

Mutual Information Aspects of Scale Space Images

Arjan Kuijper

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. In this paper a combination of both methods is investigated. This results in a parametrized mutual information measure using local information of the image. For single modality matching a critical scale is found on which the parametrized mutual information has an extremum. First results on multi-modality matching show that different scales for the images yield the best response of the parametrized mutual information.

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 camera's 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 between the two images, say A and B .

This correspondence is expressed as a transformation T that applied to A , yields image B . If T contains one rotation and translation that holds 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 measure, using the entropies of the images and their joined entropy, has proven to perform well [8, 9, 16, 17, 19]. 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. 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 [1, 4, 5, 10, 18]. 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 [7].

Since both methods start of by taking minimal assumptions and it makes sense to investigate the combination of them. In this paper the usage of scale space in mutual

information image registration is investigated. This essentially boils down to applying the mutual information measure to the images at various scales. Since images at increasing scales become more and more blurred, combining the methods is a way to embed local information of the image in the mutual information measure.

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 [20]. Pluim et al. [11] 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.

The MI measure, scale space and their combination are described in section 2. In section 3 the results of this approach are presented on MR, CT, and PET images, as well as their combinations. In section 4 results are summarised and discussed.

2 Theory

Next I will briefly describe the ideas behind and theory of the concepts of mutual entropy in section 2.1 and scale space in section 2.2. The combination, the application of MI to scale space images, will be described in section 2.3

2.1 Mutual Information

The Mutual Information measure [8, 16] is based on the shared information of the overlapping part of two images. This information is obtained using the Shannon entropy [13], known as a measure of uncertainty. Let A be a random variable and $p_A(a)$ its marginal probability distribution, then the entropy $H(A)$ is given by

$$H(A) = \sum_{a \in A} p_A(a) \log \frac{1}{p_A(a)} = - \sum_{a \in A} p_A(a) \log p_A(a).$$

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

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

The Mutual Information (MI) is defined as

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

Expanding the division and multiplication in the logarithm and the sum over the terms yield

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

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

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

and thus

$$MI = -H(A, B) + H(A) + H(B) \quad (1)$$

If A and B are completely uncorrelated, $H(A, B) = H(A) + H(B)$ and therefore $MI = 0$. In the opposite case, if A and B are fully correlated, i.e. $A = B$, $H(A, B) = H(A) = H(B)$ and $MI = H(A)$. Since this is a priori unknown value, in many tasks the Normalised Mutual Information (NMI) is used. It is defined by

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

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

For images the marginal probability functions are derived from the histogram of the intensities in the image. The histogram contains n bins b_k containing the relative number (#) of intensities of A , $I(A)$, between (generally equidistant) boundary values L_{k-1} and L_k :

$$b_k = \#\{I(A) \mid L_{k-1} < I(A) \leq L_k\} \quad (3)$$

Obviously, $\sum_{k=1}^n b_k = 1$. Similarly, a histogram for B with boundary values N_m is obtained. The joint marginal probability is represented by a 2D histogram containing bins $b_{k,m}$. In these bins the relative number of corresponding image points (pixels, voxels, etc.) are stored. These image points have the same spatial location (after translation and if necessary interpolation) and are stored in the bins b_k and b_m of the histograms of A and B , respectively.

$$b_{k,m} = \#\{\{I(A) \mid L_{k-1} < I(A) \leq L_k\} \cap \{I(B) \mid N_{m-1} < I(B) \leq N_m\}\} \quad (4)$$

As a consequence, if $A = B$, and both histograms have equal binwidth ($L_k - L_{k-1} = N_m - N_{m-1}$ for all (k, m)), then $b_{k,m} = 0$ if $k \neq m$ and the 2D histogram has only non-zero entries on the diagonal.

2.2 Scale Space

The important addition of a scale space image to an ordinary image is the notion of scale. Its appearance lies in the fact that all images are discrete. Generally, there is no a priori reason to treat the image on pixel scale, since they are only a result of the acquisition of the image by an aperture with a certain scale or sampling range. Moreover, the assumption that the image can be regarded as some kind of more or less continuous function of its spatial variables is mathematically justified, only if this discrete set is convolved with a so-called test function (Theory of Distributions [12]).

One of the most simplest and most used test functions is the Gaussian filter with zero mean and variable variance [1]. The latter is very important: since there is no a priori reason to fix the variance, all possible variances are taken. Therefore the image $L(\mathbf{x})$ is extended to a scale space image $L(\mathbf{x}; t)$ by

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

From this formulation it is clear that when regarding scale σ , $\sigma = \sqrt{2t}$, scale and spatial location have the same dimensionality. It thus makes sense to speak of ‘a scale of two pixels’. This corresponds to regard the image blurred with $t = 2$.

2.3 Mutual Information in Scale Space

The basic idea of mutual information 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 partly determined by the intensity of that point in the original image, and partly by the 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 smoothed. The result on the histogram is twofold. It also blurs, albeit not necessarily Gaussian, due to smoothing of edges. On the other hand, the suppression of noise yields ideally removal of irrelevant (noise induced) intensities, i.e. sharpening, or at least more pronouncing, of relevant (non-noisy) clusters.

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

In 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. [2, 3, 6]. 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 mutual information becomes a parametrized 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,

$$MI(A_\sigma, B_\tau) = -H(A_\sigma, B_\tau) + H(A_\sigma) + H(B_\tau)$$

and

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

Now two different routes to investigate the behaviour of these parametrized 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.



Figure 1: Used test images: a) MR b) CT c) PET

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 these parametrized measures are investigated. Firstly, the case $A = B$ is taken. This gives insight in the behaviour of the mutual information under the influence of scale. Then $MI(A_\sigma, A_\tau) = -H(A_\sigma, A_\tau) + H(A_\sigma) + H(A_\tau)$. The special case $\sigma = \tau$, yielding a one parameter measure $MI = H(A_\sigma)$ is also dealt with. It describes the behaviour of the entropy under blurring. Secondly, the general case $A \neq B$, is investigated.

3 Results

In this section I consider the test images as shown in Figure 1. They are, from left to right, Magnetic Resonance (MR), Computed Tomography (CT), and Positron Emission Tomography (PET) images. These images are already aligned and have integer values ranging from 0 to 255. Consequently, histograms are taken with bandwidth 1. In subsequent sections firstly the single modality behaviour MR-MR, CT-CT, and PET-PET is investigated.

In all cases 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 $\{T(A_\sigma), A_\sigma\}$). 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)).

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)).

3.1 MR-MR registration

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

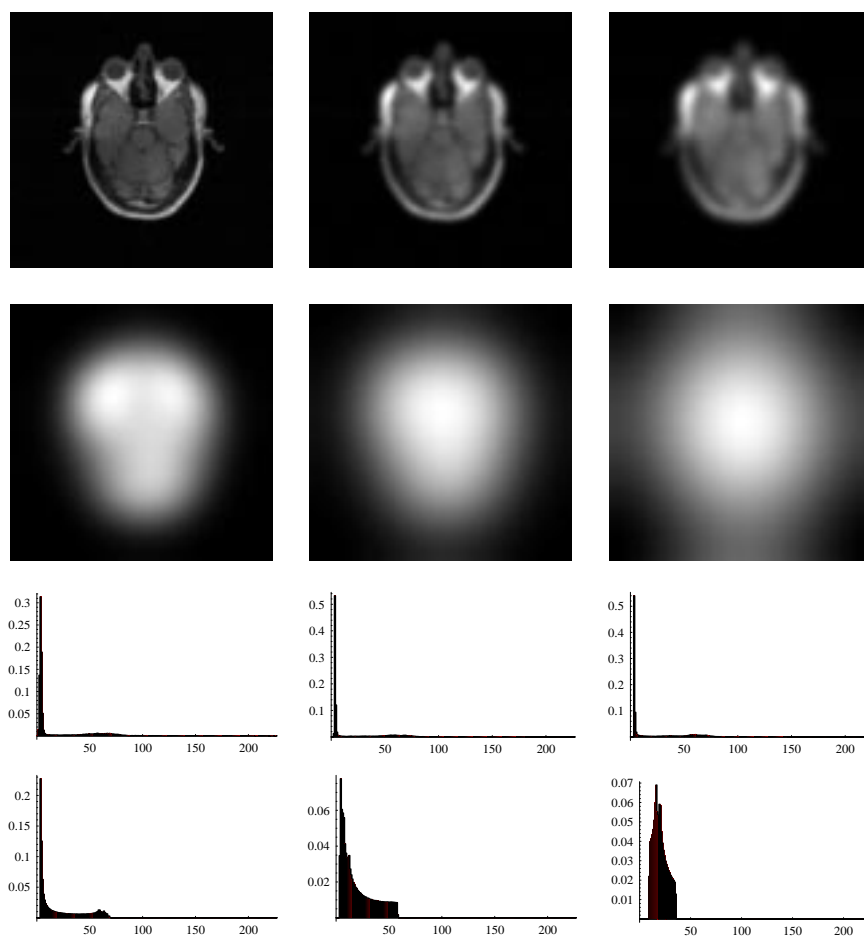


Figure 2: MR image at scales 1, 3, 5 (row 1), 20, 33, and 60 (row 2). Histograms of the MR image at scales 1, 3, 5 (row 3), 20, 33, and 60 (row 4).

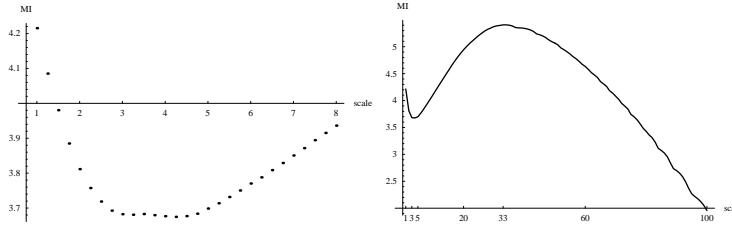


Figure 3: MI of the MR image A) Small scale behaviour: scales 1, 1.25, \dots , 8. B) Large scale behaviour: scales 1, 2, \dots , 100.

Firstly a scale space containing small scales is analysed. 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 Figure 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 is shown in the third row of Figure 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 reorganised 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.

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 Figure 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 Figure 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 two images only a white blob occurs. The histograms of these images, shown in the bottom row of Figure 2 clearly visualise 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.

Next a translation of n pixels, with $n \in [-13, 13]$ an integer, is taken. The MI of the set $\{T_n(A_\sigma), A_\sigma\}$ for $\sigma = 1, 1.25, \dots, 16$ is shown in Figure 4a. Obviously, the ridge at the translation $n = 0$ corresponds to the MI shown in Figure 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 Figure 3. Secondly the entropy at such a level moves up vertically for all translation artifacts, see e.g. the slope on the boundary of Figure 4a.

In the NMI, shown in Figure 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 thus details of the image (i.e. small peaks) vanish, and the images become more alike since, for example, a mismatch of one pixel is relatively small to blurring with, say, 10 pixels.

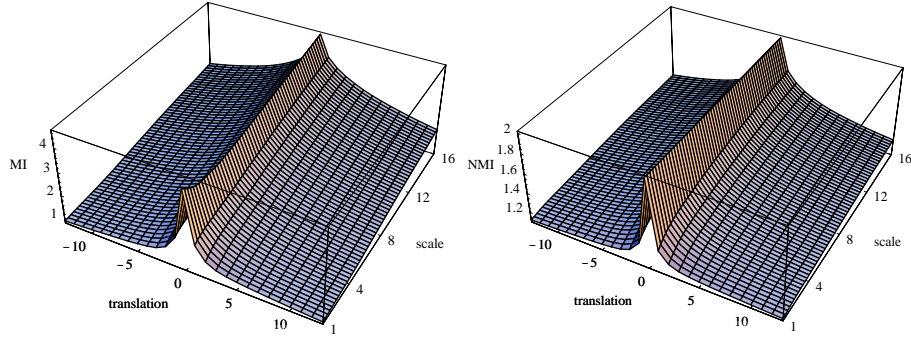


Figure 4: Registration at scales $1, 1.25, \dots, 16$ of the MR image and itself translated with n pixels, $n = -13, -12, \dots, 13$ for a) MI. b) NMI.

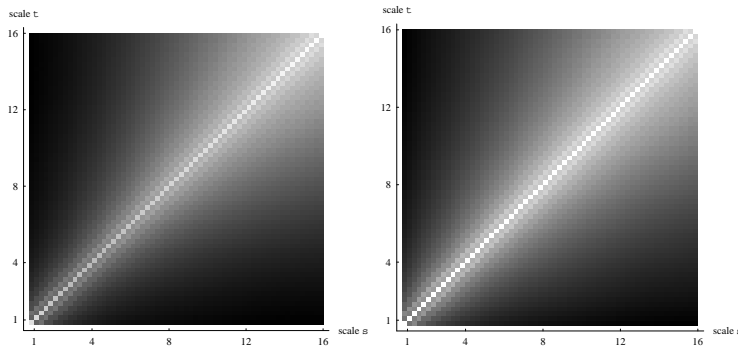


Figure 5: MI of MR-MR at scales $1, 1.25, \dots, 16$. NMI.

For an comparison of the case $\{A_\sigma, A_\tau\}$, see Figure 5. It shows the MI (left) and NMI (right) of the image perfectly aligned but compared at different scales. Obviously, best result is obtained at equal scales, $\sigma = \tau$, the diagonal of both images. The values on the diagonal equal that of Figure 3 for the MI image, and 2 for the NMI image.

3.2 CT-CT

The CT image, see Figure 1b, behaves more or less the same, the main difference is that the MI for the CT image does *not* contain a local minimum for small scales, but starts monotonically increasing, see Figure 6, 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 Figure 7, where the image at various scales and corresponding histograms are shown. The structure is similar to the MR image, albeit that here only one large peak taking a fraction of 0.62 is visible for small intensities. The temporal decrease of MI is thus no to be expected.

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, Figure 8, shows a ridge with all values increasing. The NMI, if not equal to 2, increases too.

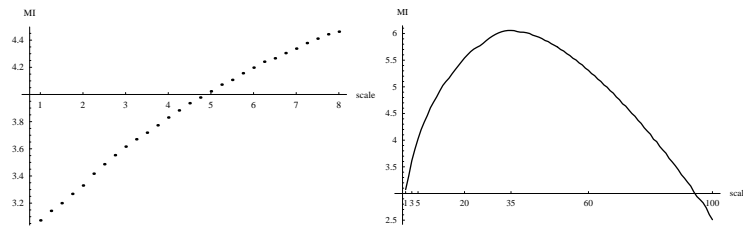


Figure 6: MI of the CT image. Left: Small scale behaviour: scales 1, 1.25, \dots , 8. Right: Large scale behaviour: scales 1, 2, \dots , 100.

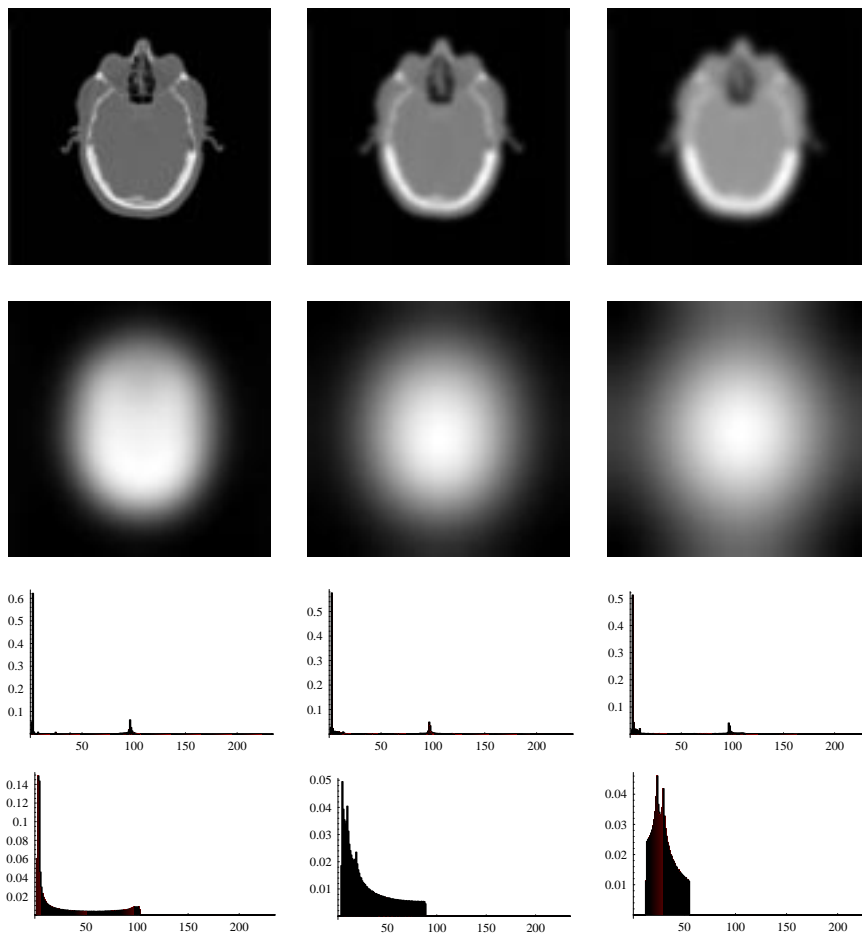


Figure 7: CT image at scales 1, 3, 5, 20, 35, and 60. Histograms of the CT image at scales 1, 3, 5, 20, 35, and 60.

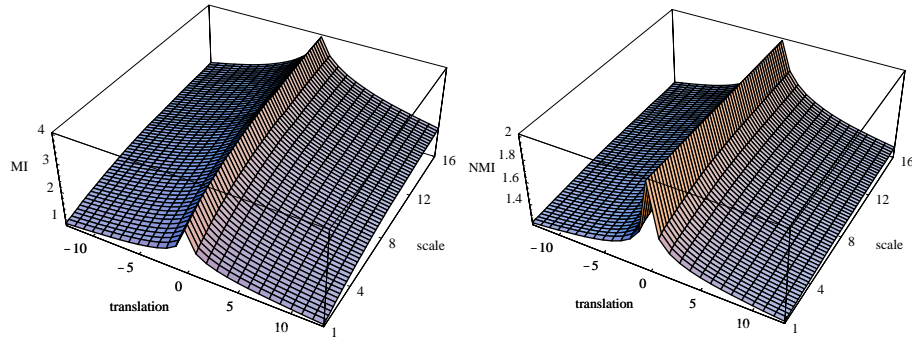


Figure 8: Registration at scales $1, 1.25, \dots, 16$ of the CT image and itself translated with n pixels, $n = -13, -12, \dots, 13$ for a) MI. b) NMI.

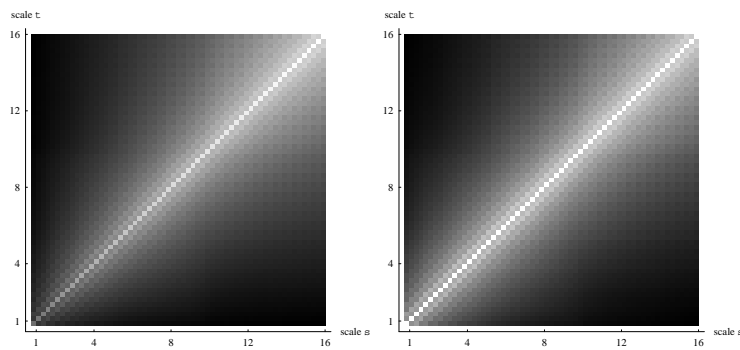


Figure 9: MI of CT-CT at scales $1, 1.25, \dots, 16$. NMI.

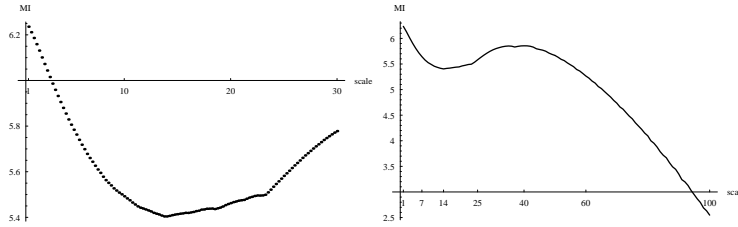


Figure 10: MI of the PET image. Left: small scale behaviour: scales 1, 1.25, \dots , 30. Right: Large scale behaviour: scales 1, 2, \dots , 100.

Comparing the MI and the NMI of the CT image at different scales, Figure 9, one finds the strongest responses at identical scales.

3.3 PET-PET

For the PET image, Figure 1c, similar results as for the MR image are obtained. Figure 10 shows the behaviour of the MI under blurring. Here also a local minimum occurs, 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 Figure 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 to 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.

The registration of the PET image with itself, translated n pixels, $n = -13, -12, \dots, 13$, as shown in Figure 12 shows the same behaviour as the MR image, albeit that the ridge for the MI is less pronounced.

Comparing the MI and the NMI of the PET image at different scales, Figure 13, one finds the strongest responses at identical scales.

3.4 MR-CT

In the CT - MR registration it may seem by intuition that registration at identical scales (the $\sigma = \tau$ case) gives the best result. However, since the MR image shows a non-monotonic behaviour for the MI for small scales, one may expect non-trivial results. Figure 14 supports this hypothesis. It shows the MI (left) and NMI (right) of the MR at scales $s = 1, 1.25, \dots, 10$ and the CT at scales $t = 1, 1.25, \dots, 10$. The second row shows the 3D image. As one can see, the highest response is not on the diagonal with similar scales, but roughly a line with a somewhat higher slope.

Basically this image is build-up by the combination of two functions with a maximum 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. For the CT at scale 1 pixel, the highest response is given by the MR at scale 2.5, moving to CT at 10 pixels vs. MR at 9 pixels.

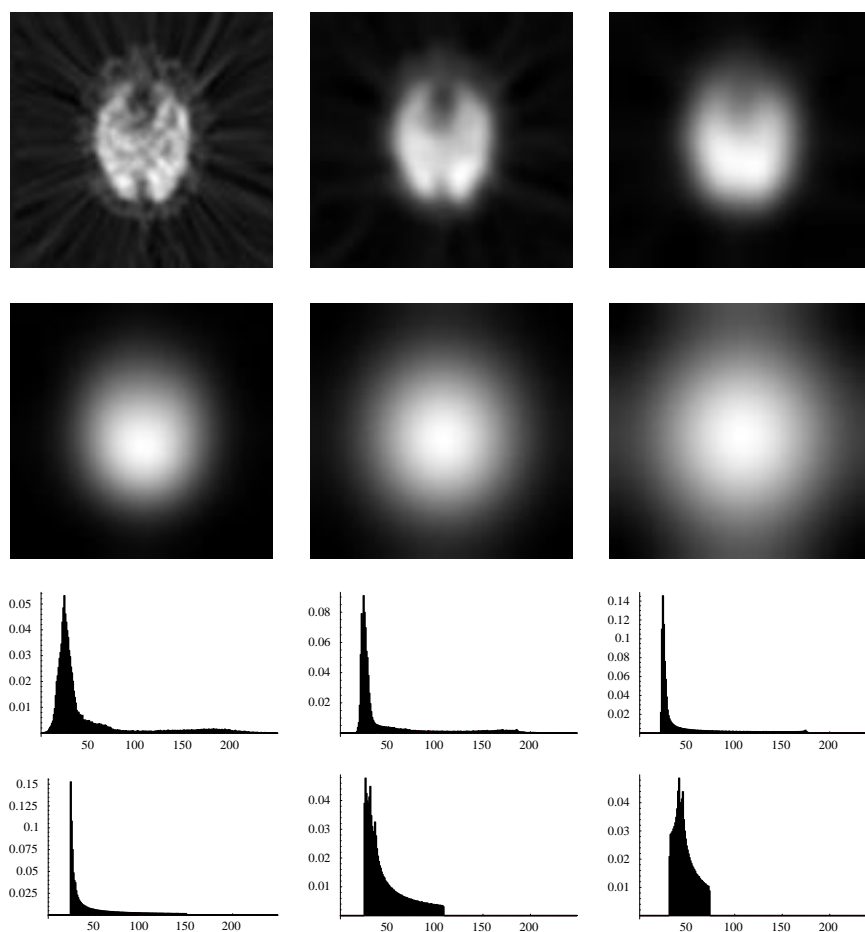


Figure 11: PET image at scales 1,7,14,25,40,60. Histograms of PET image at scales 1,7,14,25,40,60.

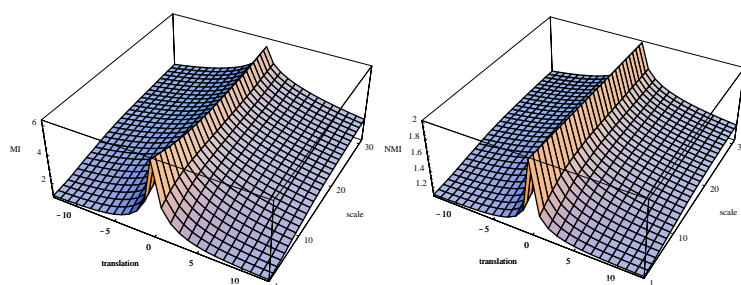


Figure 12: Registration at scales 1, 2, ..., 30 of the PET image and itself translated with n pixels, $n = -13, -12, \dots, 13$ for a) MI. b) NMI.

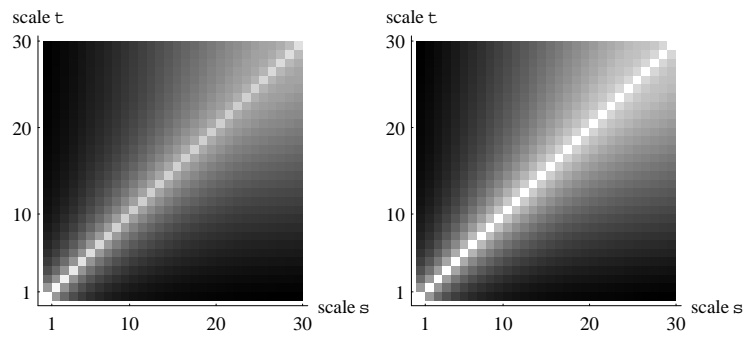


Figure 13: PET-PET at scales $1, 2, \dots, 30$. Left: MI. Right: NMI.

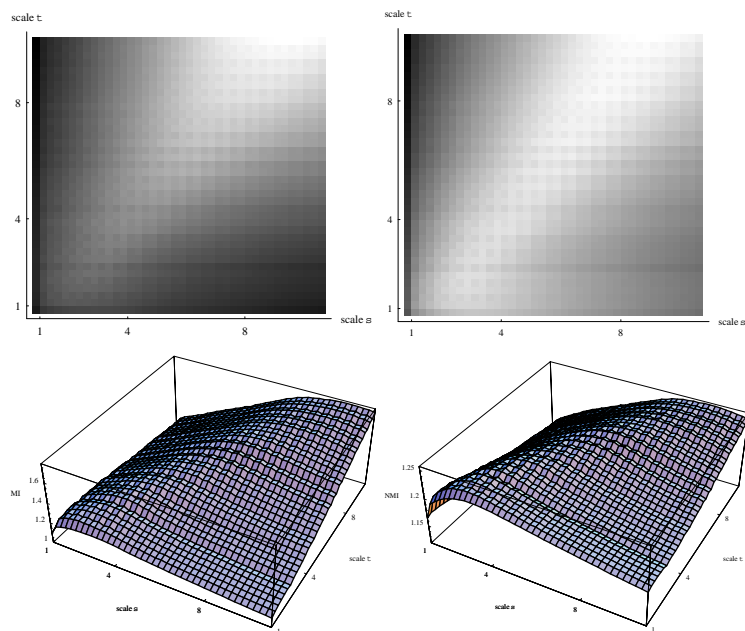


Figure 14: MR-CT at scales $1, 1.25, \dots, 10$. Left: MI. Right: NMI.

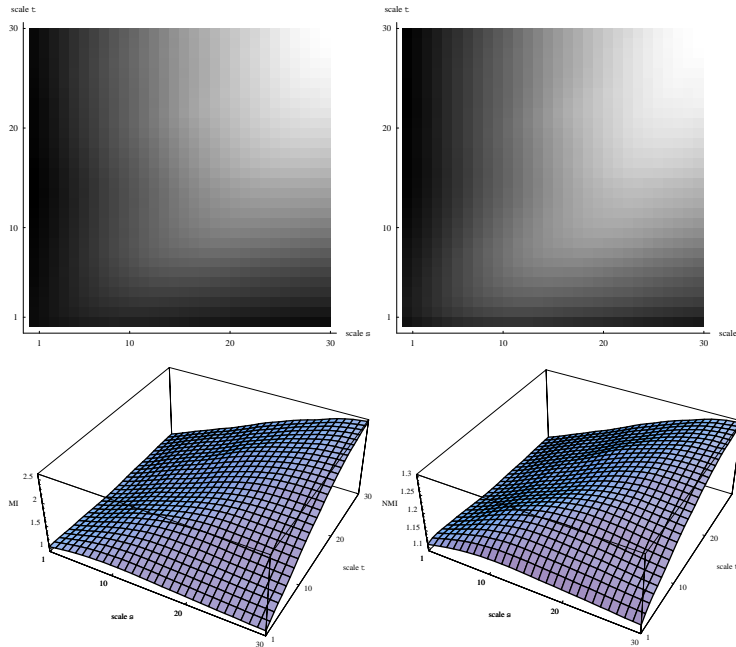


Figure 15: MI of MR-PET at scales $1, 2, \dots, 30$. NMI.

3.5 MR-PET

For the MR - PET registration also the 2D case $\sigma \neq \tau$ is considered, since $\sigma = \tau$ boils down to investigating the diagonal of this 2D image. Scales σ (PET) and τ (MR) are taken from 1 to 30 with stepsize 1 pixel. In Figure 15 clearly two ridges can be seen, one for each parameter. 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 the combination of the highest scales.

3.6 CT-PET

The CT-PET registration is similarly taken for scales σ (PET) and τ (CT) from 1 to 30 with stepsize 1 pixel. The 2D images of the case $\sigma \neq \tau$ are shown in Figure 15. Again, they show two ridges, one for each parameter. They meet and merge at the highest scales.

4 Discussion

The investigation of the mutual information measure in scale space images 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 number of bins in 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

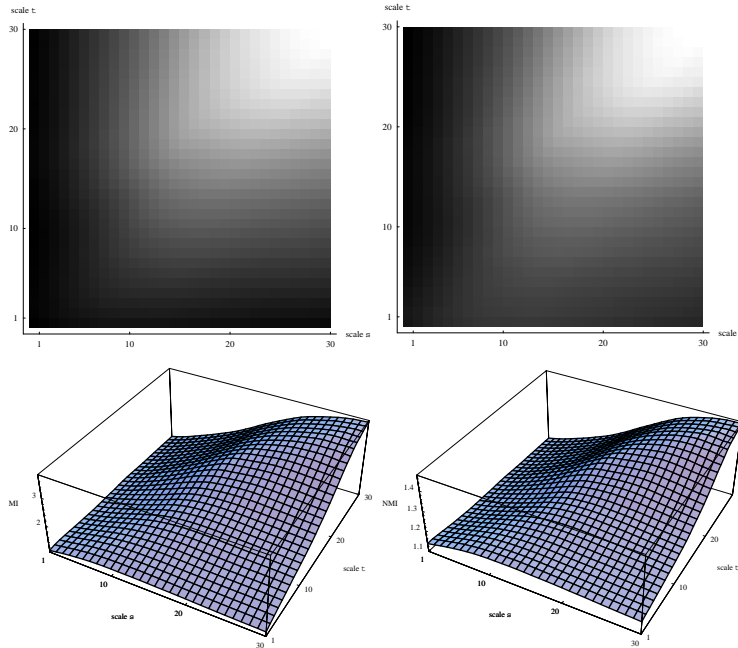


Figure 16: MI of PET-CT at scales 1, 2, . . . , 30. NMI.

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 occur. 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 two 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 more smooth CT image. For the PET image similar arguments hold. Here the axial strips influence the registration, so removing them by taking the image at some larger scale may improve the registration.

Further research may investigate if this property can be used for better or faster registration of two images. Since blurred images can be registered faster [11], one can have a good starting point for aligning the images at the initial scales and only need to refine the result.

It also needs to be investigated in howfar scale influences the position of the maximum response of the MI and NMI. Preliminary experiments show a small deviation which is, however, smaller than the scales used. Since blurring introduces an uncertainty in location, this is also to be expected. Still, further research is recommended.

As stated in section 2.3, 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 non-linear scale space [18]. A disadvantage is that in these scale spaces the

relation between scale and spatial location is less obvious.

References

- [1] L. M. J. Florack. *Image Structure*, volume 10 of *Computational Imaging and Vision Series*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.
- [2] B. van Ginneken and B. M. ter Haar Romeny. Applications of locally orderless images. In *Nielsen et al. [10]*, pages 10–21, 1999.
- [3] L. D. Griffin. Scale-imprecision space. *Image and Vision Computing*, 15:369–398, 1997.
- [4] B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink, and M. A. Viergever, editors. *Scale-Space Theory in Computer Vision: Proceedings of the First International Conference, Scale-Space'97, Utrecht, The Netherlands*, volume 1252 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, July 1997.
- [5] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [6] J. J. Koenderink and A. J. Van Doorn. Blur and disorder. In *Nielsen et al. [10]*, pages 1–9, 1999.
- [7] A. Kuijper. *The Deep Structure of Gaussian Scale Space Images*. PhD thesis, Utrecht University, 2002.
- [8] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality Image Registration by Maximization of Mutual Information. *IEEE Transaction On Medical Imaging*, 16(2):187–198, April 1997.
- [9] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373–86, 1999.
- [10] M. Nielsen, P. Johansen, O. Fogh Olsen, and J. Weickert, editors. *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*. Springer -Verlag, Berlin Heidelberg, 1999.
- [11] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. A multiscale approach to mutual information matching. In K.M. Hanson, editor, *Medical Imaging: Image Processing*, volume 3338 of *Proc. SPIE*, pages 1334–1344. SPIE Press, Bellingham, WA, 1998.
- [12] L. Schwartz. *Théorie des Distributions*, volume I, II of *Actualités scientifiques et industrielles; 1091,1122*. Publications de l’Institut de Mathématique de l’Université de Strasbourg, Paris, 1950–1951.
- [13] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948. Reprinted with correction.
- [14] J. Sporring. The entropy of scale-space. In *ICPR*, volume A, pages 900–904, 1996.
- [15] J. Sporring and J. Weickert. On generalized entropies and scale-space. In *Ter Haar Romeny et al. [4]*, pages 53–64, 1997.
- [16] C. Studholme, D.L.G Hill, and D.J. Hawkes. Automated 3D registration of MR and CT images of the head. *Medical Image Analysis*, 1(2):163–175, 1996.
- [17] P.A. Viola and W.M. Wells III. Alignment by maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):137–154, 1997.
- [18] J. A. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.
- [19] W.M. Wells III, P.A. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-Modal Volume Registration by Maximization of Mutual Information. *Medical Image Analysis*, 1(1):35–51, 1996.
- [20] A. Winter, H. Maitre, N. Cambou, and E. Legrand. Entropy and multiscale analysis: a new feature extraction algorithm for aerial images. In *IEEE ICASSP97, Munich, April 1997.*, 1997.