

# CONTINUOUS IMAGE REPRESENTATIONS AVOID THE HISTOGRAM BINNING PROBLEM IN MUTUAL INFORMATION BASED IMAGE REGISTRATION

*Ajit Rajwade, Arunava Banerjee, Anand Rangarajan*

Dept. of CISE, University of Florida, Gainesville, FL

## ABSTRACT

Mutual information (MI) based image-registration methods that use histograms are known to suffer from the so-called binning problem, caused by the absence of a principled technique for choosing the “optimal” number of bins to calculate the joint or marginal distributions. In this paper, we show that foregoing the notion of an image as a set of discrete pixel locations, and adopting a continuous representation is the solution to this problem. A new technique to calculate joint image histograms is proposed, which makes use of such a continuous representation. We report results on affine registration of a pair of 2D medical images under high noise, and demonstrate the smoothness of various information-theoretic similarity measures such as joint entropy (JE) or MI w.r.t. the transformation, when our proposed technique (referred to as the “**robust histogram**”) is adopted to compute the required probability distributions.

## 1. INTRODUCTION

MI has become a popular similarity measure for medical image registration, ever since the pioneering work of Viola *et al* [1]. A comprehensive survey of existing MI based image registration techniques has been presented by Pluim *et al* in [2]. Density-estimation techniques (both marginal and joint) lie at the core of all these methods. So far, histograms have been the most popular strategy for density estimation in this context, owing to their simplicity and efficiency. It has been empirically seen that MI performs better if the number of histogram bins is “small” in the marginal as well as the joint histograms, due to the accompanied noise reduction. However, there is no principled method to choose the “correct” number of bins, especially for images of different sizes. Traditionally, images are represented as a set of discretized locations called pixels, which are locally flat in terms of intensity. In this paper, we demonstrate that the binning problem can be solved by computing histograms that consider a continuous representation for an image, regarding it as a continuum of points. We empirically show how these “robust” histograms yield an increasingly better approximation to the joint density, as the number of chosen intensity levels increases, in contrast

to standard histograms, where the distribution is known to become ever sparser and noisier.

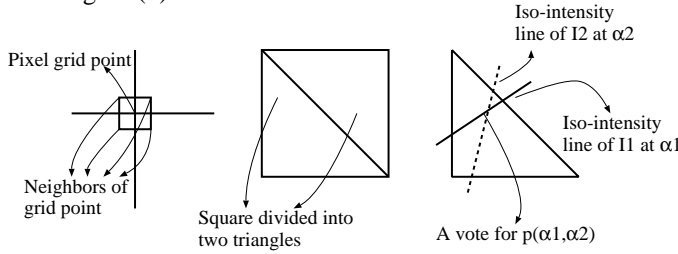
## 2. THEORY: ROBUST HISTOGRAMS FOR JOINT AND MARGINAL DISTRIBUTIONS

Consider an image  $I$  of size  $N$  by  $N$  with intensity values quantized into  $P$  bins. In standard image histograms, each pixel is regarded as “flat” and contributes to exactly one bin. If the number of bins is too large, one would obtain a distribution in which most bins are empty. There exists no principled way to predict the exact number of bins as a function of the image size. Most MI-based methods using histograms adopt a smaller number of bins, for greater noise resistance [3]. Although this improves the results, it can cause over-smoothing and fails in cases where a large number of bins is actually essential for correct registration. For instance, such a situation can arise in the case of registration of MR data to CT, where very small changes in CT values could correspond to large changes in MR values, or in the case of non-rigid image registration. Some studies suggest using a combination of three different MI values computed using three different numbers of bins [4], which would partly remedy the problem. However that may not generalize to all data-sets, and would require manual selection of the three (or more) bin-width values.

In our approach, we consider a spatially continuous representation of an image. In the experiments described here, we have employed a locally linear representation of the intensity function, though this could be trivially replaced by any other method. Although we consider image intensity to be a continuous entity, we estimate the probability distribution only at a fixed number ( $Q$ ) of equi-spaced intensity levels. Later on in the paper, we show that as  $Q$  increases, we obtain an increasingly accurate estimate of the distribution, quite unlike standard histograms. As such, our approach does not involve direct calculation of the marginal histograms of an image. Rather, we first compute the joint probability between a pair of images and use that to estimate the marginal probabilities. This is described in more detail in the following sections.

## 2.1. Calculating the Joint Distribution

We calculate the joint distribution of a given pair of images (named  $I_1$  and  $I_2$ ), whose intensity is quantized into  $P$  bins, as follows. (In all symbols used here, the subscript 1 indicates the first image, and the subscript 2 indicates the second image). Consider that the joint distribution is to be evaluated at exactly  $Q^2$  intensity levels,  $Q$  in each image. For each image grid point, we estimate the intensity values of its four neighbors lying at a horizontal or vertical distance of 0.5 pixels. We divide the square defined by these neighbors into a pair of triangles (see Figure (1)). The intensity values within the triangle can be considered to be a linear function of the coordinates of its vertices, given by the equation  $z = ax + by + c$ . Here  $z$  denotes intensity, and the coefficients  $a$ ,  $b$  and  $c$  are obtained by solving three simultaneous equations. To calculate the joint distribution of two images, we sequentially consider the  $Q^2$  different intensity pairs, denoted as  $(\alpha_1, \alpha_2)$ . Each pair of corresponding triangles from the two images is tested to see whether it contains a point  $(x, y)$  which has intensity value  $\alpha_1$  in  $I_1$  and  $\alpha_2$  in  $I_2$ . Such a point  $(x, y)$  then contributes a vote to the entry  $(\alpha_1, \alpha_2)$  in the joint distribution. See Figure (1).

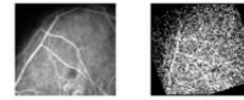


**Fig. 1.** Iso-intensity lines in an image triangle

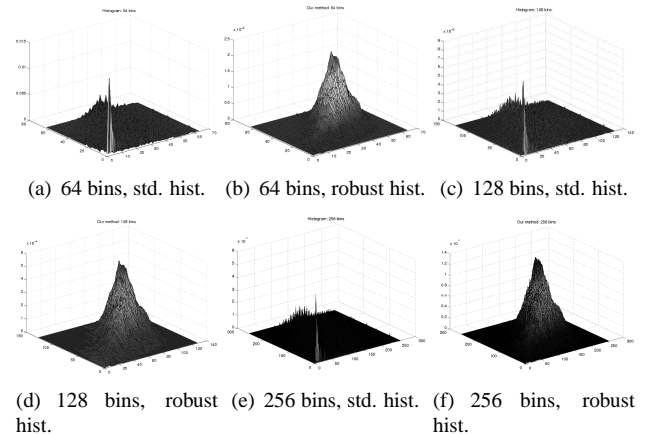
Due to the assumed local linearity of the intensity function, the iso-intensity curves within the triangle are straight lines. Any point  $(x, y)$  such that  $I_1(x, y) = \alpha_1$  and  $I_2(x, y) = \alpha_2$  can be considered to be the intersection of the iso-intensity lines at level  $\alpha_1$  in  $I_1$  and at level  $\alpha_2$  in  $I_2$ , given by the equations  $a_1x + b_1y + c_1 = \alpha_1$  and  $a_2x + b_2y + c_2 = \alpha_2$  respectively. Solving these equations simultaneously yields the values of  $x$  and  $y$ . If the point  $(x, y)$  thus calculated lies *within* the triangle under consideration, then and only then do we add its vote to the joint probability of  $(\alpha_1, \alpha_2)$ . This entire process is repeated for each such  $(\alpha_1, \alpha_2)$  within every pair of corresponding triangles in the two images, to yield the joint distribution. Note that as the number of intensity levels  $Q$  is increased, one would expect to obtain a more and more accurate estimate of the joint distribution, due to our adoption of a continuous representation for the image, completely unlike the case with standard histograms. The contrast between the two approaches is shown clearly in the surface plots in Figure (3) for  $P$  (and  $Q$ ) equal to 32, 64 and 128. The images used were a retinogram and its noisy in-plane rotated version [see Figure(2)]. The difference is due to the fact that in robust

histograms, a pixel (or triangle) contributes votes to *multiple* intensity levels.

Now consider a case in which the joint distribution is to be evaluated at exactly  $Q$  intensity levels in each image, with  $Q < P$ . In such a case, the robust histogram will yield a poor estimate of the distribution due to scarcity of votes. This is because robust histograms collect votes only *at*  $Q$  levels (and  $Q < P$ ), whereas standard histograms collect votes for a *range* of levels. To remedy this problem, the images should be requantized so as to now contain exactly  $Q$  intensity levels. Upon requantization, the robust histogram will now acquire many more votes than the standard histogram. This is because the former performs sub-pixel interpolation and therefore counts level curve intersections that occur *in between* the grid points [see Figure (1)], apart from the grid-points themselves. Standard histograms will count the grid-points *alone*.



**Fig. 2.** A retinogram and its noisy, rotated version



**Fig. 3.** Joint Probability plots, with standard histogramming and our method for 64, 128 and 256 bins in each image.

## 2.2. Calculating the Marginals

Since each vote in the joint histogram stems from an intersection of the corresponding level curves from the two images, the marginal probabilities can be obtained by tracing out such iso-intensity contours at  $Q$  intensity levels in the individual images and calculating their length. This process, however, would be time-consuming. Therefore we abandon this strategy, and instead calculate the marginals by row- and column-wise addition of the joint probabilities (which is analogous to integration). This process also ensures consistency between the joints and the marginals.

## 2.3. Effect of Noise

It is evident that noise in the pair of images will alter their probability distribution obtained by this method (as by any

Metric	Average Error	
	Robust Hist.	Std. Hist.
JE	0.5°	4.8°
MI	0.5°	3.6°
NMI	0.5°	3.6°

**Table 1.** Error using JE, MI and NMI

other). However the change due to noise would be far less significant in our case, as compared to standard histogramming. This is because ordinary binning erratically switches the votes from one bin to another. Our method avoids this, because of the continuous image representation, and therefore is inherently much more noise-resistant. Our experimental results support this.

### 3. EXPERIMENTS AND RESULTS

We applied our method of histogram calculation for the purpose of pairwise affine multimodality registration. The joint and marginal distributions were calculated using the method just described and were then used to compute MI and Normalized Mutual Information (NMI). All or some of these values were used as energy functions in the search for the optimal transformation.

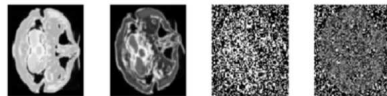
#### 3.1. Smoothness of MI w.r.t Transformation

The first experiment was performed on an MR-PD slice and an MR-T2 slice from a brain volume dataset that was obtained from [5]. Both slices were initially registered. The T2 image was rotated in-plane by  $-20^\circ$ . Gaussian noise of zero mean and a variance of 1 (using the ‘imnoise’ function of MATLAB) was added onto it (see Figure (4)). Now, we sought to iteratively move the PD image so that it aligned with the T2 image, while maximizing the MI, or minimizing the JE between the two images. The search for the optimal transformation was done in a brute-force manner within a range of  $-30^\circ$  to  $+10^\circ$ , and the results were compared to JE and MI as computed using the standard histogram. The number of bins in both cases was chosen to be 128, assuming 256 intensity levels in the original images. From Figure (6), we clearly observe the smoothness of both MI and JE when calculated by our method. Smooth energy functions can come in handy when abandoning brute-force search for the more efficient gradient-based search methods. Over 50 trials between the same pair of images, the average angle error for MI with robust histograms was  $0.5^\circ$ , as against  $3.6^\circ$  with standard histogramming (see Table(1)). Note that we treated the image that was initially rotated through  $20^\circ$  as the fixed image, and the other one as the floating image. This helped avoid the interpolation artifacts reported by Pluim in [6], namely the “dip at the ground truth” effect of MI calculated using standard histograms, due to interpolation issues. Also, note that a

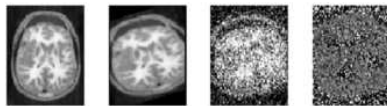
Metric	Predicted Transformation					
	MI			NMI		
	$\theta$	$s$	$t$	$\theta$	$s$	$t$
Robust Hist.	-18	-0.3	-0.3	-18	-0.3	-0.3
Std. Hist.	-17	0.6	0.6	-17	0.6	0.6

**Table 2.** Predicted parameters, ground truth is  $\theta = -20$ ,  $s = -0.3$ ,  $t = -0.1$

brute-force search was employed only in order to ensure that the results of either method were not affected by optimization issues such as local minima. Our method is in no way tied to only a brute-force search.



**Fig. 4.** (a) An MR-PD image, (b) An MR-T2 image (formerly aligned with the PD image) now rotated by  $-20^\circ$ , (c) A noisy version of (b), (d) The difference between (c) and (b)



**Fig. 5.** (a) Anatomical slice of the brain, (b) Deformed version of (a), (c) A noisy version of (b), (d) The difference between (c) and (b)

#### 3.2. Affine Image Registration

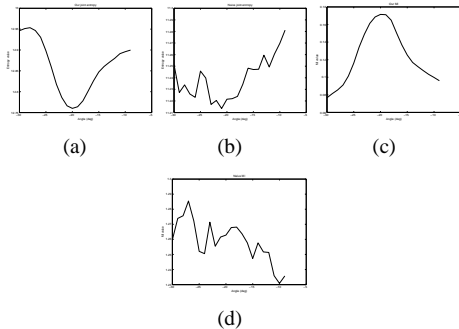
We also tested the robust histogram for affine image registration, involving non-uniform scaling and a single rotation. We used an axial anatomical slice of the brain from the VHD Male Dataset (slice 1080), as shown in Figure (5). The original slice was synthetically warped by a single rotation of  $\theta = -20^\circ$  and non-uniform scaling factors  $s = -0.3$  and  $t = -0.1$ . Gaussian noise of variance 0.25 and mean 0 was added onto it. The details of the affine transformation matrix are given in [7]. In our experiments, the value of  $\phi$  was set to 0 for simplicity. A multi-resolution brute-force search was performed to estimate the parameters that maximized MI or NMI between the two images. The range of search was  $-30^\circ$  to  $+10^\circ$  for  $\theta$  and  $-0.6$  to  $0.6$  for  $s$  and  $t$ . The predicted parameters are shown in Table (2). Clearly, the robust histogram outperformed the standard one when using both MI and NMI.

In yet another experiment, we also included a translation of 2 and 3 pixels in the  $X$  and  $Y$  directions respectively, besides a  $-20^\circ$  rotation, and a scaling of  $s = -0.3$ ,  $t = -0.3$ . The images used were a pre-registered MR-PD and MR-T2 slice as obtained from [5]. All transformations were applied to the T2 slice and zero mean Gaussian noise of variance 0.25 was added to it. The PD slice was transformed between the range  $\theta = [-24^\circ, -12^\circ]$ ,  $s = [-0.5, 0.5]$ ,  $t = [-0.5, 0.5]$ ,

	Results with MI				
	$\theta$	$s$	$t$	$t_x$	$t_y$
Robust Hist.	-20	-0.3	-0.3	3	3
Std. Hist.	-18	-0.5	0.4	0	0
	Results with NMI				
	$\theta$	$s$	$t$	$t_x$	$t_y$
Robust Hist.	-18	-0.3	-0.3	3	3
Std. Hist.	-11	-0.1	-0.1	3	3

**Table 3.** Predicted parameters, ground truth is  $\theta = -20$ ,  $s = -0.3$ ,  $t = -0.3$ ,  $t_x = 2$ ,  $t_y = 3$

$t_x = [-3, 3]$ ,  $t_y = [-3, 3]$  to search for optimal alignment. The results summarized in Table (3) clearly show that the robust histograms performed far better.



**Fig. 6.** Trajectories of JE and MI w.r.t. rotation (128 bins): (a) JE, robust hist., (b) JE, std. hist., (c) MI, robust hist., (d) MI, std. hist.

#### 4. DISCUSSION

We have presented a new histogramming technique using a continuous representation of an image. This technique has shown good results for affine image registration under high image noise, regardless of the chosen number of intensity levels  $Q$ . As  $Q$  increases, we would expect a more and more accurate estimate of the joint probability distribution.

Notice that our technique will avoid the “**small sample size problem**” in the estimation of local MI over small-sized image windows. Though this can be accomplished by Parzen windows as well, our technique requires *no* parameter tuning, unlike the latter, and is more efficient in terms of time complexity. Some techniques such as [8] use a MAP estimate with a uniform prior histogram for sparse data, but these require careful task-dependent selection of an optimal number of samples and coefficients to weigh the likelihood and prior. Others such as [9] which use a combination of floating and prior joint probabilities also require careful parameter tuning. We also wish to point out an important difference between our method and Partial Volume Interpolation (PVI) proposed by Maes *et al* [10]. The latter assigns (partial) votes to only those intensity values that occur at the corners of a pixel, in the

case that image intensities are assigned to non-grid positions (which can occur as images are being successively warped during registration). As such, it does not assign any votes to intensities that lie in between grid locations, and hence is *not* a method to solve the binning problem.

Our method is easily extensible to the pairwise registration of volume datasets, where each voxel could be split into two tetrahedra. Joint probabilities can be computed by finding the lengths of the intersection of iso-intensity planes in the corresponding tetrahedra of the two images, which turns out to be straight line segments. However, it should be noted that our method is not extensible to groupwise image registration in 2D because it requires the concurrency of three or more (iso-intensity) lines, one from each image. This is an over-constrained condition both in 2D as well as in 3D.

A striking feature of our technique is that it favours sampling in regions of high gradient. This is because the intensity levels at which the joints are computed, are chosen at equally spread-out intervals, and the regions with high gradient values accommodate many more intensity levels than the flatter regions of equal size. This issue of non-uniform sampling (dictated by gradient magnitude, a measure of feature saliency) demands deeper theoretical investigation. Besides studying these issues, we also plan to use our technique in a non-rigid setting. Here, superior performance could well be expected with a greater value of  $Q$ , as smaller  $Q$  values could reduce the accuracy of computation of small local deformations due to oversmoothing.

#### 5. REFERENCES

- [1] P. Viola and W. M. Wells, “Alignment by maximization of mutual information,” *Int. J. Comput. Vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [2] J. P. W. Pluim, J. B. A. Maintz, and M. Viergever, “Mutual information based registration of medical images: A survey,” *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.
- [3] Y. Zhu and S. Cochoff, “Influence of implementation parameters on registration of MR and SPECT brain images by maximization of mutual information,” *Journal of Nuclear Medicine*, vol. 43, no. 2, pp. 160–166, 2002.
- [4] R. Shekhar and V. Zagrodsky, “Mutual information-based rigid and nonrigid registration of ultrasound volumes,” *IEEE Trans. Med. Imaging*, vol. 21, no. 1, pp. 9–22, 2002.
- [5] D. L. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. J. Kabani, C. Holmes, and A. Evans, “Design and construction of a realistic digital brain phantom,” *IEEE Trans. Med. Imaging*, vol. 17, no. 3, pp. 463–468, 1998.
- [6] J. P. W. Pluim, J. B. A. Maintz, and M. Viergever, “Interpolation artefacts in mutual information-based image registration,” *Comput. Vis. Image Underst.*, vol. 77, no. 9, pp. 211–232, 2000.
- [7] J. Zhang and A. Rangarajan, “Affine image registration using a new information metric,” in *CVPR (1)*, 2004, pp. 848–855.
- [8] M. Toews, D. L. Collins, and T. Arbel, “Maximum a posteriori local histogram estimation for image registration,” in *MICCAI (2)*, 2005, pp. 163–170.
- [9] B. Likar and F. Pernus, “A hierarchical approach to elastic registration based on mutual information,” *Image Vision Comput.*, vol. 19, no. 1-2, pp. 33–44, 2001.
- [10] F. Maes and A. Collignon *et al*, “Multimodality image registration by maximization of mutual information,” *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 187–198, 1997.