Continuous Sampling in Mutual-Information Registration

Mika Seppä

***Abstract*—Mutual information is a popular and widely used metric in retrospective image registration. This metric excels especially with multimodal data due to the minimal assumptions about the correspondence between the image intensities. In certain situations, the mutual-information metric is known to produce artifacts that rule out subsample registration accuracy. Various methods have been developed to mitigate these artifacts, including higher order kernels for smoother sampling of the metric. This study introduces a novel concept of continuous sampling to provide new insight into the mutual-information methods currently in use. In particular, the connection between the partial volume interpolation and the recently introduced higher order partial-volume-type kernels is revealed.**

***Index Terms*—Artifacts, interpolation, mutual information (MI), registration.**

I. Introduction

In image registration, information is spatially aligned from two or more images representing the same object. The images are typically acquired at different times or by different imaging modalities. In the case of medical imaging, the images might convey information about different aspects of the anatomy and physiology in the underlying tissue.

For multimodal image registration, the mutual-information (MI) metric is very popular due to the minimal assumptions about the correspondence between the images; it only assumes that such a correspondence exists. MI is an information-theoretic measure of the dependence of two random variables [1], and it was first proposed and brought to medical imaging field in 1995 [2], [3]. Since then, it has drawn a lot of attention and it has been successfully employed in numerous applications (see survey [4]).

For registering images A and B, the two most common MI registration metrics are

$$I(A,B) = H(A) + H(B) - H(A,B) \tag{1}$$

$$Y(A,B) = \frac{H(A) + H(B)}{H(A,B)} \tag{2}$$

where $H(A,B)$ is the joint (Shannon) entropy [5] derived from the joint-probability distribution $p(a,b)$ of image intensities. This distribution estimates the probability of having intensity a in A and intensity b in B, at the same location. The marginal distributions $p(a)$ and $p(b)$ are computed by summation from $p(a,b)$ and lead to entropies $H(A)$ and $H(B)$. $Y(A,B)$ is the normalized MI measure [6] that is less sensitive to the extent of overlap between the registered images than the regular metric I. The correct registration is sought by maximizing the selected metric. The MI registration typically delivers accurate results [7], although the match metrics can in certain situations suffer from sampling-related artifacts [8]–[10]. With artifacts, the global optimum of the metric becomes more difficult to locate and the optimization could get trapped in a local optimum. Even if the global optimum was reached, these artifacts typically rule out subsample registration accuracy.

Manuscript received October 25, 2007; revised February 6, 2008. This work was supported in part by the Academy of Finland (National Center of Excellence Program). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter C. Doerschuk.

The author is with the Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, FIN-02015 TKK, Espoo, Finland (e-mail: mika.seppa@tkk.fi).

Digital Object Identifier 10.1109/TIP.2008.920738

Several approaches to mitigate these artifacts have been suggested. Smoothness of the match metric can be improved by simple resampling of one of the registered images, by adding sampling jitter to the computation of the metric, or by using histogram smoothing [8], [10]. Oversampling of the images, intensity clustering, and histogram equalization have also been found helpful [9]. Further suggestions include stochastic sampling [11] and combining gradient information in the match metric [12]. Recently, the often used partial volume interpolation [2] method was generalized into higher-order B-spline kernels to reduce the amount of artifacts [13].

This study introduces a novel concept of continuous sampling and shows its connection to the previous methods thereby providing new insight into the higher order kernels and to the symmetry of the registration process. Below, Section II gives background details necessary for the development of the novel continuous sampling concept in Section III. Finally, the interpretations provided by this new sampling scheme are summarized in Section IV.

II. Background

The notation here follows Chen and Varshney [13] although some aspects are shown more explicitly and the variable names are different to better suit the novel theoretical derivation in Section III. For a detailed review of the MI registration and medical image registration in general, see [4] and [14].

A. Coordinates and Transformation

Let A and B be two images that can be considered as two mappings

$$A : \mathbf{x}_\mathrm{d} \rightarrow A(\mathbf{x}_\mathrm{d}), \mathbf{x}_\mathrm{d} \in \mathcal{A}$$

$$B : \mathbf{y}_\mathrm{d} \rightarrow B(\mathbf{y}_\mathrm{d}), \mathbf{y}_\mathrm{d} \in \mathcal{B}$$

where $\mathcal{A}$ is the discrete domain of A, and $\mathcal{B}$ is the discrete domain of B. Vectors $\mathbf{x}_\mathrm{d}$ and $\mathbf{y}_\mathrm{d}$ are the integer coordinates of the samples and the values $A(\mathbf{x}_\mathrm{d})$ and $B(\mathbf{y}_\mathrm{d})$ represent the image intensities at grid-points $\mathbf{x}_\mathrm{d}$ and $\mathbf{y}_\mathrm{d}$, respectively. For both images, a general noninteger coordinate vector $\mathbf{z}$ can be divided into integer $\mathbf{z}_\mathrm{d} = \lfloor \mathbf{z} \rfloor$ and fractional $\mathbf{z}_\mathrm{f} = \mathbf{z} - \lfloor \mathbf{z} \rfloor$ parts where $\lfloor \mathbf{z} \rfloor$ (floor) is a vector of the largest integers that are not greater than the respective components of $\mathbf{z}$.

Finally, let $\mathbf{y} = T_\theta(\mathbf{x})$ be a geometrical transformation with the parameter set θ that maps the coordinates $\mathbf{x}$ of image A into the coordinates $\mathbf{y}$ of image B.

B. Joint Histogram

To evaluate the MI measure $I(A,B)$ or $Y(A,B)$ for the images A and B, the joint probability distribution $p(a,b)$ needs to be estimated. Parzen windowing is one possible method for this estimation, but the majority of the implementations [4] use a straightforward alternative employing a joint histogram. The joint histogram $h(a,b)$ represents the number of times when intensity a in the image A coincides with intensity b in the image B. It is typically accumulated by sampling through all the grid-points of one image and collecting the values at respective places in the other image. Dividing the histogram bins by the sum of all entries normalizes the values and yields the desired joint probability distribution $p(a,b)$.

C. Interpolation

In a registration process, a grid-point in one image does not generally coincide with a grid-point in the other image and one possibility for computing the joint histogram $h(a,b)$ is to employ interpolation. In this case, the intensity value b of the image B is interpolated at location $\mathbf{y} = T_\theta(\mathbf{x}_\mathrm{d})$ where $\mathbf{x}_\mathrm{d}$ is a grid-point of image A. Consequently, the

histogram bin $h(A(\mathbf{x}_\mathrm{d}), b(T_\theta(\mathbf{x}_\mathrm{d})))$ is incremented by 1, and the full histogram is computed by processing through all the pixels $\mathbf{x}_\mathrm{d}$ of the image A that overlap the image B. Typically, simple linear interpolation (LI) is employed for speed although higher-order interpolators can also be used. For more information about different interpolation methods and interpolation kernels, see surveys [15]–[17].

D. Fractional Incrementation

The methods classified here as fractional-incrementation methods are particularly designed for estimating the joint histogram of the MI metric. They increment several histogram bins with fractional amounts instead of calculating an interpolated intensity value and incrementing that bin by 1. This idea was first introduced by the well-known partial volume interpolation (PVI) [2] method which uses linear weights. The idea was later generalized into higher order B-spline kernels [13].

As noted by Chen and Varshney [13], a kernel function $\varphi_\mathrm{fi}(x)$ applied in a fractional-incrementation method needs to fulfill the following two conditions for all real values $x \in \Re$:

$$\varphi_\mathrm{fi}(x) \geq 0 \tag{3}$$

$$\sum_{n=-\infty}^{\infty} \varphi_\mathrm{fi}(n+x) = 1. \tag{4}$$

The first condition (3) guarantees that the histogram contains only positive values and, therefore, leads to a valid joint probability distribution $p(a, b)$. As the weights are drawn from the kernel $\varphi_\mathrm{fi}(x)$ at unit intervals, the second condition (4) guarantees that the same total weight is always distributed. Although nearest-neighbor and linear interpolation kernels fulfill these requirements, interpolating kernels do not in general satisfy the first condition. Higher order kernels, such as Keys's cubic [18] kernel, typically assume also negative values.

The procedure for the fractional incrementation can be mathematically stated as

$$h(A(\mathbf{x}_\mathrm{d}), b_{i,j}(\mathbf{x}_\mathrm{d})) \mathrel{+}= w_{i,j}(\mathbf{y}_\mathrm{f}) \quad \forall\, i,j \in \mathcal{Z}, \quad \forall\, \mathbf{x}_\mathrm{d} \in \mathcal{A} \tag{5}$$

where the operator "+=" denotes incrementing the left-hand side with the value on the right, and $\mathcal{Z}$ is the set of all integers. The intensities $b_{i,j}(\mathbf{x}_\mathrm{d})$ and weights $w_{i,j}(\mathbf{y}_\mathrm{f})$ are defined as

$$\mathbf{y} = T_\theta(\mathbf{x}_\mathrm{d}), \quad \mathbf{y}_\mathrm{f} = \mathbf{y} - \lfloor \mathbf{y} \rfloor \tag{6}$$

$$w_{i,j}(\mathbf{y}_\mathrm{f}) = \varphi_\mathrm{fi}(\mathbf{y}_\mathrm{f} \cdot \mathbf{e}_1 - i)\varphi_\mathrm{fi}(\mathbf{y}_\mathrm{f} \cdot \mathbf{e}_2 - j) \tag{7}$$

$$b_{i,j}(\mathbf{x}_\mathrm{d}) = B(\lfloor \mathbf{y} \rfloor + i\mathbf{e}_1 + j\mathbf{e}_2). \tag{8}$$

In these equations, $\mathbf{e}_k$ is a unit vector along dimension k and $\mathbf{y}_\mathrm{f} \cdot \mathbf{e}_1$ is the dot-product (inner-product) of vectors $\mathbf{y}_\mathrm{f}$ and $\mathbf{e}_1$. As is obvious, $\mathbf{y}$ and $\mathbf{y}_\mathrm{f}$ are functions of $\mathbf{x}_\mathrm{d}$ although not explicitly written out in the equations. The support of the kernel, i.e., the interval where $\varphi_\mathrm{fi}(x)$ assumes nonzero values, determines the number of histogram bins that receive nonzero incrementation for each $\mathbf{x}_\mathrm{d}$ and, therefore, affects the computational complexity of the method.

E. Artifacts

In certain situations, the two most used joint histogram estimation methods, LI and PVI, produce artifacts in the MI metric [8]–[10]. These artifacts manifest especially when the axes of the sample grids are parallel and the sample spacing is equal in the images; these conditions are called "grid alignment" from here on. Fig. 1 shows an example of the MI measure $I(A, B)$ for translation (left) and for rotation (right) under such conditions. As can be seen in the translational case, LI has a peak between every sample-aligned position whereas PVI peaks at the sample-aligning positions. For rotation, similar artifacts occur close to the rotation angle 0° where the axes become parallel. Such local peaks decrease the registration accuracy in these situations. The bottom graph in both panels shows the value for a support-4 third-order B-spline kernel [13] mentioned in Section IV.

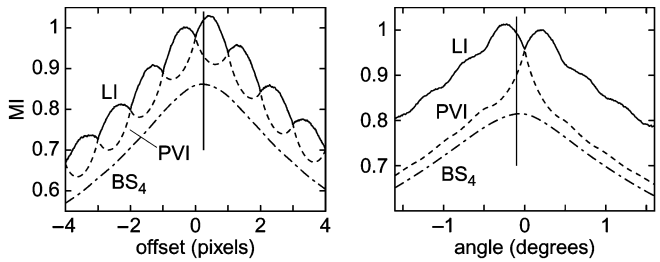

Fig. 1. Artifacts in the MI metric as a function of (left) translation and (right) rotation with different joint-histogram sampling methods: linear interpolation (LI), partial volume interpolation (PVI), and support-4 B-spline method (BS_4). The vertical lines show the true offset at 0.25 pixels for translation and true angle at -0.1 degrees for rotation where the maxima should occur.

The origin of these artifacts in the interpolation methods (such as LI) and in the fractional-incrementation methods (such as PVI) is different. For the interpolation methods, the artifacts are believed to be induced by interpolation errors that blur the images and affect small details and noise [8], [10]. Interpolation can also create completely new intensity values that do not occur in the original image and thereby spread the joint histogram. Therefore, the artifacts cannot be totally removed even by better quality interpolation methods. These effects vary with the grid alignment.

The fractional-incrementation methods do not create new intensity values but they transfer the histogram weights gradually from one original sample to another. In this case, the artifacts are attributed to the nonlinearity of the entropy H [8]; for a nonzero probability p, the entropy of a single histogram bin $H(p)$ is smaller than the sum of the entropies of the two bins into which the probability is distributed, for example $H(p) < 2H(p/2)$. Once again, the amount of the joint-histogram and joint-probability spread depends on the grid alignment.

F. Symmetry

Although the MI measures $I(A, B)$ and $Y(A, B)$ are theoretically symmetrical with respect to the images, the implementational details cause asymmetry to the MI registration. As mentioned above, the histogram accumulation usually processes through all the grid-points $\mathbf{x}_\mathrm{d} \in \mathcal{A}$ that overlap the image B. Therefore, the roles of the images A and B are not interchangeable; assuming that the inverse transform $\mathbf{x} = T_\theta^{-1}(\mathbf{y})$ exists, the evaluation of the registration measure at grid-points $\mathbf{y}_\mathrm{d} \in \mathcal{B}$ of the image B with the corresponding locations $\mathbf{x} = T_\theta^{-1}(\mathbf{y}_\mathrm{d})$ in the image A typically yields a slightly different result.

Symmetry of the registration metric is conceptually desirable since the roles of the registered images should be equal. If the inverse transform $T_\theta^{-1}(\mathbf{y})$ exists, symmetrical treatment can be achieved by employing metrics that interchange the roles of the images, for example $Y(A, B) + Y(B, A)$. However, the registration results between the symmetric and the asymmetric treatment are typically small and, therefore, do not alone justify the extra computational cost. The registration errors due to the sampling-related artifacts are much larger and are not a consequence of the symmetry or the lack thereof.

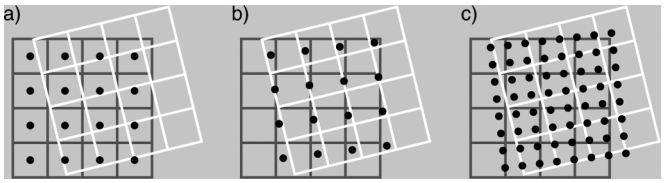

Fig. 2. Illustration of sampling the MI metric of two images (gray and white grids) (a) at centers of pixels, (b) at off-center points, and (c) with double density at off-center points.

III. Continuous Fractional Incrementation

One known trick [8], [10] to reduce MI artifacts is to replace one of the registered images with a resampled version so that the sample spacings become different. As mentioned, the typical joint-histogram estimation methods sample through all the original grid-points $\mathbf{x}_\mathrm{d}$ of image A. The resampling trick suggests that a similar improvement could be possible by changing these histogram sampling locations instead of preprocessing one or both of the images. Ultimately, these sampling locations could be spaced infinitely tight to produce continuous sampling.

Fig. 2 illustrates the change of sampling locations from pixel centers of image A (a) to off-center points (b) and (c). Let $\mathcal{S}$ be the set of the sampling locations used for histogram accumulation and not necessarily coinciding with the pixel centers of either image. In this case, both images have to be treated equally and the fractional-incrementation procedure becomes

$$h(a_{k,l}(\mathbf{x}), b_{i,j}(\mathbf{x})) \mathrel{+}= w^A_{k,l}(\mathbf{x}_\mathrm{f}) w^B_{i,j}(\mathbf{y}_\mathrm{f}) \quad \forall\, i,j,k,l \in \mathcal{Z}, \quad \forall\, \mathbf{x} \in \mathcal{S} \tag{9}$$

where

$$\mathbf{y} = T_\theta(\mathbf{x}), \quad \mathbf{x}_\mathrm{f} = \mathbf{x} - \lfloor \mathbf{x} \rfloor, \quad \mathbf{y}_\mathrm{f} = \mathbf{y} - \lfloor \mathbf{y} \rfloor \tag{10}$$

$$w^A_{k,l}(\mathbf{x}_\mathrm{f}) = \varphi_\mathrm{fi}(\mathbf{x}_\mathrm{f} \cdot \mathbf{e}_1 - k)\varphi_\mathrm{fi}(\mathbf{x}_\mathrm{f} \cdot \mathbf{e}_2 - l) \tag{11}$$

$$w^B_{i,j}(\mathbf{y}_\mathrm{f}) = \varphi_\mathrm{fi}(\mathbf{y}_\mathrm{f} \cdot \mathbf{e}_1 - i)\varphi_\mathrm{fi}(\mathbf{y}_\mathrm{f} \cdot \mathbf{e}_2 - j) \tag{12}$$

$$a_{k,l}(\mathbf{x}) = A(\lfloor \mathbf{x} \rfloor + k\mathbf{e}_1 + l\mathbf{e}_2) \tag{13}$$

$$b_{i,j}(\mathbf{x}) = B(\lfloor \mathbf{y} \rfloor + i\mathbf{e}_1 + j\mathbf{e}_2). \tag{14}$$

Obviously, the amount of calculations and the number of histogram bins updated are significantly increased. If the support of the selected kernel $\varphi_\mathrm{fi}(x)$ is L and data dimensions are D, the normal fractional-incrementation procedure updates L^D bins for each sample location and the procedure above updates L^{2D} bins. Furthermore, if the sampling density in $\mathcal{S}$ is increased, as in Fig. 2(c), the computational burden increases drastically.

An analytical solution equivalent to continuous sampling can be sought by decreasing the sample spacing to infinitesimally small. The solution depends on the selected geometrical transformation $T_\theta(\mathbf{x})$ and its parameter vector θ and is, therefore, difficult to obtain in a general case. However, since the worst artifacts manifest when the image axes are parallel and the sample spacing is equal, we can search for a solution in this special case with a simple translational transformation.

For simplicity, let us discard possible image edge and overlap effects and assume A and B to be infinitely long 1-D images. Furthermore, let the sampling set $\mathcal{S}$ to consist of equispaced locations at interval Δx. For pure translation, (9)–(14) become

$$h(a_k(x), b_i(x)) \mathrel{+}= \varphi_\mathrm{fi}(x_\mathrm{f} - k)\varphi_\mathrm{fi}(y_\mathrm{f} - i)\Delta x \quad \forall\, i,k \in \mathcal{Z}, \quad \forall\, x \in \mathcal{S} \tag{15}$$

$$y = T_\theta(x) = x + \theta \tag{16}$$

$$x_\mathrm{f} = x - \lfloor x \rfloor, \quad y_\mathrm{f} = y - \lfloor y \rfloor \tag{17}$$

$$a_k(x) = A(\lfloor x \rfloor + k) \tag{18}$$

$$b_i(x) = B(\lfloor y \rfloor + i). \tag{19}$$

Note that in (15), the spacing Δx is added as an extra weight. Since the histogram is normalized to obtain the joint-probability distribution $p(a,b)$, this weight does not affect the MI metric but it simplifies the calculations. As mentioned, we are interested what happens when $\Delta x \to 0$.

Let us focus on the histogram bins that a single pixel $A(0)$ contributes to, together with any pixel $B(v)$. In such a case, (18) and (19) yield $k = -\lfloor x \rfloor$ and $i = v - \lfloor y \rfloor$. Inserting these, (15) simplifies to

$$h(A(0), B(v)) \mathrel{+}= \varphi_\mathrm{fi}(x)\varphi_\mathrm{fi}(x + \theta - v)\Delta x \quad \forall\, v \in \mathcal{Z}, \quad \forall\, x \in \mathcal{S}. \tag{20}$$

Now, summing the incrementation over all $x \in \mathcal{S}$ when the sampling density is increased as $\Delta x \to 0$, results in integral

$$h_c(A(0), B(v)) \mathrel{+}= \int_{-\infty}^{\infty} \varphi_\mathrm{fi}(x)\varphi_\mathrm{fi}(x + \theta - v)dx \quad \forall\, v \in \mathcal{Z} \tag{21}$$

$$h_c(A(0), B(v)) \mathrel{+}= \varphi_\mathrm{fi} \star \varphi_\mathrm{fi}(\theta - v), \quad \forall\, v \in \mathcal{Z}. \tag{22}$$

Here, $\varphi_\mathrm{fi} \star \varphi_\mathrm{fi}$ denotes the cross correlation of kernel φ_fi with itself (i.e., auto-correlation) and the subscript c indicates that the equation is for the continuous sampling. In a similar fashion, the normal fractional-incrementation method (5) can be simplified in the 1-D case of a simple translational transformation. The incrementation related to the pixel $A(0)$ then becomes

$$h(A(0), B(v)) \mathrel{+}= \varphi_\mathrm{fi}(\theta - v), \quad \forall\, v \in \mathcal{Z}. \tag{23}$$

The comparison of (22) and (23) reveals that, in this special case of equal sample spacing with pure translational transformation, continuous sampling with kernel φ_fi is equivalent to the normal fractional-incrementation procedure with kernel $\varphi_\mathrm{fi} \star \varphi_\mathrm{fi}$.

The above result, obtained in one dimension, generalizes in a straight-forward fashion into higher dimensional cases. Although derived for infinitely long images, (22) holds also for the vast majority of the pixel pairs $A(\mathbf{x}_\mathrm{d})$ and $B(\mathbf{y}_\mathrm{d})$ in normal images as long as these pixels are further away from the image overlap borders (or from each other) than the support size of the kernel $\varphi_\mathrm{fi}(x)$. Typically, image sizes are much larger than the kernel support size.

Since the cross correlation of two φ_fi kernels yields a valid kernel function for the fractional incrementation method (see Appendix for a proof), that kernel can be employed with any transformation T_θ and not only with translational ones. Whenever the applied transformation T_θ approaches the grid alignment, the auto-correlated kernel $\varphi_\mathrm{fi} \star \varphi_\mathrm{fi}$ can be interpreted as performing continuous sampling with the φ_fi kernel. In a general case, this interpretation is lost or is only "approximately valid" when the transformation is close to the grid alignment, but the auto-correlated kernel can be nevertheless employed.

IV. Summary

The introduced continuous-sampling scheme allows the creation of new kernels and provides some insight into the methods already in use. In the grid alignment, the use of the well-known PVI method is equivalent to employing nearest-neighbor kernel with continuous sampling. Therefore, the two registered images are actually treated symmetrically under these conditions and their roles are thus interchangeable,

although that is not clear from the procedure itself (Section II-D). In a similar fashion, any kernel that is an auto-correlation result of another valid kernel treats the images symmetrically in the grid alignment. As is now clearly evident, the lack of symmetry is not related to the sampling artifacts since the PVI method experiences its most severe artifacts when it is actually symmetric.

Starting from the nearest-neighbor kernel, auto-correlation provides PVI and the other higher-order B-spline kernels. In general, all the B-spline kernels of even support can be created by auto-correlation from a B-spline kernel with half of that support. Although these kernels have been suggested before [13], the derivation above brings new insight into their symmetry and helps to better understand their behavior by connecting them directly to the lower-order kernels. In particular, the auto-correlation of the PVI kernel produces the support-4 third-order B-spline kernel. As was experimented by Chen and Varshney [13] and also illustrated in Fig. 1, this kernel (BS_4) produces significantly smaller artifacts than the LI and the PVI methods. Examples of the ideas in this paper applied to registration accuracy tests can be found in [19].

APPENDIX
CROSS-CORRELATION OF KERNELS

Let $\varphi_a(x)$ and $\varphi_b(x)$ be two kernel functions that satisfy the conditions of (3) and (4). Their cross correlation gives

$$\varphi_c(x) = \varphi_a \star \varphi_b = \int_{-\infty}^{\infty} \varphi_a(t)\varphi_b(x+t)dt.$$

Clearly, $\varphi_c(x) \geq 0$ if both $\varphi_a(x) \geq 0$ and $\varphi_b(x) \geq 0$, for all x (3). It is easy to see that the second condition (4) is also true for φ_c

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} \varphi_c(n+x) &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi_a(t)\varphi_b(n+x+t)dt \\
&= \int_{-\infty}^{\infty} \varphi_a(t) \left\{ \sum_{n=-\infty}^{\infty} \varphi_b(n+x+t) \right\} dt \\
&= \int_{-\infty}^{\infty} \varphi_a(t)dt \\
&= \sum_{n=-\infty}^{\infty} \int_{0}^{1} \varphi_a(n+t)dt \\
&= \int_{0}^{1} \left\{ \sum_{n=-\infty}^{\infty} \varphi_a(n+t) \right\} dt = 1.
\end{aligned}$$

Above, the summations inside the curly braces are equal to one due to the condition of (4), and the second to last row was acquired by dividing the previous integral into unit length pieces.

ACKNOWLEDGMENT

The author would like to thank Prof. R. Hari, L. Parkkonen, and Dr. K. Uutela for comments on the manuscript and the anonymous reviewers for their valuable suggestions to improve the paper.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
[2] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," in *Information Processing in Medical Imaging*, Y. Bizais, C. Barillot, and R. Di Paola, Eds. Dordrecht, The Netherlands: Kluwer, 1995, pp. 263–274.
[3] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," in *Proc. 5th Int. Conf. Comput. Vision*, Jun. 1995, pp. 16–23.
[4] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
[5] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul. 1948.
[6] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.
[7] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl, "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations," *Med. Image Anal.*, vol. 1, no. 3, pp. 195–206, Apr. 1997.
[8] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Interpolation artefacts in mutual information-based image registration," *Comput. Vis. Image Understand.*, vol. 77, no. 2, pp. 211–232, Feb. 2000.
[9] J. X. Ji, H. Pan, and Z.-P. Liang, "Further analysis of interpolation effects in mutual information-based image registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1131–1140, Sep. 2003.
[10] J. Tsao, "Interpolation artifacts in multimodality image registration based on maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 854–864, Jul. 2003.
[11] M. Unser and P. Thévenaz, "Stochastic sampling for computing the mutual information of two images," in *Proc. 5th Int. Workshop Sampl. Theory Appl.*, Strobl, Austria, May 2003, pp. 102–109.
[12] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Trans. Med. Imag.*, vol. 19, no. 8, pp. 809–814, Aug. 2000.
[13] H. Chen and P. K. Varshney, "Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1111–1119, Sep. 2003.
[14] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.*, vol. 46, no. 3, pp. R1–R45, Mar. 2001.
[15] T. M. Lehmann, C. Gönner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049–1075, Nov. 1999.
[16] T. M. Lehmann, C. Gönner, and K. Spitzer, "Addendum: B-spline interpolation in medical image processing," *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 660–665, Jul. 2001.
[17] E. H. Meijering, W. J. Niessen, and M. A. Viergever, "Quantitative evaluation of convolution-based methods for medical image interpolation," *Med. Image Anal.*, vol. 5, no. 2, pp. 111–126, Jun. 2001.
[18] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.
[19] M. Seppä, "Quality improvements for multi-modal neuroimaging," Ph.D. dissertation, Dept. Eng. Phys. Math., Helsinki Univ. Technol., Espoo, Finland, 2007.