

# Local Similarity Measures for Multimodal Image Matching

Peter Rogelj, Stanislav Kovacic  
Faculty of Electrical Engineering, University of Ljubljana  
Tržaška 25, 1000 Ljubljana  
*peter.rogelj@fe.uni-lj.si, stanek@fe.uni-lj.si*

## Abstract

*In this paper we focus on local similarity measures based on Shannon entropy which can be used for multimodal image matching employing deformations. The advantage of our approach is that global similarity or similarity of a larger image region can be computed from the similarities of its constitutive parts or individual voxels. We also discuss the interpolation artefacts in entropy based similarity measures caused by linear and partial volume interpolation.*

**Keywords:** similarity measure, multimodality, entropy, mutual information, elastic matching

## 1. Introduction

Image matching is usually used to spatially align one image to another. For that purpose first image, sometimes called target, has to be geometrically transformed, or aligned, to achieve high spatial correspondence with the second image, called reference image. Similarity measure, which serves to evaluate the spatial correspondence of images, plays a crucial role in this process. When we need to select the measure of similarity, there are lots of issues to consider, but the most important ones are domain of transformation, modalities involved, and optimization method used to align the images. Selection of an appropriate similarity measure becomes even more important when matching involves deformations (Bajcsy and Kovacic, 1989, Lester and Arridge 1999). In this case similarity measure is used not only to estimate the differences between the images, but also to define the direction and strength of the forces, which deform one image to make it more similar to the other image. The domain, i.e. the size of an image region used for similarity determination, also influences the rigidity of the model. As the region size increases the elasticity of model decreases, and it is more difficult to detect and correct small local differences. In some sense, the region size effectively limits the maximum possible elasticity that can be achieved by elastic model or the maximum possible viscosity of viscous fluid model. When using such models, properties of deformations are expected to be defined by spatial model, and therefore, the similarity measure is desired to have small influence on them. This motivated us to focus on similarity measures which could be suitable for image matching, in particular multimodal image matching, involving local deformations. Thus, we limit ourselves to local multimodal similarity measures. More specifically, we investigate how widely used multimodal entropy based similarity measures could be modified to reflect local image properties.

## 2. Multimodal similarity measure

The most commonly used multimodal similarity measure is mutual information. It was first proposed and brought to the medical imaging field by Viola and Wells (Viola and Wells,

1995). The definition of mutual information is based on statistics and its origins are in information theory. When used as a similarity measure it measures the statistical dependence between the image intensities. In principle, it reveals how much one image tells us about the other image, and if suitably defined, it takes on maximum value when the images are geometrically aligned. Given two images  $A$  and  $B$  mutual information  $I(A,B)$  between them is defined as

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

where  $H(A)$  and  $H(B)$  denote marginal entropies of  $A$  and  $B$ , and  $H(A,B)$  is their joint entropy. The entropies can be calculated using well-known Shannon definition:

$$H(\cdot) = -\sum p(\cdot) \log p(\cdot), \quad (2)$$

where  $p(\cdot)$  stands for either marginal or joint probability distributions, estimated from image intensities. There are also other types of multimodal similarity measures like energy similarity measure (Buzug et al., 1997). Various types of generalised entropies (Taneja, 1989) could be used as well. Nevertheless, due to their statistical nature they are all inherently global in the sense that they can only be estimated from large enough image regions, and as such they can not be used directly to estimate local image properties. If the region covered by the similarity measure is small, the statistical significance of joint probability is low and similarity function is not well defined. To alleviate this problem one can reduce the number of histogram bins or resort to similarity measures based on 1D histograms, e.g.  $H(A-B)$ . Another approach is to use a prior joint probability (Likar and Pernuš, 1999). In this case joint probability used for calculating mutual information is a combination of the joint probability, derived from the image region, and the prior joint probability, which can be obtained from a pre-registered training set. Local similarity measure based on global conditional probabilities was also used (Maintz et al., 1998). In the next section we describe our approach.

### 3. Local multimodal similarity measure

We propose a local similarity measure derived from Shannon entropy that can be also used as a voxel based similarity measure. It is noted (Baens et al., 1998, Maintz et al., 1998) that matching based on individual voxels is ill-posed if displacements are calculated separately for each image voxel. Thus, matching schemes using voxel based similarity measures must incorporate spatial model, which in our case is the model of elastic deformation. Let us rewrite Eq. (2) in the following form,

$$H = -\sum_{i,j} p_{i,j} \log p_{i,j} = -\sum_{i,j} \frac{N_{i,j}}{N} \log p_{i,j}, \quad (3)$$

where  $i$  and  $j$  denote image intensities of reference and target image, and  $p_{i,j}$  is the joint probability of intensity pair  $(i, j)$ .  $N_{i,j}$  is the number of occurrences of this intensity combination, and  $N$  is the total number of intensity pairs in the image, which is usually equal to the number of image voxels. Let us assume that the images are divided into smaller non-overlapping regions  $R_r$ , each containing, say,  $N_{i,j}^r$  occurrences of intensity combination  $(i, j)$ . The total number of occurrences  $N_{i,j}$  in the whole image can be obtained by summing  $N_{i,j}^r$  over all regions,

$$N_{i,j} = \sum_r N_{i,j}^r. \quad (4)$$

Substituting Eq. (4) in Eq. (3) we obtain

$$H = -\sum_r \sum_{i,j} \frac{N_{i,j}^r}{N} \log p_{i,j}. \quad (5)$$

If the image is divided into individual voxels  $v = (x,y,z)$ , then

$$N_{i,j}^r = N_{i,j}^v = \begin{cases} 1 & ; i = A(v), j = B(v) \\ 0 & ; \text{otherwise} \end{cases}, \quad (6)$$

and Eq. (5) can be simplified to

$$H = -\sum_r \frac{1}{N} \log p_{i,j} = \sum_v -\frac{1}{N} \log p_v = \frac{1}{N} \sum_v h(v), \quad (7)$$

where  $p_v$  is the probability  $p_{i,j}$  with  $i$  equal  $A(v)$  and  $j$  equal  $B(v)$ , and  $h(v) = -\log p_v$  is the ‘‘uncertainty’’ of this intensity combination. Thus, it is possible to calculate the entropy, which is global similarity measure, from voxel based similarities  $h(v)$  by summing up over all image regions, i.e. voxels. Of course, we can also sum up over image regions of arbitrary size. Thus, local similarity measure  $H^r$  of region  $R_r$  can be calculated from voxel similarities in this region,

$$H^r = \frac{1}{N_r} \sum_{v \in R_r} h(v). \quad (8)$$

If the selected region covers the whole image, the result is the entropy of entire image. Notice that  $h(v)$  is always estimated out of the whole image.

### 3.1 Generalisation

Observing Eq. (7) it is easy to see its close relation with *log likelihood* similarity measure described in (Leventon and Crimson, 1998). The only difference is in the way how the probability  $p(\cdot)$  is derived. In our case, it is estimated from the image pair, while in case of likelihood similarity measure the probability  $p(\cdot)$  is a prior information. It is also possible to use a combination of both (Likar and Pernuš, 1999). Moreover, instead of joint intensity probability  $p(A,B)$ , it is possible to use probability of intensity difference  $p(A-B)$ , or conditional intensity probability  $p(A/B)$ . Finally, it is possible to replace the *log* function with any other function  $f(p)$  that meets the requirement that the first derivative of function  $p \neq f(p)$  is strictly monotonically increasing or decreasing. If it is increasing/decreasing, higher/lower value means better correspondence. When linear function  $f(p)=p$  is used similarity measure is computed as the sum of squared histogram values and therefore it is often called *energy* similarity measure.

Based on observations mentioned above the voxel based local similarity measure derived from entropy, can be more generally written as

$$S^v = f(p); \quad S^r = \sum S^v = \sum f(p), \quad (8)$$

where  $S^v$  is voxel based similarity measure,  $S^r$  is similarity measure for region  $r$ , and  $f(p)$  can be any function that meets the requirement mentioned before. Probability  $p$  may be joint probability  $p(A,B)$ , difference probability  $p(A-B)$ , conditional probability  $p(A/B)$ , which are derived from image pair, or combination of intensity probability derived from image pair  $p_{image}$  and prior probability  $p_{prior}$  that is given in advance. In the most general case we have

$$p = I \cdot p_{image} + (1 - I) \cdot p_{prior} ; I \in [0,1] \quad (9)$$

where  $\lambda$  is weighting parameter.

#### 4. Comparative study of voxel based similarity measures

A reasonable approach to compare global similarity measures is to apply them on various transformations and to evaluate their properties, such as the number of local extrema, their smoothness, position of global maximum, capture range, etc. However, comparing voxel based similarity measures turns to be more problematic, because, as we mentioned before, the results do not depend only on similarity measure but also on spatial model of elasticity. We argue that voxel based similarities cannot be tested in isolation without considering spatial model, and it appears that the only meaningful way to compare them is their treatment in the context of complete matching systems.

To show the relative performance of our local similarity measures we used two spatially aligned MRI sets of human head. The first image (T1) was transformed with a known transformation and then deformed back to match the second image (PD) using different versions of local similarity measure within the same elastic deformation model. The overall matcher performance was evaluated by using the average voxel displacement (AVD) as performance criterion, defined as the absolute difference between voxel displacements achieved by matcher and the displacements derived by known inverse transformation. Lower AVD value means better performance. The results obtained using different selections of  $p$  are tabulated below.

Table 1: Comparison results of voxel based similarity measures.

		AVD			CR		
$\lambda$	$f(p)$	$p(A,B)$	$p(A-B)$	$p(A \hat{B})$	$p(A,B)$	$p(A-B)$	$p(A \hat{B})$
1	$\log(p)$	1.5725	1.8121	1.5725	0.9278	0.9332	0.9278
	$p$	2.6153	2.3150	1.9323	0.7182	0.8283	0.8271
0	$\log(p)$	1.5180	1.7981	1.5180	0.9339	0.9106	0.9339
	$p$	2.5600	2.4834	1.6861	0.7260	0.8024	0.8915

Table 1 shows that prior information ( $\lambda=0$ ) in general gives better result than the probability derived from image pair ( $\lambda=1$ ). This was expected, because in our experiment prior probability distribution was correctly estimated from already aligned images. Unfortunately, in general this is not the case. Nevertheless, when estimated properly, it can improve matching results. Furthermore, the logarithmic function shows better performance than linear function. Among different probabilities conditional probability  $p(A \hat{B})$  performed in all cases best.

For illustrative purposes normalized cross correlation ratio (CR) between the original T1 image and recovered T1 image in addition to AVD was used as performance criterion. Ideally, CR should be 1. While both criteria show similar results, we need to mention that using CR as performance measure might be misleading, because aligned images can correlate well although transformation is not correct. For example,  $CR(\log p(A - B), \lambda=1)$  shows good correlation, while AVD indicates higher residual differences than in some other cases.

There is another problem that needs to be mentioned. Because we are dealing with discrete images, as one of the images is exposed to the deformation process, the positions of voxels do not necessarily lie on the regular grid and an interpolation method has to be employed to

determine the image intensity probabilities. This problem is present in global as well as local approaches. Fig. 1 shows similarity functions obtained for different multimodal similarity measures with respect to translational misalignment, using linear interpolation and partial volume interpolation (Maes et al., 1999). It is evident that linear interpolation (Fig. 1 – top row) causes high interpolation artefacts and many local extrema, making this kind of interpolation inappropriate for practical use. It is possible to improve its performance by random resampling of one of the images (Pluim et al., 1999). While random resampling improves the performance of global similarity measures, it becomes less appropriate as the region size in which similarity is estimated gets smaller.

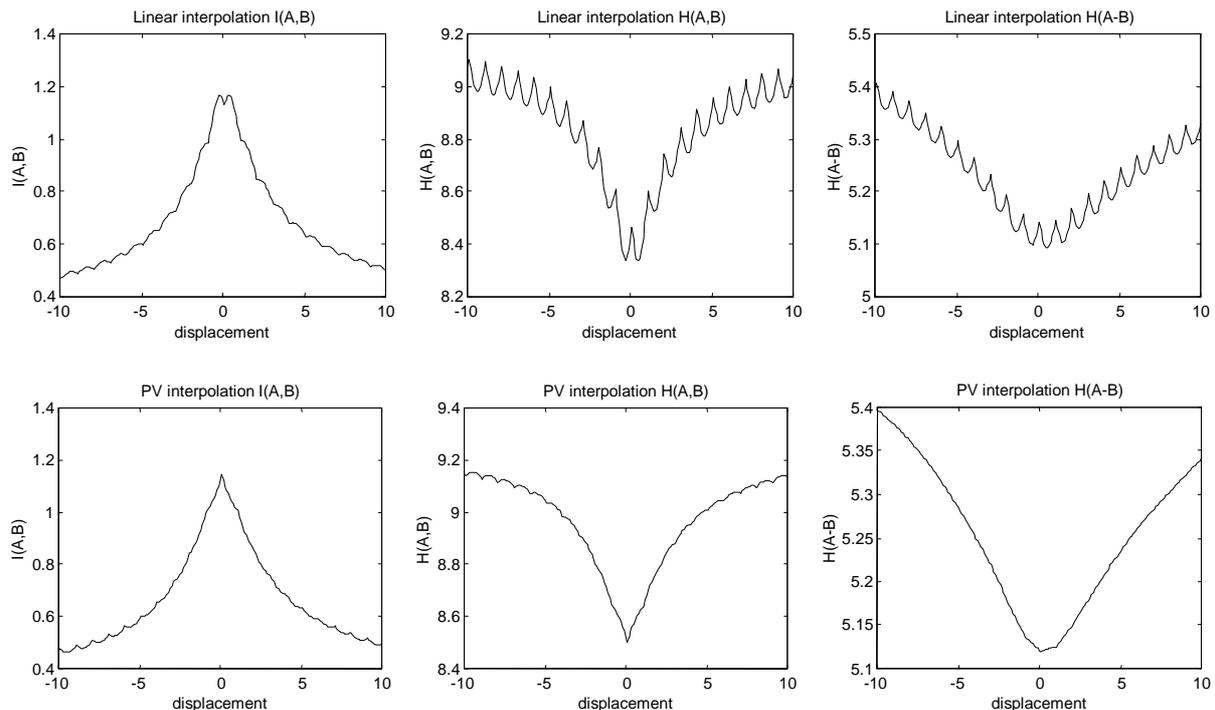


Fig. 1: Similarity functions for translation, using MRI PD and MRI T1 image of human head.

Partial volume interpolation (PV interpolation shown in Fig. 1 – bottom row) results in a much smoother similarity function with lower interpolation artefacts. We have found that interpolation artefacts do not appear in the marginal entropies  $H(A)$  and  $H(B)$  if partial volume interpolation is used. This makes joint entropy  $H(A,B)$  appropriate to be used individually as the similarity measure. From the examples in Fig. 1 it is evident that even PVI causes some interpolation artefacts that cause local extrema in  $H(A,B)$  and therefore in  $I(A,B)$ . The interpolation artefacts have smaller influence on  $H(A-B)$  as a similarity measure, but the function itself has wider global extremum and as such it is more appropriate for the initial steps of the matching process. In general, partial volume interpolation is recommended when entropy based similarity measures are used.

## 5. Conclusions

We have presented various similarity measures for local multimodal image matching. They are all based on probability distribution. To achieve high statistical significance, probability is determined using whole image content and prior information, if available. If prior information can not be used, high initial correspondence of images is required to prevent incorrect matching, caused by incorrect determination of probability distribution. Different similarity

measures give different results. Which one to use depends on the problem wished to solve. Generally, probability of image difference  $p(A-B)$  is more appropriate for initial steps of image matching while joint probability  $p(A,B)$  and conditional probability  $p(A/B)$  are more appropriate for final steps as they better define intensity correspondences. In all cases partial volume interpolation is recommended.

## References

1. R. Bajcsy and S. Kovacic (1989), "Multiresolution elastic Matching", in *Computer Vision, Graphics and Image Processing*, 46, pp 1-21.
2. T. M. Buzug, J. Weese, C. Fassnacht and C. Lorenz (1997), "Elastic matching based on motion vector fields obtained with a histogram based similarity measure for DSA-image correction", in *Computer Assisted Radiology and Surgery*, ed. by H.U. Lemke, M. W. Vannier and K. Inamura, pp 139-144.
3. T. Gaens, F. Maes, D. Vandermeulen, and P. Suetens (1998), "Non-rigid Multimodal Image Registration Using Mutual Information", in *Medical Image Computing and Computer-Assisted Intervention - MICCAI'98*, ed. by W.M. Wells, A. Colchester, S. Delp, Cambridge, MA, Springer LNCS 1496, pp 1099.
4. H. Lester and S. R. Arridge (1999), "A survey of hierarchical non-linear medical image registration", in *Pattern Recognition*, Vol. 32, No. 1, pp 129-149.
5. M.E. Leventon and W.E.L. Crimson (1998), "Multi-modal Volume Registration Using Joint Intensity Distributions", in *Medical Image Computing and Computer-Assisted Intervention - MICCAI'98*, ed. by W.M. Wells, A. Colchester, S. Delp, Cambridge, MA, Springer LNCS 1496, pp 1057.
6. B. Likar and F. Pernuš (1999), "A Hierarchical Approach to Elastic Registration Based on Mutual Information", in *Proc. of Workshop on Biomedical Image Registration – WBIR'99*, ed. by F. Pernuš, S. Kovacic, H. S. Stiehl, M. A. Viergever, Bled, Slovenia, pp 24-45.
7. F. Maes, D. Vandermuelen and P. Suetens (1999), "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information", in *Medical Image Analysis*, Vol. 3, No. 4, pp 373-386.
8. J. B. A. Maintz, H.W. Meijering and M. A Viergever (1998), "General multimodal elastic registration based on mutual information", in *Medical Imaging 1998 - Image Processing*, Vol. 3338, ed. by Hanson, K. M., pp 144-154.
9. J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever (1999), "Interpolation artefacts in mutual information based image registration", in *Computer vision and image understanding*, In press.
10. I. J. Taneja (1989), "On Generalized Information Measures and Their Applications", in *Advances in Electronics and Electron Physics*, Academic Press (USA), 76, pp 327–413.
11. P. Viola and W. M. Wells III (1995), "Alignment by Maximization of Mutual Information", in *Proceedings of the 5th International Conference on Computer Vision*, pp 16-23