Compression-based Image Registration

Anton Bardera, Miquel Feixas, Imma Boada, Mateu Sbert

IiiA, Universitat de Girona

Campus Montilivi, P4, 17004 Girona, Spain

Email: {anton.bardera, miquel.feixas, imma.boada, mateu.sbert}@udg.es

Abstract— **Image registration is an important component of image analysis used to align two or more images. In this paper, we present a new framework for image registration based on compression. The basic idea underlying our approach is the conjecture that two images are correctly registered when we can maximally compress one image given the information in the other. The contribution of this paper is twofold. First, we show that the image registration process can be dealt with from the perspective of a compression problem. Second, we demonstrate that the similarity metric, introduced by Li et al., performs well in image registration. Two different versions of the similarity metric have been used: the Kolmogorov version, computed using standard real-world compressors, and the Shannon version, calculated from an estimation of the entropy rate of the images.**

I. INTRODUCTION

The registration of two images consists in finding the transformation that brings one image into the best possible spatial correspondence with the other one. A common method of solving the registration task is to treat it as a mathematical optimization problem, using a similarity measure to quantify the quality of the alignment of the two images for any given transformation.

Some information-theoretic measures, such as *mutual information* (MI) [1], [2] and *normalized mutual information* (NMI) [3], have become a standard reference, mainly in medical imaging, due to their accuracy and robustness. On the other hand, the normalization of *information distance* [4], based on the non-computable notion of Kolmogorov complexity, has been introduced for measuring similarity between sequences [5], [6]. It has been successfully applied in areas such as genome phylogeny [5], language phylogeny [6] and classification of music pieces [7]. In essence, the idea behind it is that two objects are similar if we can significantly compress one given the information in the other.

In this paper, a normalized version of the information distance, called *the similarity metric* [6], is applied to image registration using two different versions. In the first case, the Kolmogorov complexity is computed using standard real-world compressors and, in the second case, an estimation of the entropy rate of the images is used. In both cases, experimental results demonstrate that *the similarity metric* performs well in image registration. However, the entropy rate approach is more accurate and robust than the Kolmogorov version due to the compressor imperfections.

This paper is organized as follows. In Section II we survey background and related work, and in Section III we present our framework for image registration. Experimental results are given in Section IV.

II. BACKGROUND

In this section we review *the similarity metric* based on the Kolmogorov complexity [6], some basic information-theoretic measures [8], [9] and their application to image registration [1], [2], [10].

A. The Similarity Metric

The *Kolmogorov complexity* $K(x)$ of a string x is the length of the shortest program to compute x on an appropriate universal computer. Essentially, the Kolmogorov complexity of a string is the length of the ultimate compressed version of the string. The conditional complexity $K(x|y)$ of x relative to y is defined as the length of the shortest program to compute x given y as an auxiliary input to the computation. The joint complexity $K(x, y)$ represents the length of the shortest program for the pair (x, y) [6].

In [4], the *information distance* is defined as the length of the shortest program that computes x from y and y from x. It was shown there that, up to an additive logarithmic term, the information distance is given by

$$E(x, y) = \max\{K(y|x), K(x|y)\}. \tag{1}$$

It was also shown that $E(x, y)$ is a metric. It is interesting to note that long strings that differ by a tiny part are intuitively closer than short strings that differ by the same amount. Hence, there arises the necessity to normalize the information distance. In [6], a normalized version of $E(x, y)$, called the *normalized information distance* (NID) or *the similarity metric*, is defined by

$$
\begin{aligned}
NID(x, y) &= \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}} \\
&= \frac{K(x, y) - \min\{K(x), K(y)\}}{\max\{K(x), K(y)\}}. \tag{2}
\end{aligned}
$$

Li et al. have shown that $NID(x, y)$ is a metric and takes values in $[0, 1]$. It is also universal in the sense that if two strings are similar according to the particular feature described by a particular normalized admissible distance (not necessarily metric), then they are also similar in the sense of the normalized information metric [11].

Due to the non-computability of Kolmogorov complexity, a feasible version of the normalized information distance (2),

called the *normalized compression distance*, is defined as

$$NCD(x,y) = \frac{C(x,y) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}. \quad (3)$$

NCD is computed from the lengths of compressed data files. Thus, $C(x)$ (or $C(y)$) represents the length of compressed string x (or y) and $C(x,y)$ the length of the compressed pair (x,y). Thus, NCD approximates NID by using a standard real-world compressor.

B. Information-Theoretic Measures

Let $\mathcal{X}$ be a finite set, let X be a random variable taking values x in $\mathcal{X}$ with distribution $p(x) = Pr[X = x]$. Likewise, let Y be a random variable taking values y in $\mathcal{Y}$. The *Shannon entropy* $H(X)$ of a random variable X is defined by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \quad (4)$$

The Shannon entropy $H(X)$ measures the average uncertainty of random variable X. If the logarithms are taken in base 2, entropy is expressed in bits. The *conditional entropy* is defined by

$$H(X|Y) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log p(x|y), \quad (5)$$

where $p(x,y) = Pr[X = x, Y = y]$ is the joint probability and $p(x|y) = Pr[X = x|Y = y]$ is the conditional probability. The conditional entropy $H(X|Y)$ measures the average uncertainty associated with X if we know the outcome of Y. The *mutual information* between X and Y is defined by

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (6)$$

It is a measure of the shared information between X and Y.

We review now the definition of entropy rate. The notation used here is inspired by the work of Feldman and Crutchfield [9]. Given a chain $\ldots X_{-2}X_{-1}X_0X_1X_2\ldots$ of random variables X_i taking values in $\mathcal{X}$, a block of L consecutive random variables is denoted by $X^L = X_1\ldots X_L$. The probability that the particular L-block x^L occurs is denoted by $p(x^L)$. The Shannon entropy of length-L sequences or *L-block entropy* is defined by

$$H(L) = -\sum_{x^L \in \mathcal{X}^L} p(x^L) \log p(x^L), \quad (7)$$

where the sum runs over all possible L-blocks. The *entropy rate* is defined by

$$h^x = \lim_{L \to \infty} \frac{H(L)}{L} \quad (8)$$

and measures the average amount of information per symbol x [8]. It can also be rewritten as

$$h^x = \lim_{L \to \infty} h^x(L), \quad (9)$$

where $h^x(L) = H(X_L|X_{L-1}X_{L-1}\ldots X_1)$ is the entropy of a symbol conditioned on a block of $L-1$ adjacent symbols. The entropy rate of a sequence is a measure of its compressibility [12].
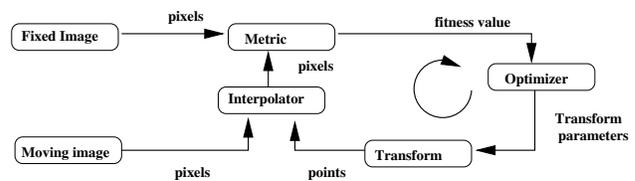

Fig. 1. Main components of the registration process.

C. MI-based Image Registration

Image registration is treated as an iterative optimization problem with the goal of finding the spatial mapping that will bring two images into alignment resulting in a *fused image*. This process is composed of four elements (see Fig. 1). As input, we have both fixed X and moving Y images. The *transform* represents the spatial mapping of points from the fixed image space to points in the moving image space. The *interpolator* is used to evaluate moving image intensity at non-grid positions. The *metric* provides a measure of how well the fixed image is matched by the transformed moving one. This measure forms the quantitative criterion to be optimized by the *optimizer* over the search space defined by the parameters of the transform.

The crucial point of image registration is the choice of a metric. The registration between two images X and Y can be represented by a channel $X \to Y$, where its marginal and joint probability distributions are obtained by simple normalization of the corresponding intensity histograms of the overlap area of both images [1]. The most successful automatic image registration methods are based on the maximization of MI. This method, almost simultaneously introduced by Maes et al. [1] and Viola et al. [2], is based on the conjecture that the correct registration corresponds to the maximum MI between the overlap areas of the two images. Later, Studholme et al. [3] proposed a normalization of mutual information defined by

$$NMI(X,Y) = \frac{I(X,Y)}{H(X,Y)} = 1 - \frac{H(X|Y) + H(Y|X)}{H(X,Y)}, \quad (10)$$

which is more robust than MI, due to its greater independence of the overlap area. Another justification about its good behavior is the following. In information theory, the most basic information metric is given by $H(X|Y) + H(Y|X)$. This measure fulfills the requirements for a true distance. The normalization of this distance by the joint entropy is also a true distance [6], [13]. In image registration literature, this fact has not been sufficiently stressed. It has to be noted that $1 - NMI$ is very close to the Shannon version of NID (see [6], [13]).

III. REGISTRATION BASED ON THE SIMILARITY METRIC

Three different approaches for image registration based on NID are proposed. The two first ones are based on the Kolmogorov complexity (Sect. III-A) and the third one on the Shannon information theory (Section III-B).

437

Fig. 2. Lena image and its R, G and B channels. The similarity metric is evaluated within the dashed window.

A. Kolmogorov Version

In this section, *the similarity metric* is implemented using both image and text compressors.

1) Image compressor-based registration: A color image is represented as a finite number of color image planes where each color is obtained by filtering the image spectrum and by measuring the resulting luminosity energy. Usually, digital color images use three filters, corresponding to the red (R), green (G) and blue (B) channels. In Fig. 2, we show a color image with its RGB channels.

Our proposal is to use an image compressor to register monochrome images. To obtain the fused image, the method places each one of the monochrome images in a different color channel of an RGB image. Then, for each arrangement of images, the fused image and the original ones are compressed. In our experiments, we used the JPEG 2000 compressor, that is based on wavelet technology. $C(x,y)$, $C(x)$, and $C(y)$ are given by the lengths of the JPEG 2000 files corresponding to the compressed fused image and the original images X and Y, respectively. The correct registration would be achieved when NCD is minimum.

Since the application of an image compressor requires rectangular images and, in general, the overlap area is not rectangular, different strategies can be considered to evaluate the similarity. In our experiments, we reduce the area of the image to be processed to the maximal rectangular area contained in the overlap region.

2) Text compressor-based registration: Our proposal is now to convert the registration process into a text compression problem. To do this, the overlap area of the two images to be registered is scanned (see Fig. 3). From the scan path, strings $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are obtained by taking in consecutive order the intensity value of the pixels of both fixed and moving images, respectively. In a similar way, string $\binom{x}{y} = \left(\binom{x_1}{y_1}, \binom{x_2}{y_2}, \ldots, \binom{x_n}{y_n}\right)$ is obtained by taking alternately the intensity value of the pixels of both superimposed images. This notation has been extracted
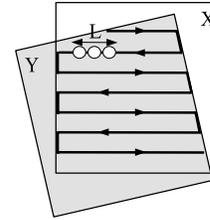

Fig. 3. Scan path on the overlap area between images X and Y.

from [14]. For each iteration in the registration process, the resulting string files are compressed. In our experiments, we use the bzip2 compressor based on the Burrows-Wheeler block-sorting text compression algorithm and Huffman coding. The registration position is achieved minimizing NCD. In this case, $C(x,y)$, $C(x)$, and $C(y)$ are given by the length of files compressed by bzip2. This method has no restrictions on the shape of the overlap area.

B. Shannon Version

From (2), our proposal is to substitute the Kolmogorov complexity by the entropy rate, which gives the average information per intensity value. We assume here that strings x, y and $\binom{x}{y}$ are generated by finite-order stationary Markov sources [14]. Thus, the Shannon version of *the similarity metric* is defined by

$$NED(x,y) = \frac{h^{\binom{x}{y}} - \min\{h^x, h^y\}}{\max\{h^x, h^y\}}, \tag{11}$$

where

$$h^{\binom{x}{y}} = \lim_{L\to\infty} \frac{H\binom{X^L}{Y^L}}{L} = \lim_{L\to\infty} H\left(\begin{array}{c} X_L \\ Y_L \end{array} \middle| \begin{array}{ccc} X_{L-1} & \ldots & X_1 \\ Y_{L-1} & \ldots & Y_1 \end{array}\right) \tag{12}$$

represents the entropy rate of $\binom{x}{y}$, i.e., the degree of compressibility of the two superimposed strings. With this proposal, we make use of neighbor information by considering the correspondence between blocks of pixels instead of the correspondence between individual pixels.

Similar to the bzip case, to generate the blocks of pixels we follow the scan path on the overlap area of the two images to be registered (see Fig. 3). Each block is composed of L consecutive pixels. Note that computations are influenced by the size of the blocks, since the sparsity of the joint histogram and also the computational cost of the process increases with it. To overcome these limitations, reductions of the original alphabet (0..255) to 8 or 16 symbols are appropriate.

IV. RESULTS

In this section, we analyze the performance of the compression-based registration measures.

The first experiment analyzes the JPEG 2000-based registration approach. For this test we use the Lena image of Fig. 2, moving the red channel through the X axis from -10 pixels to 10 pixels around the origin. To guarantee a rectangular overlap, we define a rectangular window as shown in Fig. 2.
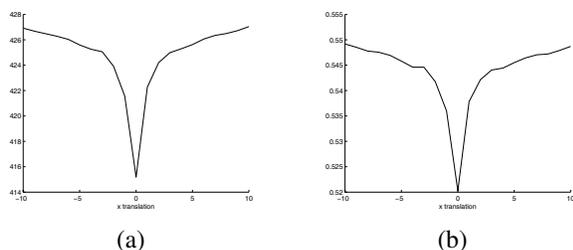
438

Fig. 4. JPEG 2000-based registration of Lena image. (a) Plot of the compressed file length $C(x,y)$. (b) Plot of NCD.
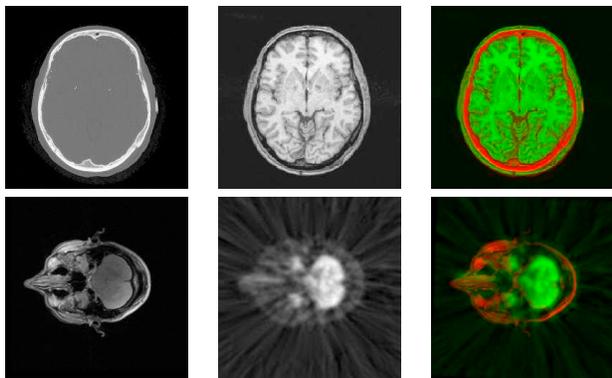

Fig. 5. Top row: CT, MR and CT-MR fused images. Bottom row: MR, PET and MR-PET fused images.

The resulting images are compressed, obtaining $C(x,y)$, $C(x)$ and $C(y)$ from the length of the compressed RGB image, the length of the green-blue image and the length of the red image, respectively. Fig. 4 depicts the obtained compression $C(x,y)$ (a) and the values of the NCD measure (b). Observe that both plots behave similarly, reaching the minimum at the registration position.

For the next three experiments, we use two different pairs of medical images of 256×256 pixels obtained from a patient's brain (see Fig. 5). The first pair consists of a computed tomography (CT), which provides precise anatomical characteristics of the brain, and a magnetic resonance (MR) image, which gives accurate information of the soft tissue. The second pair consists of an MR image and a positron emission tomography (PET), which provides functional information. Registration of these modalities is of special interest for medical diagnosis.

In the tests with the CT-MR pair, the CT image has been taken as a fixed image and the MR image as the moving one, and in the tests with the MR-PET pair, the PET image has been taken as the moving one. In both cases, the moving images are translated through the X axis from -10 pixels to 10 pixels around the registration position. The obtained results for each one of the methods (JPEG 2000, bzip2, and entropy rate) are illustrated, respectively, in Fig. 6 (a), (b) and (c). For the CT-MR pair, Fig. 6(i.a-b) show the compression $C(x,y)$ and Fig. 6(i.c) the entropy rate $h\binom{x}{y}$, and Fig. 6(ii.a-b) the NCD measure and Fig. 6(ii.c) the NED measure. For the MR-PET

pair, Fig. 6(iii.a-b) shows the NCD measure and Fig. 6(iii.c) the NED measure.

To apply the JPEG 2000-based registration method, we define a rectangular window in all test images in order to obtain a rectangular overlap area. Observe that for the MR-CT pair, the compression $C(x,y)$ (Fig. 6(i.a)) has several local minima and the absolute minimum is not at the registration position. On the contrary, the NCD plot (Fig. 6(ii.a)) has a smooth shape, reaching the minimum at the correct position. The undesirable behaviour obtained in the NCD plot of the MR-PET pair (Fig. 6(iii.a)) is due to the high level of noise of the PET images.

From the results obtained with the bzip2-based registration for the MR-CT pair (see Fig. 6(i-ii.b)), it can be seen that the minima of the $C(x,y)$ and NCD plots do not coincide, and nor do the different shape at the extremes of the plots. Observe that the compression $C(x,y)$ decreases with long translations (more than 5 pixels). This behaviour is due to the fact that long translations lead to small overlap regions and a reduction of the information to be compressed. As it can be seen in the NCD plot (Fig. 6(ii.b)), this measure corrects this undesirable behaviour and, moreover, the minimum of the function coincides with the registration position. Observe that the NCD plot obtained with the MR-PET pair (Fig. 6(iii.b)), although it behaves better than the JPEG 2000 method (Fig. 6(iii.a)), it still has some local minima. In our experiments with bzip2, we also observe that $NCD \geq 1$, which is a consequence of bzip2 compressor imperfections (see [11]).

We evaluate the NED measure (Fig. 6(c)) considering three different L values, 1, 2 and 3, represented by solid, dash-dotted and dashed lines, respectively. Due to the high dimensionality of the joint histogram in the L=3 case, the number of bins has been reduced to 8. For comparison purposes, this quantization has been kept in all cases. In Fig. 6(ii-iii.c), the bold plot represents the standard image registration measure NMI. As we expected, entropy rate estimation decreases with L (Fig. 6(i.c)), giving us a more approximate measure of the real entropy rate and, equivalently, the string compressibility. Observe in Fig. 6(ii-iii.c) the smoothness of the NCD curves, without local minima, and the accuracy of the registration, achieving their minimum at the correct position for both MR-CT and MR-PET pairs. Note also that NMI behaves very similar to the entropy rate approach when L=1. This behaviour was also noted by Kraskov et al. [13] in clustering applications. From our experiments, we can conclude that registration based on Shannon information shows more accuracy and robustness than real-world compressor-based registration.

V. CONCLUSIONS AND FUTURE WORK

We have presented a new compression-based framework for image registration using *the similarity metric*. The behaviour of this measure has been analyzed using two different perspectives: the Kolmogorov complexity and the Shannon information theory. Experimental results reveal the good performance of both versions, computed using standard real-world compressors and an estimation of entropy rate, respectively. Due to the
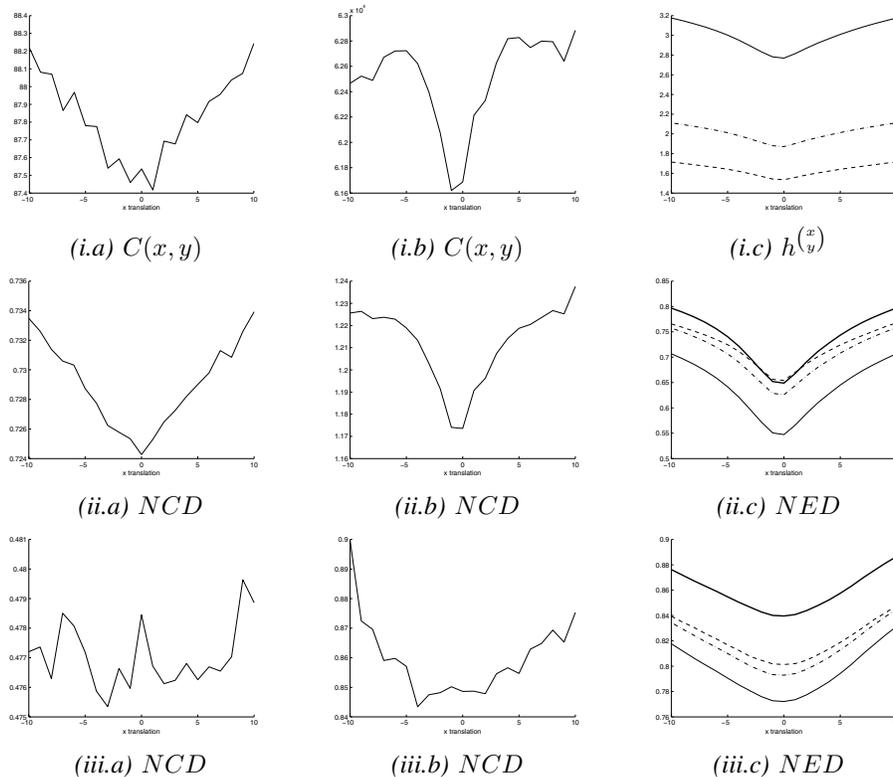
439

(i.a) $C(x, y)$ *(i.b)* $C(x, y)$ *(i.c)* $h\binom{x}{y}$

(ii.a) NCD *(ii.b)* NCD *(ii.c)* NED

(iii.a) NCD *(iii.b)* NCD *(iii.c)* NED

Fig. 6. *(i)* Compression and *(ii)* distance results of MR-CT images, and *(iii)* distance results of MR-PET images, using *(a)* JPEG 2000-based method, *(b)* bzip2-based method, and *(c)* entropy rate-based method.

compressor imperfections, more accurate and robust results are obtained in the second case. In contrast to traditional MI-based registration algorithms, which only use individual pixel information to perform computations, our approaches take into account spatial or structural information contained in the images.

Future work will be addressed to analyze the Shannon version of *the similarity metric* using different scan space-filling curves and uniformly distributed random lines [15] in order to improve the capture of structural information.

ACKNOWLEDGMENTS

REFERENCES

[1] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[2] P. A. Viola, "Alignment by maximization of mutual information," Ph.D. dissertation, Massachusetts Institute of Technology, Massachusetts (MA), USA, 1995.

[3] C. Studholme, "Measures of 3D medical image alignment," Ph.D. dissertation, University of London, London, UK, August 1997.

[4] C. Bennett, P. Gács, M. Li, P. Vitányi, and W. Zurek, "Information distance," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1407–1423, July 1998.

[5] M. Li, J. Badger, X. Chen, S.Kwong, P. Kearney, and H. Zang, "An information-based sequence distance and its applications to whole mitochondrial genome phylogeny," *Bioinformatics*, vol. 17, no. 2, pp. 149–154, 2001.

[6] M. Li, X. Chen, X. Li, B. Ma, and P. Vitányi, "The similarity metric," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3250–3264, December 2004.

[7] R. Cilibrasi, R. de Wolf, and P. Vitányi, "Algorithmic clustering of music," 2003. [Online]. Available: http://arxiv.org/archive/cs/0303025

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[9] D. P. Feldman and J. P. Crutchfield, "Structural information in two-dimensional patterns: Entropy convergence and excess entropy," December 2002. [Online]. Available: arxiv.org/abs/cond-mat/0212078

[10] J. P. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 986–1004, 2003.

[11] R. Cilibrasi and P. M. B. Vitányi, "Clustering by compression," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1523–1545, April 2005.

[12] P. Grassberger, "Data compression and entropy estimates by non-sequential recursive pair substitution." [Online]. Available: http://arxiv.org/abs/physics/0207023

[13] A. Kraskov, H. Stögbauer, R. G. Andrzejak, and P. Grassberger, "Hierarchical clustering based on mutual information," 2003. [Online]. Available: http://arxiv.org/abs/q-bio/0311039

[14] A. Kaltchenko, "Algorithms for Estimating Information Distance with Application to Bioinformatics and Linguistics," April 2004. [Online]. Available: http://arxiv.org/archive/cs/0404039

[15] A. Bardera, M. Feixas, I. Boada, and M. Sbert, "Medical image registration based on random line sampling," in *IEEE International Conference on Image Processing (ICIP'05), Proceedings*, Genova, Italy, September 2005.