PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Pattern Recognition 37 (2004) 2361–2373

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Mutual information aspects of scale space images

A. Kuijper*

IT University of Copenhagen, Rued LanggaardsVej 7, DK-2300 Copenhagen S2300, Denmark

Received 15 October 2003; accepted 15 April 2004

Abstract

In image registration, mutual information is a well-performing measure based on principles of uncertainty. Similarly, in image analysis the Gaussian scale space, based on minimal assumptions of the image, is used to derive intrinsic properties of an image. This paper starts an investigation of a combination of both methods. This combination results in a double parameterized mutual information measure using local information of the image. For single modality matching best response is found for coinciding parameters. Then critical values are found for which the parameterized mutual information has extrema. First results on multi-modality matching show that different parameter values instead of coinciding values yield the best response for the parameterized mutual information.

Keywords: Scale space; Mutual information; Entropy; Registration; Image analysis; Multiresolution; Image structure

1. Introduction

If two images are derived from a scene, they will mostly be (slightly) different due to change of, for instance, illumination, position, and the scene itself. If the images are taken with different kind of cameras the differences will even be larger, since then the correspondence in intensity is not an a priori relevant correspondence. Although a difference in position is sometimes an advantage and used in stereo images, in most cases these differences are not desired and it is a non-trivial task to find the correspondence (registration) between the two images, say A and B. As an example, one can think of medical images of different modality, like magnetic resonance (MR) images (showing the "soft" structures), computed tomography (CT) images (showing the "hard" structures), and positron emission tomography (PET) images (showing "activity"). The reason for taking both kinds of images of the same patient is the need for determining a specific location of soft tissue (e.g. a brain tumor) in correspondence to the skull.

This correspondence is obtained in a threefold manner, containing transformation, comparison and evaluation. Usually an optimization scheme starting with an initial guess of the correspondence, is used to obtain the best solution.

Firstly, a *transformation*, say T, is applied to image A, that is supposed to match image B as close as possible in some sense. If T contains one rotation and translation that hold for the complete image, T is called a ridged transformation.

Secondly, the transformation $T(A)$ needs to be *compared* to B. Since images are discrete by nature, transformations imply comparison between the pixels and one obviously needs to use some interpolation scheme. In other words, the images are assumed to be continuous.

The third step is to *evaluate* the correctness of the transformation. This requires an appropriate measure on the set $(T(A), B)$. The *mutual information* (MI) measure, using the entropies of the images and their joint entropy, has proven to perform well [1–6] in the evaluation of the registration of two images. Generally, one of the images is transformed to obtain the best result and one needs to use some

* Tel.: +45-38-16-8888; fax: +45-38-16-88-99.
E-mail address: arjan@itu.dk (A. Kuijper)
URL: http://www.itu.dk/people/arjan/.

0031-3203/$30.00
doi:10.1016/j.patcog.2004.04.014

interpolation scheme, thus assuming that the images are continuous. The use of the entropy implies the use of uncertainty, or "*lack of knowledge*".

On the other hand, the acquisition of an image introduces the notion of *scale*. One can think of it in some sense as the resolution of the image: a binary (newspaper) image at a certain scale, for instance, looks like a grayscale image. The evaluation of an image under (almost) trivial assumptions, derived from the statement "*we know nothing of the image*", leads to the notion of *scale space* [7–11]. One advantage in this context is that the *discrete image* becomes a *continuous scale space image*. The scale space image contains sufficient information for an uncommitted hierarchy and segmentation without a priori knowledge [12,13].

Research has been done on the use of scale space and entropy, but these investigations are based on the scale space of the histogram of the image [14,15], or the Gaussian pyramid [16]. Pluim et al. [17,18] implemented a multi-scale approach to mutual information matching, aiming for an acceleration of the matching process while considering the accuracy and robustness of the method. They found an acceleration up to a factor of around 3.

Since both methods start of by taking "minimal assumptions", it makes sense to investigate the combination of them. The research will lead to the underlying reasons for the speed up mentioned by Pluim et al. [17,18]. Furthermore, the use of scale space allows one to embed local neigbourhood information (image structure) into the registration measure, since images at increasing scales become more and more blurred. In this way, there is no need for changing the registration measure with derivative information, as proposed by Pluim et al. [19].

In this paper, we start this study and see how the MI measure applied to single images as well as multimodality images, is affected by using a scale space. We will focus on this aspect, since the behaviour of the MI measure at a particular scale on the optimization procedure has already been studied by Pluim et al. [17,18]. So this study complements theirs by investigating the method itself. The MI measure and scale space are briefly summarized in Section 2. Their combination is described in Section 3. In Section 4 the results of this approach are presented on MR, CT, and PET images, as well as their combinations. In Section 5 results are summarized and discussed.

2. Theory

Next I will briefly describe the ideas and theory behind the concepts of mutual entropy in Section 2.1 and scale space in Section 2.2. The combination and the application of MI to scale space images will be described in Section 3.

2.1. Mutual information

The MI measure [1,4,3] is based on the shared information of the overlapping part of two images. This information is obtained using the Shannon entropy [20], known as a measure of uncertainty. Alternative paradigms, based on the difference of conditional entropy of one image gives a second one, and the latter, as well as based on the so-called Kullback–Leibner distance, can be given and yield the same definition [3].

Let A be a random variable and $p_A(a)$ its marginal probability distribution, then the entropy $H(A)$ is given by

$$H(A) = -\sum_{a\in A} p_A(a)\log\, p_A(a).$$

Let B, $p_B(b)$, and $H(B)$ be similarly defined and $p_{AB}(a,b)$ be the joint marginal probability distribution, then the joint entropy $H(A,B)$ is given by

$$H(A,B) = -\sum_{a\in A,b\in B} p_{AB}(a,b)\log\, p_{AB}(a,b)$$

and $\mathrm{MI}(A,B)$—in terms of the Kullback–Leibner distance—is defined by

$$\mathrm{MI}(A,B) = \sum_{a\in A,b\in B} p_{AB}(a,b)\log\frac{p_{AB}(a,b)}{p_A(a)p_B(b)}.$$

Expanding the division and multiplication in the logarithm and the sum over these terms yields

$$\begin{aligned}\mathrm{MI}(A,B) = &\sum_{a\in A,b\in B} p_{AB}(a,b)\log p_{AB}(a,b)\\ &-\sum_{a\in A,b\in B} p_{AB}(a,b)\log\, p_A(a)\\ &-\sum_{a\in A,b\in B} p_{AB}(a,b)\log\, p_B(b).\end{aligned}$$

Since $\sum_{a\in A,b\in B} p_{AB}(a,b)=\sum_{b\in B} p_B(b)=\sum_{a\in A} p_A(a)=1$, the latter reduces to

$$\begin{aligned}\mathrm{MI}(A,B) = &\sum_{a\in A,b\in B} p_{AB}(a,b)\log\, p_{AB}(a,b)\\ &-\sum_{a\in A} p_A(a)\log\, p_A(a)\\ &-\sum_{b\in B} p_B(b)\log\, p_B(b)\end{aligned}$$

and thus—in terms of the joint entropy

$$\mathrm{MI}(A,B) = -H(A,B) + H(A) + H(B). \tag{1}$$

Since the conditional entropy of B given A is given by $H(B|A) = H(A,B) - H(A)$, one can as well write $\mathrm{MI}(A,B) = -H(B|A) + H(B)$, being the conditional paradigm.

If A and B are completely uncorrelated, $H(A,B) = H(A) + H(B)$ and therefore $\mathrm{MI}(A,B) = 0$. In the opposite case, if A and B are fully correlated, i.e. $A = B$, $H(A,B) = H(A) = H(B)$ and $\mathrm{MI}(A,A) = H(A)$.

Since this is a priori unknown value, in many tasks the normalized mutual information (NMI) is used. It is defined by

$$\mathrm{NMI}(A,B) = \frac{H(A)+H(B)}{H(A,B)} = \frac{\mathrm{MI}(A,B)}{H(A,B)} + 1 \qquad (2)$$

and is bounded by 1 (uncorrelated) and 2 (correlated).

This can easily be applied to the analysis of image registration. Let A and B denote the images that need to be registered, after the alignment has taken place. The marginal probability functions are derived from the *relative histogram* of the intensities in each image: this is the histogram of all the pixel (or voxel) values of the image where the height of each bin (entry) of the histogram is divided by the total number of considered values. The joint marginal probability is represented by a 2D relative histogram in which each bin (a,b) the relative number of corresponding image points (pixels, voxels, etc.) are stored: the fraction of pixels with intensity a in image A and intensity b in image B. These image points obviously need to have the same spatial location (after translation, rotation, and, if necessary, interpolation). As a consequence, if $A = B$ and both histograms have equal binwidth, the 2D histogram has only non-zero entries on the diagonal.

More details on the MI and NMI measures can be found in aforementioned references, [1–6,17–19,21], of which [21] gives a nice extended introduction of both measures and their impact on medical image alignment, and [3] an extensive survey of the use of MI in medical imaging.

2.2. Scale space

Generally, there is no a priori reason to treat the image on pixel scale, since it is only a result of the acquisition of the image by an aperture with a certain scale or sampling range. This becomes clearly an important factor when two different images are compared. Even when the same image is considered, but with different resolutions, the need of "equal resolution" is evident: the best match is obtained when they have equal resolution.

Moreover, the assumption that the image can be regarded as a continuous function of its spatial variables is only mathematically justified, if this discrete image is convolved with a so-called test function. This highly mathematical result is obtained within the "Theory of Distributions", started by Schwartz in the 1950s [22].

One of the simplest and mostly used test functions is the Gaussian filter with zero mean and variable variance [7]. The latter is very important: since there is no a priori reason to fix the variance, *all* possible variances (i.e. scales) are to be taken. Therefore the n-dimensional image $L(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}_0^+$ is extended to an $(n+1)$-dimensional scale space image $L(\mathbf{x};t) : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}_0^+$ by

$$L(\mathbf{x};t) = \int_{\mathbb{R}^n} \frac{1}{\sqrt{4\pi t}^n} \mathrm{e}^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{4t}} L(\mathbf{y})\,\mathrm{d}\mathbf{y}. \qquad (3)$$

As a result, the scale space image satisfies the well-known heat equation, $\partial_t L(\mathbf{x};t) = \Delta L(\mathbf{x};t)$, where ∂_t denotes the partial derivative with respect to t, and Δ the Lapacean, defined by $\partial_{x_1x_1} + \partial_{x_2x_2} + \cdots + \partial_{x_nx_n}$.

Commonly, instead of t, the variance σ, with $\sigma = \sqrt{2t}$, is considered. From Eq. (3) it is clear that when regarding "scale σ", scale and spatial location have the same dimensions. It thus makes sense to speak of 'a scale of three pixels'. This corresponds to regarding the image blurred with $\sigma = 3$ and thus $t = 4.5$. Therefore, besides the advantage of transferring the *discrete* data into a *continuous* setting, one is also able to embed *local information* into the image observation, and therefore into the registration measure.

For more details on (Gaussian) scale space and its properties, the reader is referred to the ample literature on this subject, e.g. in Refs. [7–13,23].

3. Mutual information in scale space

The basic idea of MI in scale space is that applying a Gaussian filter to an image followed by binning, introduces spatial information into the histograms. The intensity of an image point that is put into a certain bin is partially determined by the intensity of that point in the original image, and partially by the (intensity-) environment around that image point. The larger the scale of the filter, the larger the local environment of a pixel that is taken into account. Since convolution with a Gaussian is a way of low-pass filtering, noise is suppressed and the image is smoothened. The result on the histogram is twofold. It also blurs, albeit not necessarily Gaussian, due to smoothening of edges. On the other hand, the suppression of noise yields ideally elimination of irrelevant (noise induced) intensities, i.e. sharpening, or at least more pronouncing, of relevant (non-noisy) clusters.

Note that changing the bin size is a completely different approach. In that case the neighbouring values in the *histogram* are taken together, regardless where they are neighbouring pixels in the *image*. So in this case non-local spatial information is taken into account.

In the literature, applying a scale space on intensities themselves is also proposed. In this case the intensity is partly put in its bin and partly in its neighbouring bins, depending on the width of the Gaussian filter used. This may be called fuzzy binning: the measured values are stored in bins determined by the *blurring of the values* themselves, cf. [24–26]. Although this may give interesting results for single images, the transfer of this idea to the joint histogram—and thus the mutual information measure—is complicated.

In these both case the values in the one dimensional signal called *histogram* are blurred, while in the presented approach the values in the *image* are blurred. It is therefore essentially different.

In the scale space image, the MI becomes a parameterized measure. Let A and B become the scale space images A_σ and B_τ, then the entropies and joint entropy become $H(A_\sigma)$, $H(B_\tau)$, and $H(A_\sigma, B_\tau)$. Consequently, Eqs. (1) and (2) become

$$\mathrm{MI}(A_\sigma, B_\tau) = -H(A_\sigma, B_\tau) + H(A_\sigma) + H(B_\tau) \tag{4}$$

and

$$\mathrm{NMI}(A_\sigma, B_\tau) = \frac{H(A_\sigma) + H(B_\tau)}{H(A_\sigma, B_\tau)}. \tag{5}$$

Now two different routes to investigate the behaviour of these parameterized measures seem to be possible. The first one is to construct the scale space images, derive the histograms at all scales and compare the histograms and joined histogram at all these scales. An alternative way is to construct the scale space images and derive histograms at all scales simultaneously. In this case the scales are regarded as an extra variable adding an extra dimension to the image, instead of a parameter.

Although both methods appear different, the latter is a special case of the first. This is due to the fact that the histogram of the scale space image (the second route) equals the sum of the histograms at all the calculated scales (obtained in the first route) divided by the number of scales. It therefore suffices to take into account only the first case.

Several aspects of this parameterized measure are investigated. Firstly, the case $A = B$ is taken. This gives insight in the behaviour of the (normalized) mutual information under the influence of scale. Then Eqs. (4) and (5) become

$$\mathrm{MI}(A_\sigma, A_\tau) = -H(A_\sigma, A_\tau) + H(A_\sigma) + H(A_\tau) \tag{6}$$

and

$$\mathrm{NMI}(A_\sigma, A_\tau) = \frac{H(A_\sigma) + H(A_\tau)}{H(A_\sigma, A_\tau)}. \tag{7}$$

The special case $\sigma = \tau$, yielding the one parameter measure $\mathrm{MI}(A_\sigma, A_\sigma) = H(A_\sigma)$ for Eq. (6), since then $H(A_\sigma, A_\sigma) = H(A_\sigma)$, is also dealt with. It describes the behaviour of the entropy of the histogram under blurring. Note that in this case the NMI of Eq. (7) reduces to 2 for all σ. Secondly, the general case $A \neq B$, is investigated.

4. Results

In this Section 1 consider the 2D test images as shown in Fig. 1. They are, from left to right, a MR image, a CT image, and a PET image and are representative for the different modalities. These images have dimensions 248 × 253, are already pair-wise aligned, and have integer values ranging from 0 to 255. Consequently, histograms are taken with bandwidth 1. Scales s and t are taken in pixels. In subsequent sections firstly the single modality behaviour MR–MR, CT–CT, and PET–PET is investigated. Secondly, the bimodal results for the MR–CT, MR–PET, and CT–PET registrations are calculated. Here the images are perfectly registered, but both MI and NMI are calculated as a function of both varying scales (the set (A_σ, B_τ)).

4.1. Entropy of the scaled images

In each case the behaviour of the MI as function of increasing scale are presented. Here the images are perfectly registered (the set (A_σ, A_σ)). Then the MI reduces to the entropy of the image ($H(A_\sigma)$) and the NMI to 2 (and is thus disregarded), as was shown in Section 2.1. These results are followed by the behaviour of the MI and NMI as function of both linear translation of one of the images and increasing scale (the set $\{T(A_\sigma), A_\sigma\}$), i.e. the effect of scale on a misregistration. The results can be explained by the effects of increasing scale to each image and its histogram, which are also presented and discussed. The last unimodal registration comparison is by perfectly registered images, but as a function of both varying scales (the set (A_σ, A_τ)).

4.1.1. MR image

Of the MR image shown in Fig. 1a a scale space is calculated. The first and second row of Fig. 2 show the image at several scales. The corresponding histograms with binwidth 1 are shown in the third and fourth row of this figure.

4.1.1.1. Small scale behaviour. Firstly, a scale space containing small scales is analyzed. This scale space has 29 scales, ranging linearly from 1 to 8 pixels. At each scale the MI of the image with itself is calculated and shown in Fig. 3a.

One may note that initially the MI decreases until a scale of 3 pixels, remains constant for, say, two pixels, and then increases again. The histograms of the image at scales 1, 3, and 5 pixels was shown in the third row of Fig. 2. The decrease can be explained by noting that the bins b_1, b_2, and b_3 with fractions 0.14, 0.31, and 0.19 are reorganized to the more pronounced bin sequence 0.53, 0.12, and 0.1. This deletion of noise in the background influences the change in MI more than the decreasing number of filled bins, from 227 to 178. With increasing scale this number decreases more to 143 bins, causing an increasing value of the MI.

4.1.1.2. Large scale behaviour. To investigate the large scale behaviour, scales varying from 1 to 100 pixels are taken. The MI of the image with itself is shown in Fig. 3b. Clearly, the increase of MI changes to a decrease at a scale of 33 pixels.

An explanation of this fact can be found in Fig. 2. The second row shows the image at scales 20, 33, and 60. Where there is still some structure visible at scale 20, at the other

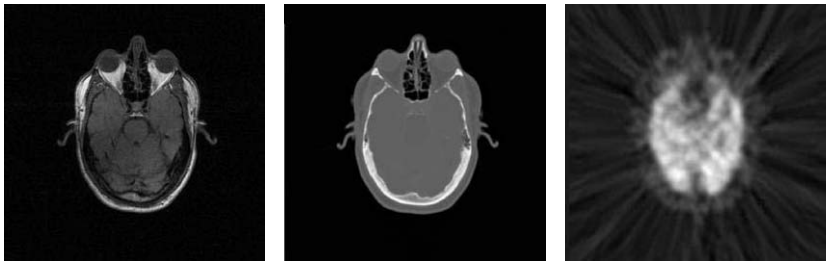

Fig. 1. Used test images: (a) MR, (b) CT, (c) PET.

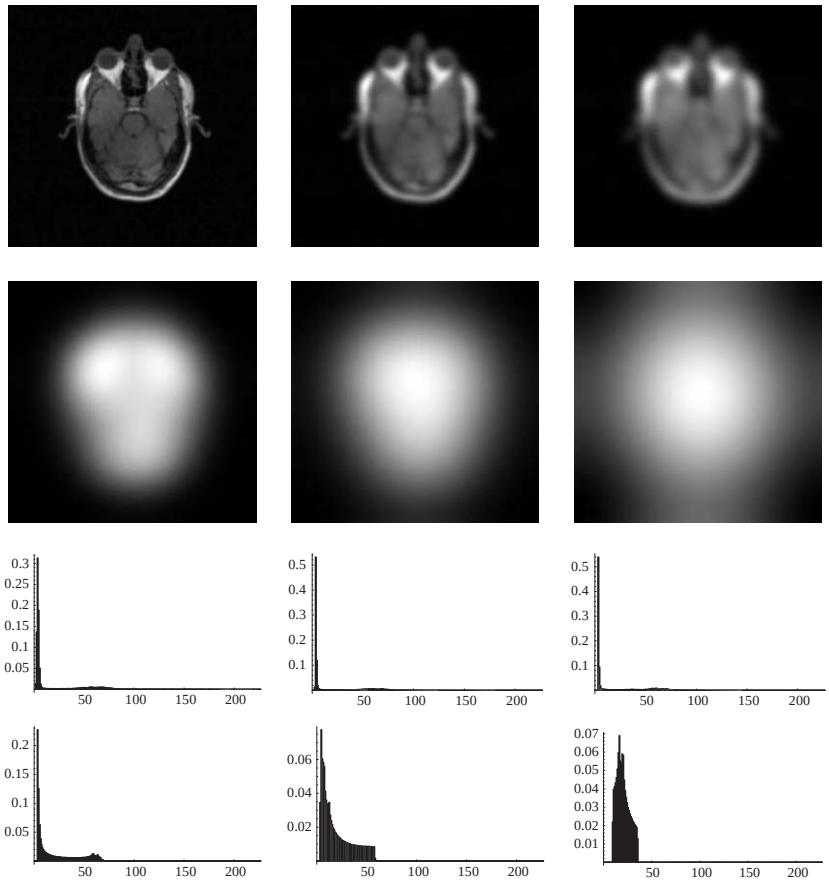

Fig. 2. MR image at scales 1, 3, 5 (row 1), 20, 33, and 60 (row 2). Histograms of the MR image at scales 1, 3, 5 (row 3), 20, 33, and 60 (row 4).

two images only a white blob occurs. The histograms of these images, shown in the bottom row of Fig. 2 clearly visualize this. The first histogram shows a (small) peak at the bins 50–70 (the latter is the last filled bin). The other histograms just present blob information in the binrange 4–59 and 10–36, respectively. Blurring this blob obviously converges to a uniformly grey image (to be expected at a scale of approx. 140, i.e. half of the image size) with only one bin filled and a MI of 0. Note that the NMI is not defined for a single filled bin.

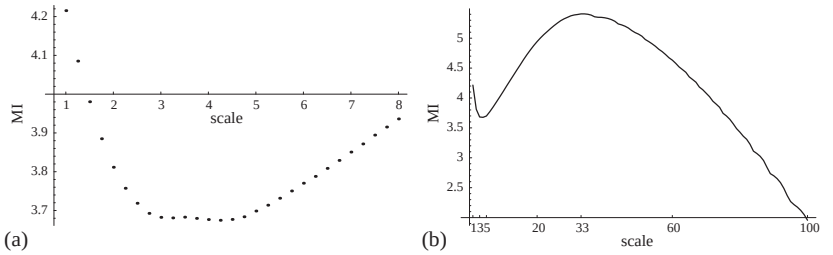

Fig. 3. MI of the MR image. (a) Small scale behaviour: scales $1, 1.25, \ldots, 8$. (b) Large scale behaviour: scales $1, 2, \ldots, 100$.

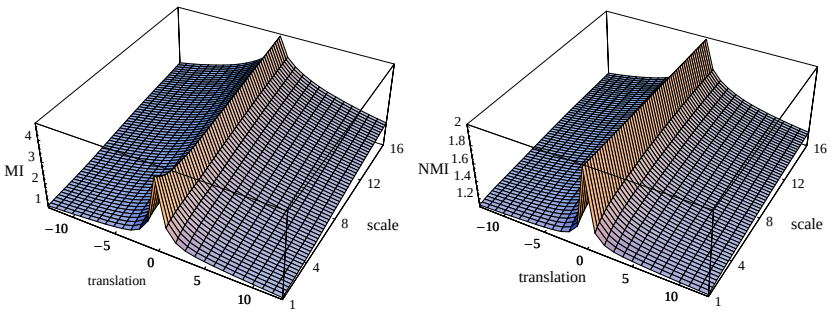

Fig. 4. Registration of the MI (left) and NMI (right) of the MR image and itself translated with n pixels, $n = -13, -12, \ldots, 13$ at scales $1, 1.25, \ldots, 16$.

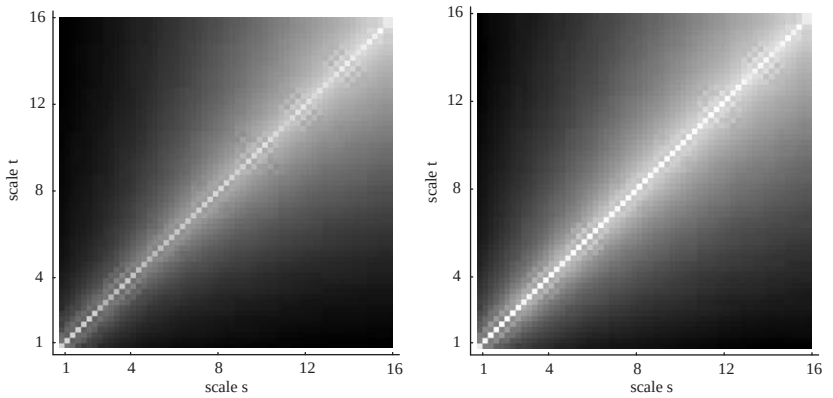

Fig. 5. Registration results of the MR image at scales $s, t = 1, 1.25, \ldots, 16$ for the $\mathrm{MI}(A_s, A_t)$ (left) and $\mathrm{NMI}(A_s, A_t)$ (right).

4.1.1.3. Translation artifact and scale. Next a translation of n pixels, with $n \in [-13, 13]$ an integer, is taken. The MI of the set $\{T_n(A_s), A_s\}$ for $s = 1, 1.25, \ldots, 16$ is shown in Fig. 4a. Obviously, the ridge at the translation $n = 0$ corresponds to the MI shown in Fig. 3. As the scale increases the ridge becomes less pronounced. This is due to the fact that details are blurred away and the number of bins decreases. Two events take place while increasing scale. Firstly the peak occurring at a scale level due to the translation artifact is damped, for small scales even strongly to the local minimum visible in Fig. 3. Secondly the entropy at such a level moves up vertically for all translation artifacts, see e.g. the slope on the boundary of Fig. 4a.

In the NMI, shown in Fig. 4b, the ridge value is a constant, viz. 2, since it corresponds to the image perfectly aligned with itself. The entropy due to misregistration increases. This is explained by the fact that due to blurring both the number of bin decreases and the details of the image (i.e. small peaks) vanish, and the images become more alike since, for example, a mismatch of

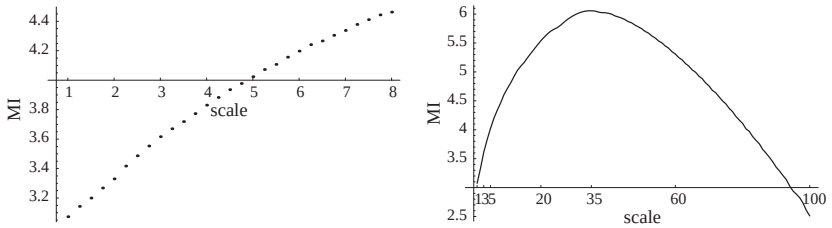

Fig. 6. MI of the CT image. Left: Small scale behaviour: scales $1, 1.25, \ldots, 8$. Right: Large scale behaviour: scales $1, 2, \ldots, 100$.

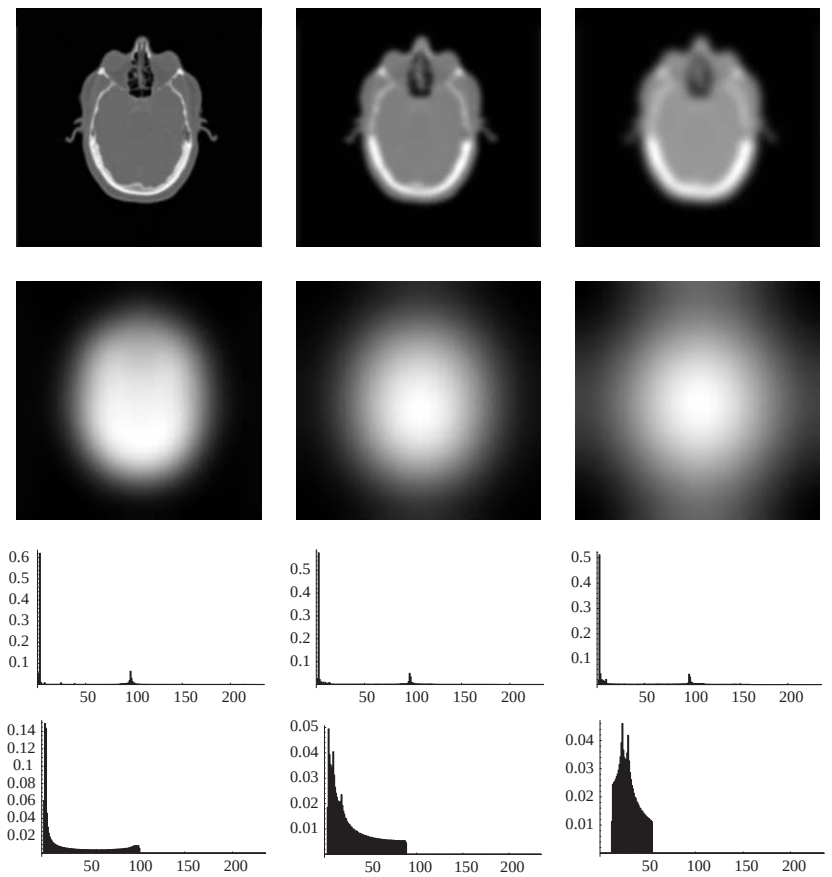

Fig. 7. CT image at scales 1, 3, 5 (row 1), 20, 35, and 60 (row 2). Histograms of the CT image at scales 1, 3, 5 (row 3), 20, 35, and 60 (row 4).

one pixel is relatively small to blurring with, say, 10 pixels.

4.1.1.4. Registration at different scales. For a comparison of the case $\{A_s, A_t\}$, see Fig. 5. It shows the MI (left) and NMI (right) of the image perfectly aligned but compared at different scales. Obviously, the best result is obtained at equal scales, $s = t$, the diagonal of both images. The values on the diagonal equal that of Fig. 3 for the MI image, and 2 for the NMI image.

4.1.2. CT image

The MI of the CT image, see Fig. 1b, does *not* contain a local minimum for small scales, but starts monotonically increasing, see Fig. 6b, until a scale of 35 pixels. For large scales there is only blurring of a white blob and consequently a decrease of MI. One may consider the CT image to be slightly blurred already, compared to the MR image.

This becomes clear from Fig. 7, where the image at various scales and corresponding histograms are shown. The structure is similar to the MR image, albeit that here only

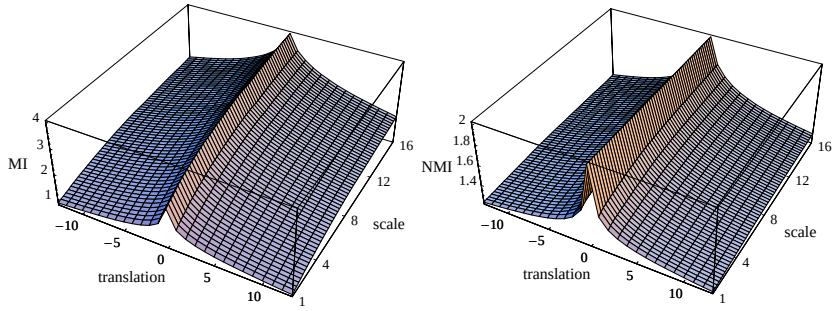

Fig. 8. Registration of the MI (left) and NMI (right) of the CT image and itself translated with n pixels, $n = -13, -12, \ldots, 13$ at scales $1, 1.25, \ldots, 16$.

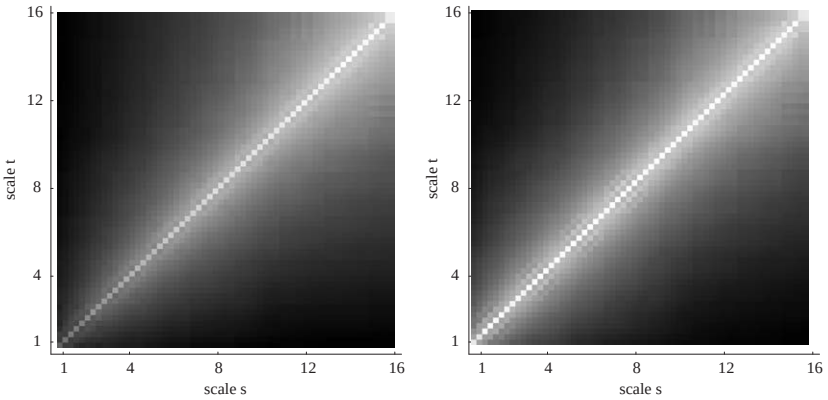

Fig. 9. $\mathrm{MI}(A_s, A_t)$ (left) and $\mathrm{NMI}(A_s, A_t)$ (right) of the CT image at scales $s, t = 1, 1.25, \ldots, 16$.

one large peak taking a fraction of 0.62 is visible for small intensities. The temporal decrease of MI is thus not to be expected.

4.1.2.1. Translation artifact and scale. As a consequence, the image of the MI of the CT–CT registration with pixel wise translation of one image under the influence of small scales, Fig. 8, shows a ridge with all values increasing. The NMI, if not equal to 2, increases too.

4.1.2.2. Registration at different scales. The MI and the NMI of the CT image at different scales, Fig. 9, one finds the strongest responses at identical scales.

4.1.3. PET image

For the PET image, Fig. 1c, similar results as for the MR image are obtained. Fig. 10 shows the behaviour of the MI under blurring. Here also a local minimum occurs due to the deletion of noise—the stripes—in the background, albeit at a relatively large scale, viz. 14 pixels. The MI increases until the local maximum at scale 35 and then decreases again until it converges to 0.

The image and its histograms at several scales are shown in Fig. 11. At scale 1 there is one big peak (fractions of approx. 0.05) spread over roughly 50 bins, with a very large tail containing one small wide peak around 175. During blurring to scale 14, the spread of this big peak is narrowed to 25 bins that contain more data (fractions of 0.04–0.14), while the tail is shortened (the stripes in the PET image disappear). After this local extremum of the MI the large peak firstly remains stable, while the tail keeps moving to the left, due to the disappearance of all structure in the image. After the second local extremum only the white blob is blurred and the histogram converges to a uniformly distributed one containing finally one bin.

4.1.3.1. Translation artifact and scale. The registration of the PET image with itself, translated n pixels, $n = -13, -12, \ldots, 13$, as shown in Fig. 12 shows the same behaviour as the MR image, albeit that the ridge for the MI is less pronounced.

4.1.3.2. Registration at different scales. Comparing the MI and the NMI of the PET image at different scales,

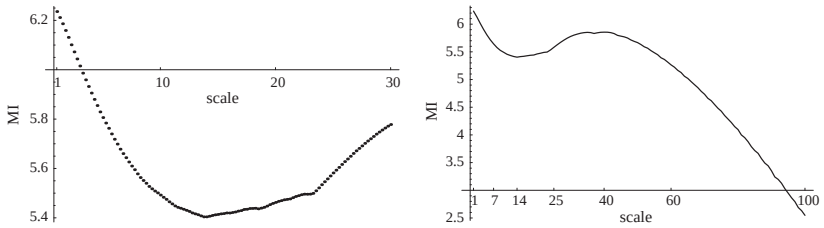

Fig. 10. MI of the PET image. Left: small scale behaviour: scales $1, 1.25, \ldots, 30$. Right: Large scale behaviour: scales $1, 2, \ldots, 100$.

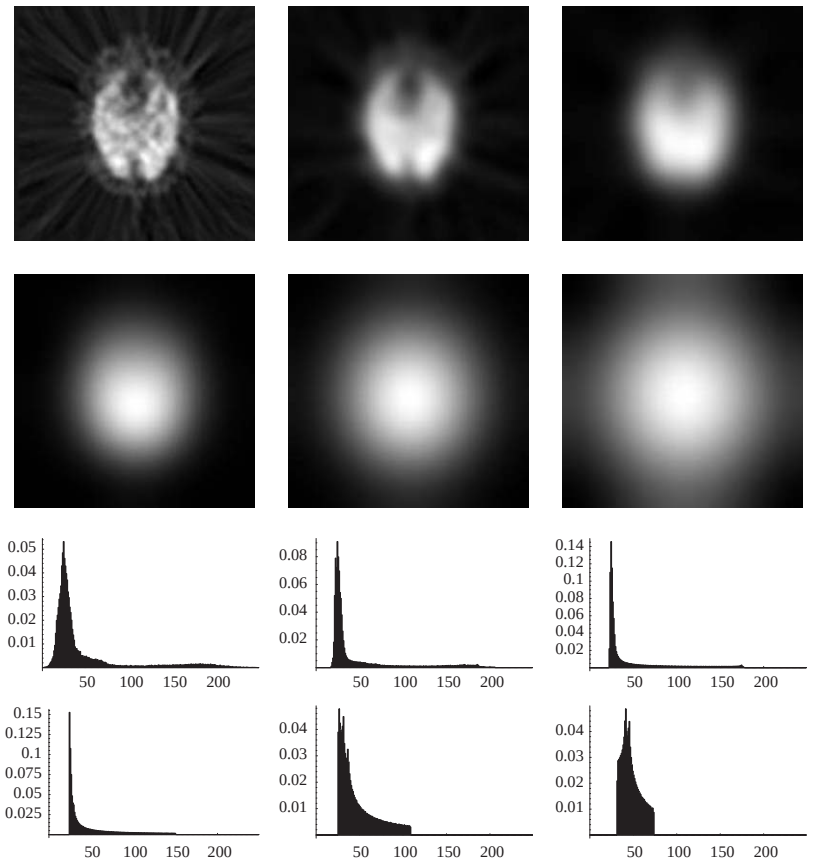

Fig. 11. PET image at scales 1, 7, 14 (row 1), 25, 40, and 60 (row 2). Histograms of the PET image at scales 1, 7, 14 (row 3), 25, 40, and 60 (row 4).

Fig. 13, one finds the strongest responses at identical scales.

4.2. Different modality, different scale registration

In this section, the bimodal results $(N)\mathrm{MI}(A_s, B_t)$ for the pairs MR–CT, MR–PET, and CT–PET, respectively, are presented. The images are pair-wise aligned at the initial scales and the behaviour of the (normalized) mutual information measures is investigated as a function of the two (different) scale parameters s and t. Since the behaviour of the entropy of each modality has its specific characteristics, it is to be expected that the bimodal outcome is also non-trivial. So, for example, a 'diagonal-like' optimal response suggesting $s = t$ as the best combination, is not likely.

4.2.1. MR–CT registration

In the MR–CT registration the registration at identical scales (the $s = t$ case) indeed does not give the best response, as shown in Fig. 14. The highest response is not on the

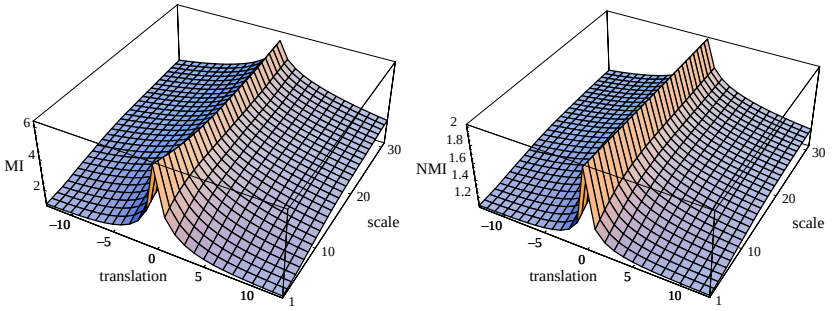

Fig. 12. Registration of the MI (left) and NMI (right) of the PET image and itself translated with n pixels, $n = -13, -12, \ldots, 13$ at scales $1, 2, \ldots, 30$.

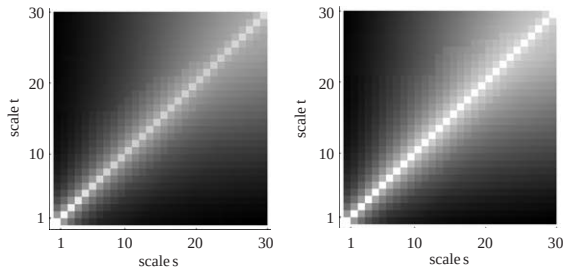

Fig. 13. MI(A_s, A_t) (left) and NMI(A_s, A_t) (right) of the PET image at scales $s, t = 1, 2, \ldots, 30$.

diagonal (i.e. similar scales), but on a line with slope that is somewhat higher than 1: a ridge is present along the scale combinations (s, t)=(2.5, 1) to (s, t)=(9, 10), where the CT image is scaled by s and the MR image by t. Basically, this image is a build-up by the combination of two functions with extrema that move with different velocities under blurring. However, since they are out of phase, the maximum of the combination—the top of the ridge—starts "out of phase" and moves while increasing scales.

4.2.2. MR–PET registration

The MR–PET registration, where the PET image is scaled by s and the MR image by t, is shown in Fig. 15. The scales are taken from 1 to 30 with stepsize 1 pixel. Two ridges are present, approximately along the scale combinations (s, t)=(1, 4) to (s, t)=(30, 30) (slightly curved), and (s, t)=(5, 15) to $(s, t) = (30, 30)$. They meet and merge at the highest scales. Given the behaviour of the entropies of both images this is to be expected. In contrast to the previous section, however, now the absolute maximum is found to be the combination of the highest scales.

4.2.3. CT–PET registration

The CT–PET registration, where the PET image is scaled by s and the CT image by t, is shown in Fig. 16. The scales are taken from 1 to 30 with stepsize 1 pixel. Again, two ridges are present, albeit not as clear as in the previous case. They are located along the scale combinations (s, t)=(10, 5) to $(s, t) = (30, 30)$ (slightly curved), and $(s, t) = (10, 15)$ to $(s, t) = (30, 30)$. They meet and merge at the combination of the highest scales.

A comparison of the output of the two measures shows that the NMI and the MI generate the same structures.

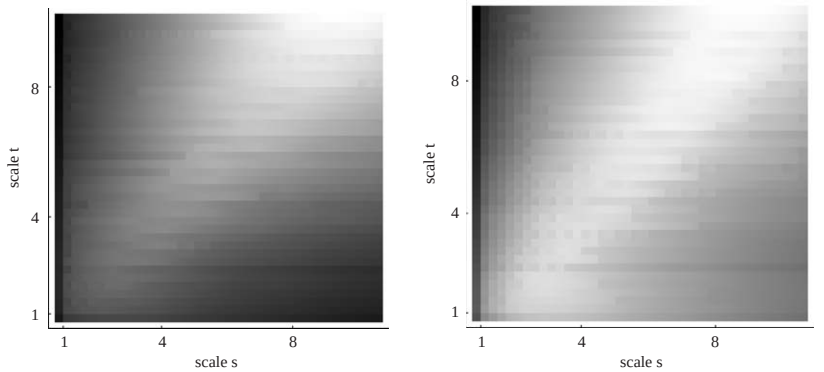

Fig. 14. MR(t)–CT(s) at scales $s, t = 1, 1.25, \ldots, 10$. Left: MI. Right: NMI.

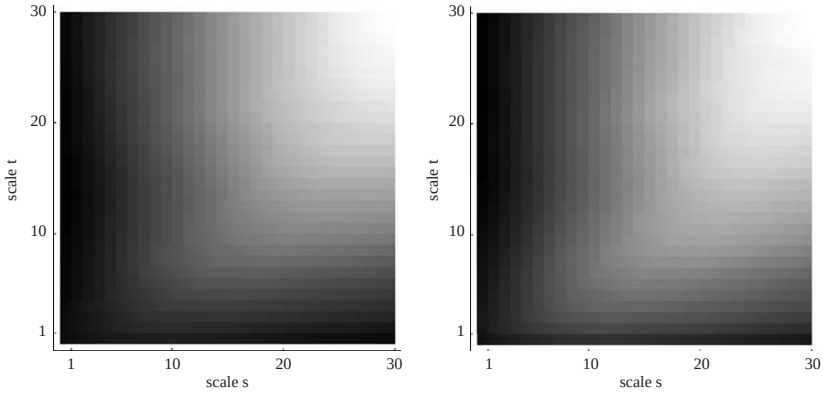

Fig. 15. MR(t)–PET(s) at scales $1, 2, \ldots, 30$. Left: MI. Right: NMI.

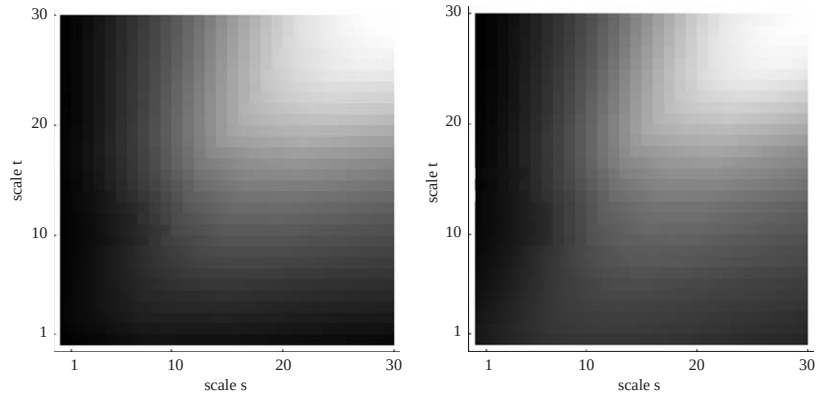

Fig. 16. PET(s)–CT(t) at scales $1, 2, \ldots, 30$. Left: MI. Right: NMI.

5. Summary

In this paper, the investigation of the MI measure applied to scale space images is considered. In this way local image structure can be embedded in the measure, can the registration be speed-up, and is the discrete image transferred into a continuous setting. The focus of the research is on the properties of this combination, since convergence of the algorithm performing the registration and practical implications in this multiscale context have been reported elsewhere [17–19]. In this paper, the combination of the methods itself is under review.

Applying mutual information to an image in scale space (multiresolution) context is not as straightforward as it would seem to be. The entropy of an image under the influence of blurring changes non-monotonically. This can be understood by examining the change of the histogram. The MR image shows extra complicated behaviour due to the temporally merging of bins for low values. This is caused by the smoothing of the background A resampling of the histogram, e.g. by adjusting the binwidth with scale, does not remove this behaviour. The registration of an image with itself under various scales, however, yields completely the desired result. The strongest response is obtained by the correct alignment and correspondence of scale.

If images of different modality are registered at various scales, complicated behaviour occurs. This is due to the non-monotonic behaviour of the separate entropies. As a consequence, a maximal response of the MI and NMI is generally obtained by different scales for the images. With the images used in mind, one can argue indeed that a highly detailed MR image may be smoothed a little bit when comparing it to a smoother CT image. For the PET image similar arguments hold. Here the strips influence the registration, so removing them by taking the image at some larger scale may improve the registration. Obviously, the maximal response of the (N)MI measure is not the selection criterion: Then

extreme high scales should be selected, which blur everything away. On the other hand, given a certain scale for one image, a scale with an optimal response for the other image can be selected.

It needs to be investigated whether this property can be used for better or faster registration of two images. Since blurred images can be registered faster [17], one has a good starting point for aligning the images at the initial scales and only need to refine the result.

As stated in Section 3, the influence of linear scale space, i.e. linear Gaussian blurring, is twofold. Removing the noise is an advantage and gives a more pronounced histogram, but the blurring of edges blurs the histogram. The latter can be avoided by using a nonlinear scale space [11]. A disadvantage is that in these scale spaces the relation between scale and spatial location is less obvious.

Acknowledgements

This work was partially supported by Senter/IOP, innovation-driven research programmes, of the Department of Economic Affairs, the Netherlands (IBV99006) and by the Deep Structure, Singularities, and Computer Vision (DSSCV) project, an IST Programme of the European Union (IST-2001-35443).

References

[1] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, Multimodality image registration by maximization of mutual information, IEEE Trans. Med. Imaging 16 (2) (1997) 187–198.

[2] F. Maes, D. Vandermeulen, P. Suetens, Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information, Med. Image Anal. 3 (4) (1999) 373–386.

[3] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, Mutual-information-based registration of medical images: a survey, IEEE Trans. Med. Imaging 22 (8) (2003) 986–1004.

[4] C. Studholme, D. Hill, D. Hawkes, Automated 3D registration of MR and CT images of the head, Med. Image Anal. 1 (2) (1996) 163–175.

[5] P. Viola, W. Wells III, Alignment by maximization of mutual information, Int. J. Comput. Vision 24 (2) (1997) 137–154.

[6] W. Wells III, P. Viola, H. Atsumi, S. Nakajima, R. Kikinis, Multi-modal volume registration by maximization of mutual information, Med. Image Anal. 1 (1) (1996) 35–51.

[7] L.M.J. Florack, Image structure, Computational Imaging and Vision Series, Vol. 10, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.

[8] B.M.t. Haar Romeny, L.M.J. Florack, J.J. Koenderink, M.A. Viergever (Eds.), Scale-Space Theory in Computer Vision: Proceedings of the First International Conference, Scale-Space'97, Utrecht, The Netherlands, Lecture Notes in Computer Science, Vol. 1252, Springer, Berlin, 1997.

[9] J.J. Koenderink, The structure of images, Biol. Cybernet. 50 (1984) 363–370.

[10] in: M. Nielsen, P. Johansen, O.F. Olsen, J. Weickert (Eds.), Scale-Space Theories in Computer Vision, Lecture Notes in Computer Science, Vol. 1682, Springer, Berlin, Heidelberg, 1999.

[11] J.A. Weickert, Anisotropic Diffusion in Image Processing, Teubner, Stuttgart, 1998.

[12] A. Kuijper, L.M.J. Florack, M.A. Viergever, Scale space hierarchy, J. Math. Imaging Vision 18 (2) (2003) 169–189.

[13] A. Kuijper, L.M.J. Florack, The hierarchical structure of images, IEEE Trans. Image Process. 12 (9) (2003) 1067–1079.

[14] J. Sporring, The entropy of scale-space, ICPR, Vol. A, 1996, pp. 900–904.

[15] J. Sporring, J. Weickert, On generalized entropies and scale-space, in: Ter Haar Romeny et al. [B.M.t. Haar Romeny, L.M.J. Florack, J.J. Koenderink, M.A. Viergever (Eds.), Scale-Space Theory in Computer Vision: Proceedings of the First International Conference, Scale-Space'97, Utrecht, The Netherlands, Lecture Notes in Computer Science, Vol. 1252, Springer, Berlin, 1997], 1997, pp. 53–64.

[16] A. Winter, H. Maitre, N. Cambou, E. Legrand, Entropy and multiscale analysis: a new feature extraction algorithm for aerial images, in: IEEE ICASSP97, Vol. 4, Munich,April 1997, pp. 2765–2768.

[17] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, A multiscale approach to mutual information matching, in: K. Hanson (Ed.), Medical Imaging: Image Processing, Proc. SPIE, Vol. 3338, SPIE Press, Bellingham, WA, 1998, pp. 1334–1344.

[18] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, Mutual information matching in multiresolution contexts, Image Vision Comput. 19 (1) (2001) 45–52.

[19] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, Image registration by maximization of combined mutual information and gradient information, IEEE Trans. Med. Imaging 19 (8) (2000) 809–814.

[20] C. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423 and 623–656, reprinted with correction.

[21] C. Studholme, D. Hill, D. Hawkes, An overlap invariant entropy measurement of 3D medical image alignment, Pattern Recognition 32 (2) (1999) 71–86.

[22] L. Schwartz, Théorie des Distributions, Vol. I, II of Actualités scientifiques et industrielles, Publications de l'Institut de Mathématique de l'Université de Strasbourg, Paris, 1950–1951, pp. 1091, 1122.

[23] T. Lindeberg, Scale-Space Theory in Computer Vision, The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Dordrecht, 1994.

[24] B.v. Ginneken, B.M.t. Haar Romeny, Applications of locally orderless images, in: Nielsen et al. [M. Nielsen, P. Johansen, O.F. Olsen, J. Weickert (Eds.), Scale-Space Theories in Computer Vision, Lecture Notes in Computer Science, Vol. 1682, Springer, Berlin, Heidelberg, 1999], 1999, pp. 10–21.

[25] L.D. Griffin, Scale-imprecision space, Image Vision Comput. 15 (1997) 369–398.

[26] J.J. Koenderink, A.J. Van Doorn, Blur and disorder, in: Nielsen et al. [M. Nielsen, P. Johansen, O.F. Olsen, J. Weickert (Eds.), Scale-Space Theories in Computer Vision, Lecture Notes in Computer Science, Vol. 1682, Springer, Berlin, Heidelberg, 1999], 1999, pp. 1–9.

About the Author—ARJAN KUIJPER received his M.Sc. degree in applied mathematics in 1995 with a thesis on the comparison of two image restoration techniques, from the University of Twente, The Netherlands. During the period 1996–1997 he worked as programmer and consultant at ELTRA Parkeergroep, Ede, The Netherlands. In the period 1997–2002 he has been a Ph.D. student and associate researcher at the Institute of Information and Computing Sciences of Utrecht University. In 2002 he received his Ph.D. degree with a thesis on "Deep Structure of Gaussian Scale Space Images" and worked as postdoc at Utrecht University on the project "Co-registration of 3D Images". Since January 1st 2003 he has been working as an assistant research professor at the IT University of Copenhagen in Denmark funded by the IST Programme "Deep Structure, Singularities, and Computer Vision (DSSCV)" of the European Union. His interest subtends all mathematical aspects of image and shape analysis, notably multiscale representations (scale spaces), catastrophe and singularity theory, medial axes and symmetry sets, and applications to medical imaging.