}\right) \approx 0.244. \quad (13)$$

The more similar two images are, the lower is their joint entropy. However, its optimisation may lead to incorrect solutions when the images do not overlap entirely during a registration process. Mutual information is a significantly better candidate for a registration criterion. It can be proved that two images are properly matched when their mutual information is maximal [11].

In order to find the optimal transformation parameters, an optimisation procedure is used to search for a global optimum (minimum or maximum, depending on convention) of the registration criterion (similarity measure). A review of the most often used methods can be found in Ref. 9. Non-deterministic algorithms (e.g., simulated annealing) are successful in many real-world tasks, including registration. However, they typically require much computing time. Deterministic ones (Powell, Davidon-Fletcher-Powell, Levenberg-Marquardt, etc., Ref. 12]) are faster, but they fail in the presence of a large number of local extrema. To address this problem and to speed up the convergence, multi-scale or sub-sampling techniques are utilised [13]. Also, a few different start points may be randomly selected.

4. Entropic thresholding

Thresholding pixel (or voxel in 3-dimensional images) intensity values method is the simplest and the most commonly applied segmentation one. It can be performed either globally (for example using the intensity-scale histogram of the image) or locally (by selecting thresholds separately for each sub-image). The partition is defined as [6]

$$j(r,c) = \begin{cases} b_0, & \text{if } i(r,c) \le t, \\ b_1, & \text{if } i(r,c) > t. \end{cases} \quad (14)$$

$i(r,c)$ is the grey-scale image, $j(r,c)$ is the resulting binary image, and t is the threshold to be found. In this way we obtain two classes of pixels: "black" b_0 and "white" b_1.

According to Ref. 14 we assume, that "the entropy of each region is always lower than entropy of the whole image or, in other words, the entropy of a region is always greater than the entropy of its subdomains". For the two classes of pixels (b_0 and b_1) in an 8-bit greyscale source image we can calculate entropies as [6]

$$H_{b_0}(t) = -\sum_{i=0}^{t} \frac{p_i}{p(b_0)} \log \frac{p_i}{p(b_0)}, \quad (15)$$

$$H_{b_1}(t) = -\sum_{i=t+1}^{255} \frac{p_i}{p(b_1)} \log \frac{p_i}{p(b_1)}, \quad (16)$$

where

$$p(b_0) = \sum_{i=0}^{t} p_i, \\ p(b_1) = \sum_{i=t+1}^{255} p_i. \quad (17)$$

The entropies H_{b_0} *and* H_{b_1} *may be analogically calculated using Eq. (9) by including spatial information. The segmentation criterion may be calculated in numerous ways. The two examples are [15,16]*

$$t = \arg\max(H_{b_0}(t) + H_{b_1}(t)), \quad (18)$$

or

$$t = \arg\min(H_{b_0}(t) + H_{b_1}(t))^2. \quad (19)$$

Multi-class segmentation is performed by iterative segmentation of classes resulting from previous steps.

5. Experimental results

Figure 4 shows the images (CT and MRI of the same patient) to be registered. Figures 5 and 6 present the registration result. The images have been registered by maximization of mutual information using the Powell's optimisation algorithm [12] with sub-sampling and multiple start points. Despite the simplicity of the Powell's method, in the case of typical datasets it led to acceptable results. The transformation parameters are: T_x = 65.0 mm, T_y = 59.5 mm, T_z = 5.9 mm, R_x = 17.8°, R_y = 1.4°, and R_z = 7.4°.

Figures 7 and 8 show 2-dimensional histograms of the images, before and after the registration process. The problem of local extrema is visualised in Fig. 9 (mutual information as a function of two transformation parameters, while the others have been set in the optimum). The global maximum corresponds to the best possible registration.

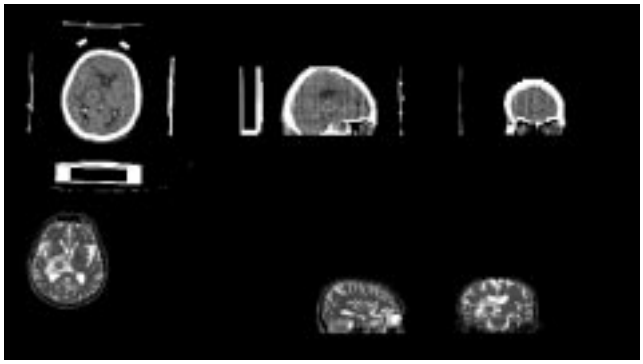


Fig. 4. Source datasets.

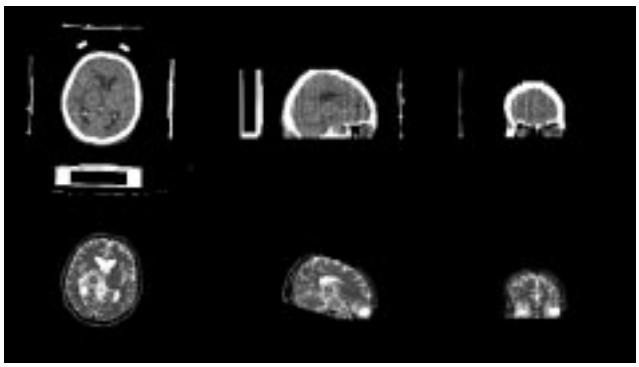


Fig. 5. Optimisation result.

Figure 10 shows a result of a CT image entropy segmentation using spatial information. The optimal value of the threshold corresponds to the maximum of the criterion [Eq. (18), Fig. 11]. One of the resulting classes has been segmented over again in order to obtain three classes (background, soft tissues and bones).

6. Conclusions

Novel, sophisticated medical routines require new advanced image processing techniques. Information theory provides means to create highly automated systems. In most cases, there is no need of pre-processing or expert assistance. Computational complexity and occurrence of local extrema are still important concerns, however on a modern PC it is possible to achieve satisfactory results within a reasonable time.

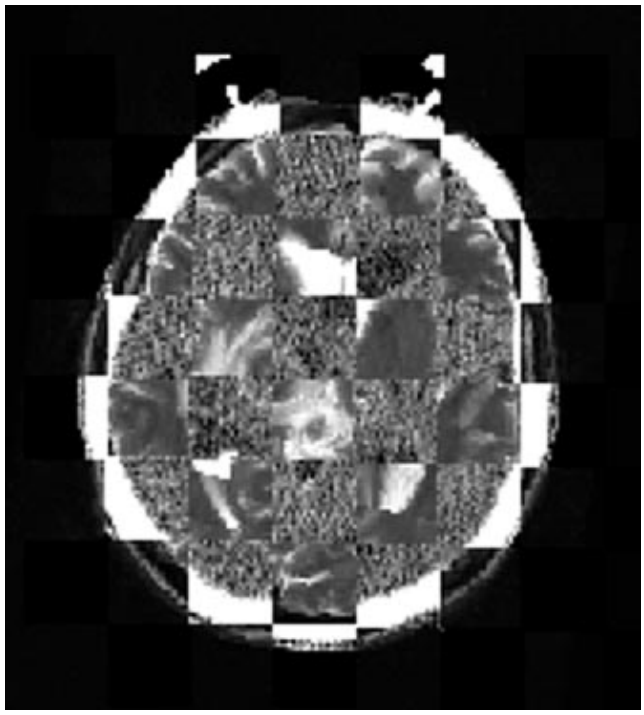


Fig. 6. Optimisation result – a checkerboard test.

Acknowledgements

This study has been partially financed by the Polish State Committee for Scientific Research (research project no. 7T11E01921).

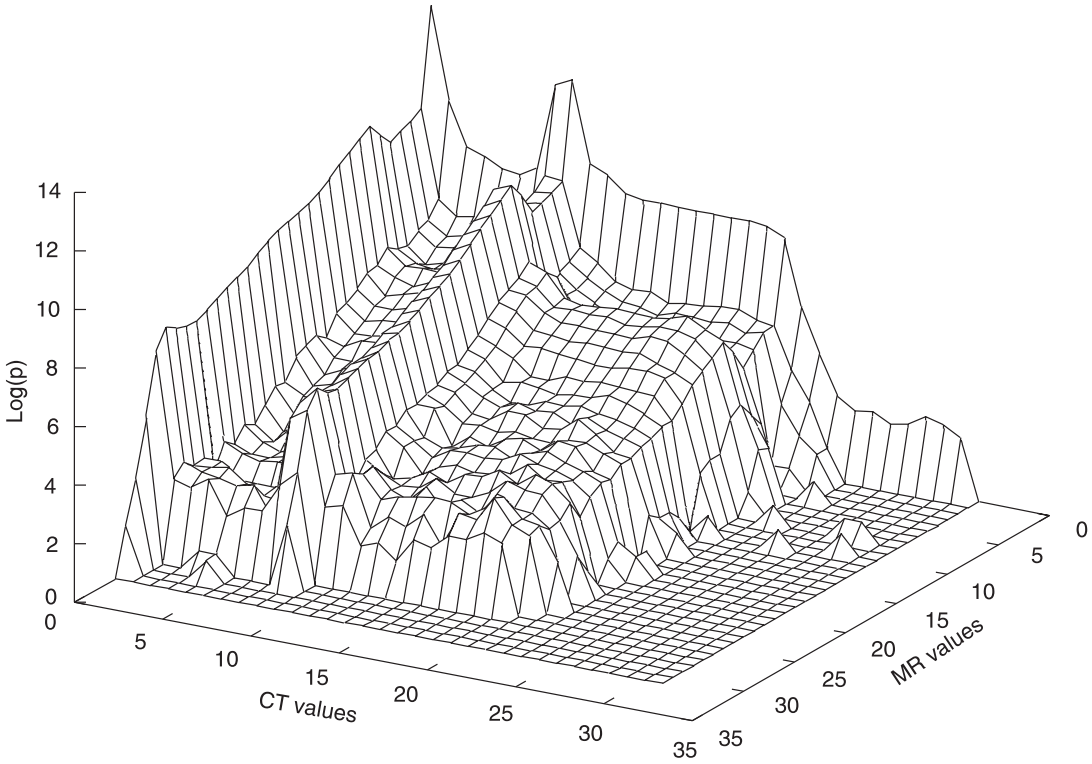


Fig. 7. Histogram of the source images.

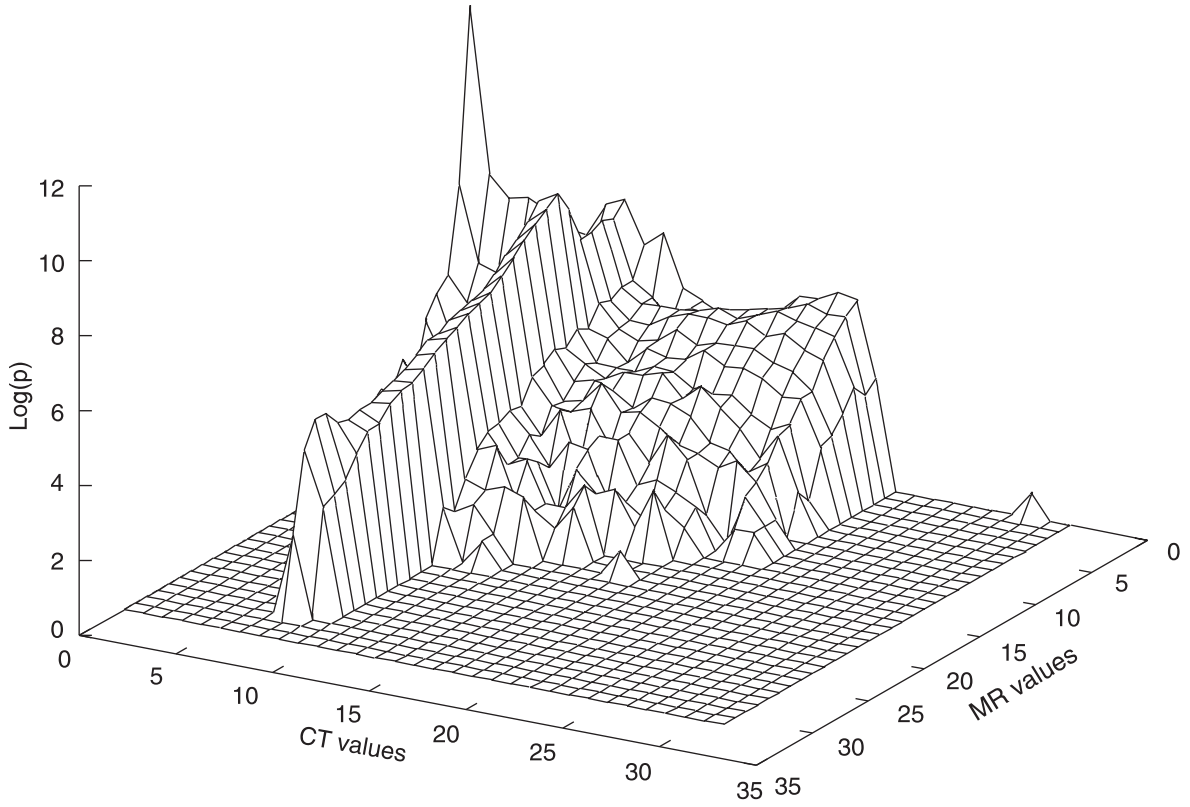


Fig. 8. Histogram of the registered images.

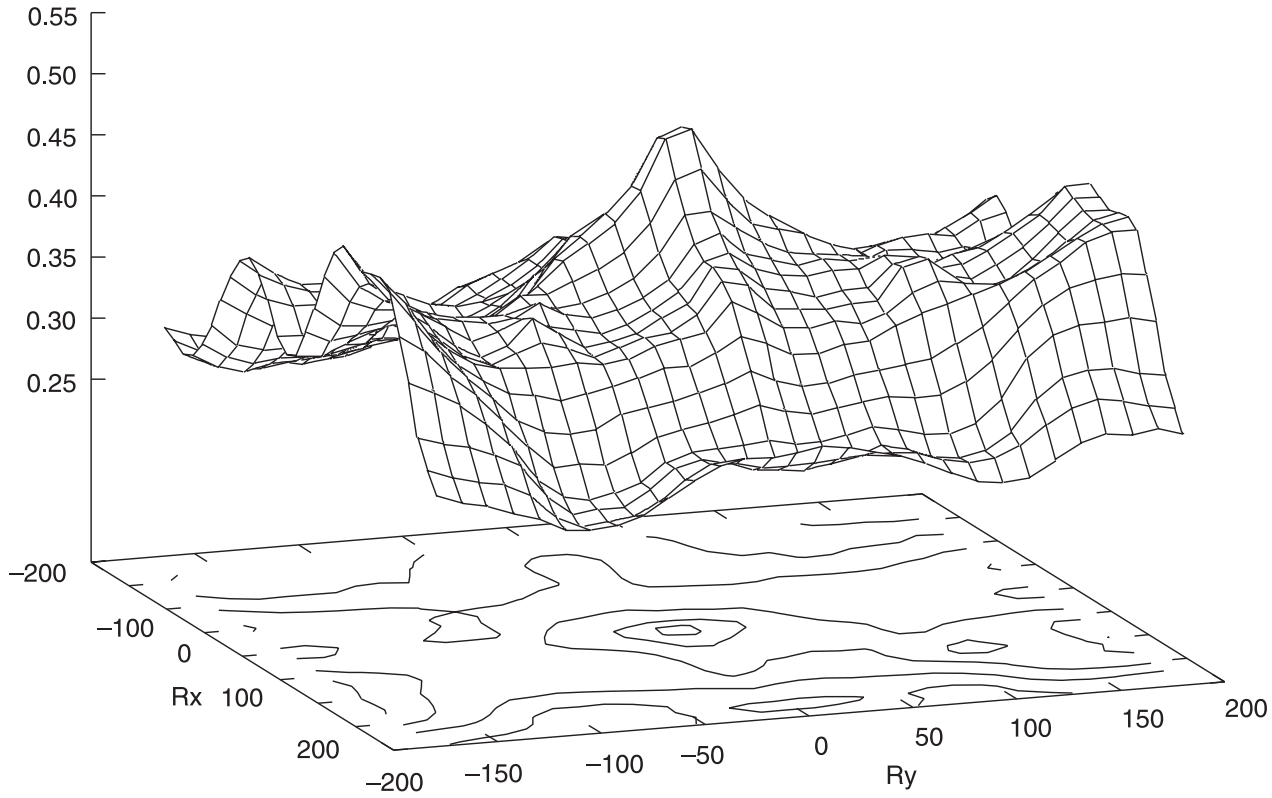


Fig. 9. Mutual information as a function of two of the transformation parameters.

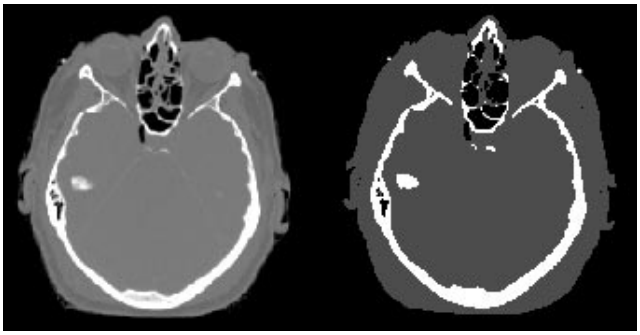


Fig. 10. CT image segmentation.

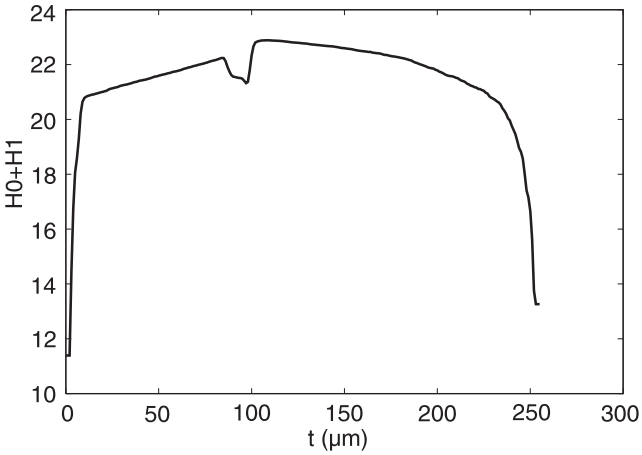


Fig. 11. Optimisation criterion [Eq. (18)] as a function of threshold.

References

1. J.P. Hornak, *The Basics of MRI*, http://www.cis.rit.edu/htbooks/mri, 2000.
2. J. Rogowska, "Overview and fundamentals of medical image segmentation", in *Handbook of Medical Imaging*, *Processing and Analysis*, Academic Press, London, 2000.
3. T. Cover and J. Thomas, *Elements of Information Theory*, Wiley–Interscience Publication, New York, 1991.
4. C.E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal* **27**, 379–423, 623–656 (1948).
5. C. Studholme, D.L.G. Hill, and D.J. Hawkes, "An overlap invariant measure of 3D image alignment", *Pattern Recognition* **32**, No 1 (1998).
6. E.D. Jansing, T.A. Albert, and D.L. Chenoweth, "Two-dimensional entropic segmentation", *Pattern Recognition Letters* **20**, 329–336 (1999).
7. A.D. Brink, "Using spatial information as an aid to maximum entropy image threshold selection", *Pattern Recognition Letters* **17**, 29–36 (1996).
8. J. Skilling, "Theory of maximum entropy image reconstruction", in *Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by J.H. Justice, Cambridge University Press, Cambridge, 1986.
9. J.B. Maintz and M.A. Viergever, "A survey of medical image registration", *Medical Image Analysis* **2**, 1–36 (1998).
10. K. Kuczyński and P. Mikołajczak, "Mutual information based registration of brain images", *Journal of Medical Informatics &* Technologies **3**, 213–219 (2002).
11. P.A. Viola, *Alignment by Maximisation of Mutual Information*, A.I. Technical Report No. 1548, MIT 1995.
12. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C++*, Cambridge University Press, Cambridge, 2002.
13. D.L.G. Hill, *Combination of 3D Medical Images from Multiple Modalities*, University of London, London, 1993.
14. S. Vitulano, C. Di Ruberto, and M. Nappi, "Different methods to segment biomedical images", *Pattern Recognition Letters* **18**, 1125–1131 (1997).
15. J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, "A new method for grey level picture thresholding using entropy of the histogram", *Computer Vision, Graphics and Image Processing* **29**, 273–285 (1985).
16. P.K. Sahoo, D.W. Slaaf, and T.A. Albert, "Threshold selection using a minimal entropy difference", *Optical Engineering* **36**, 1976–1981 (1997).