Published in final edited form as:

IEEE Trans Pattern Anal Mach Intell. 2009 March ; 31(3): 475–491. doi:10.1109/TPAMI.2008.97.

Probability Density Estimation Using Isocontours and Isosurfaces: Application to Information-Theoretic Image Registration

Ajit Rajwade [Student Member, IEEE], Arunava Banerjee [Member, IEEE], and Anand Rangarajan [Member, IEEE]

The authors are with the Department of Computer and Information Science and Engineering, University of Florida, CSE Building, Gainesville, FL 32611-6120.

Ajit Rajwade: avr@cise.ufl.edu; Arunava Banerjee: arunava@cise.ufl.edu; Anand Rangarajan: anand@cise.ufl.edu

Abstract

We present a new geometric approach for determining the probability density of the intensity values in an image. We drop the notion of an image as a set of discrete pixels and assume a piecewise-continuous representation. The probability density can then be regarded as being proportional to the area between two nearby isocontours of the image surface. Our paper extends this idea to joint densities of image pairs. We demonstrate the application of our method to affine registration between two or more images using information-theoretic measures such as mutual information. We show cases where our method outperforms existing methods such as simple histograms, histograms with partial volume interpolation, Parzen windows, etc., under fine intensity quantization for affine image registration under significant image noise. Furthermore, we demonstrate results on simultaneous registration of multiple images, as well as for pairs of volume data sets, and show some theoretical properties of our density estimator. Our approach requires the selection of only an image interpolant. The method neither requires any kind of kernel functions (as in Parzen windows), which are unrelated to the structure of the image in itself, nor does it rely on any form of sampling for density estimation.

Index Terms

Probability density estimation; isocontours; isosurfaces; level sets; transformation of random variables; computational geometry; image registration

1 INTRODUCTION

Information-theoretic tools have for a long time been established as the de facto technique for image registration, especially in the domains of medical imaging [22] and remote sensing [3], which deal with a large number of modalities. The groundbreaking work for this was done by Viola and Wells [32] and Maes et al. [17] in their widely cited papers. A detailed survey of subsequent research on information-theoretic techniques in medical image registration is presented in the works of Pluim et al. [22] and Maes et al. [18]. A required component of all information-theoretic techniques in image registration is a good estimator

of the joint entropies of the images being registered. Most techniques employ plug-in entropy estimators, wherein the joint and marginal probability densities of the intensity values in the images are first estimated, and these quantities are then used to obtain the entropy. There also exist recent methods that define a new form of entropy using cumulative distributions instead of probability densities (see [25]). Furthermore, there also exist techniques that directly estimate the entropy, without estimating the probability density or distribution as an intermediate step [1]. Below, we present a bird's eye view of these techniques and their limitations. Subsequently, we introduce our method and bring out its salient merits.

The plug-in entropy estimators rely upon techniques for density estimation as a key first step. The most popular density estimator is the simple image histogram. The drawbacks of a histogram are that it yields discontinuous density estimates and requires an optimal choice of the bin width. Too small a bin width leads to noisy sparse density estimates (variance), whereas too large a bin width introduces oversmoothing (bias). Parzen windows have been widely employed as a differentiable density estimator for several applications in computer vision, including image registration [32]. Here, the problem of choosing an optimal bin width translates to the optimal choice of a kernel width and the kernel function itself. The choice of the kernel function is somewhat arbitrary [29], and furthermore, the implicit effect of the kernel choice on the structure of the image is an issue that has been widely ignored.¹ The kernel width parameter can be estimated by techniques such as maximum likelihood. Such methods, however, require complicated iterative optimizations and also a training and validation set. From an image registration standpoint, the joint density between the images undergoes a change in each iteration, which requires reestimation of the kernel width parameters. This step is an expensive iterative process with a complexity that is quadratic in the number of samples. Methods such as the fast Gauss transform [33] reduce this cost to some extent, but they require a prior clustering step. However, the fast Gauss transform is only an *approximation* to the true Parzen density estimate, and hence, one needs to analyze the behavior of the approximation error over the iterations if a gradient-based optimizer is used. Also, as per [29, Section 3.3.2], the ideal width value for minimizing the mean squared error between the true and estimated density is itself dependent upon the second derivative of the (unknown) true density. Yet another drawback of Parzen-window-based density estimators is the well-known “tail effect” in higher dimensions, due to which a large number of samples will fall in those regions where the Gaussian has a very low value [29]. Mixture models have been used for joint density estimation in registration [15], but they are quite inefficient and require a choice of the kernel function for the components (usually chosen to be Gaussian) and the number of components. This number again will change across the iterations of the registration process, as the images move with respect to one another. Wavelet-based density estimators have also been recently employed in image registration [9] and in conjunction with mutual information (MI) [21]. The problems with a wavelet-based method for density estimation include a choice of wavelet function, as well as the selection of the optimal number of levels or coefficients, which again requires iterative optimization.

Direct entropy estimators avoid the intermediate density estimation phase. While there exists a plethora of papers in this field (surveyed in [1]), the most popular entropy estimator used in image registration is the approximation of the Renyi entropy as the weight of a minimal spanning tree [16] or a K -nearest neighbor graph [5]. Note that the entropy used here is the Renyi entropy as opposed to the more popular Shannon entropy. The drawbacks of this approach include the computational cost in construction of the data structure in each step of

¹Parzen showed in [19] that $\sup |\hat{p}(x) - p(x)| \rightarrow 0$, where $\hat{p}$ and p refer to the estimated and true density, respectively. However, we stress that this is only an asymptotic result (as the number of samples $N_S \rightarrow \infty$) and, therefore, not directly linked to the nature of the image itself, for all practical purposes.

registration (the complexity whereof is quadratic in the number of samples drawn), the somewhat arbitrary choice of the α parameter for the Renyi entropy, and the lack of differentiability of the cost function. Some work has been done recently, however, to introduce differentiability in the cost function [27]. A merit of these techniques is the ease of estimation of entropies of high-dimensional feature vectors, with the cost scaling up just linearly with the dimensionality of the feature space.

Recently, a new form of the entropy defined on cumulative distributions and related cumulative entropic measures such as cross-cumulative residual entropy (CCRE) have been introduced in the literature on image registration [25]. The cumulative entropy and the CCRE measure have perfectly compatible discrete and continuous versions (quite unlike the Shannon entropy, though not unlike the Shannon MI) and are known to be noise resistant (as they are defined on cumulative distributions and not densities). Our method of density estimation can be easily extended to computing cumulative distributions and CCRE.

All of the techniques reviewed here are based on different principles but have one crucial common point: They treat the image as a set of pixels or samples, which inherently ignores the fact that these samples originate from an underlying continuous (or piecewise-continuous) signal. None of these techniques take into account the ordering between the given pixels of an image. As a result, all of these methods can be termed *sample based*. Furthermore, most of the aforementioned density estimators require a particular kernel, the choice of which is *extrinsic* to the image being analyzed and not necessarily linked even to the noise model. In this paper, we present an entirely different approach in which the density estimate is built directly from a continuous *image* representation (as opposed to an arbitrary kernel on the density). Our approach here is based on the earlier work presented in [24], the essence of which is to regard the marginal probability density as the area between two isocontours at infinitesimally close intensity values. A similar approach to density estimation has also been taken in the work of Kadir and Brady [13]. In our work, we have also presented a detailed derivation for the joint density between two or more images and also extended the work to the 3D case, besides testing it thoroughly on affine image registration for varying noise levels and quantization widths. Prior work on image registration using such image-based techniques includes [24], [23], [10], and [14]. The work in [10], however, reports results only on template matching with translations, whereas the main focus of [14] is on the estimation of densities in vanishingly small circular neighborhoods. The formulas derived are very specific to the shape of the neighborhood. Their paper [14] shows that local MI values in small neighborhoods are related to the values of the angles between the local gradient vectors in those neighborhoods. The focus of this method, however, is too local in nature, thereby ignoring the robustness that is an integral part of more global density estimates. There also exists some related work by Hadjidemetriou et al. [12] in the context of histogram-preserving locally continuous image transformations (the so-called Hamiltonian morphisms), which relates histograms to areas between isocontours. The main practical applications discussed in [12] are histograms under weak perspective and paraperspective projections of 3D textured models.

Note that our method, based on finding areas between isocontours, is significantly different from partial volume interpolation (PVI) [17], [30]. PVI uses a continuous image representation to build a joint probability table by assigning fractional votes to multiple intensity pairs when a digital image is warped during registration. The fractional votes are assigned typically using a bilinear or bicubic kernel function in cases of nonalignment with pixel grids after image warping. In essence, the density estimate in PVI still requires histogramming or Parzen windowing.

In this paper, we also present in detail the problems that lead to singularities in the probability density as estimated by the suggested procedure and also suggested principled modifications. The main merit of the proposed geometric technique is the fact that it sidesteps the parameter selection problem that affects other density estimators and also does not rely on any form of sampling. The accuracy of our techniques will always *upper bound* all sample-based methods if the image interpolant is known (see Section 4). In fact, the estimate obtained by all sample-based methods will converge to that yielded by our method only in the limit when the number of samples tends to infinity. Empirically, we demonstrate the robustness of our technique to noise and its superior performance in image registration. We conclude with a discussion and clarification of some properties of our method.

2 MARGINAL AND JOINT DENSITY ESTIMATION

In this section, we show the derivation of the probability density function (PDF) for the marginal and the joint density for a pair of 2D images. We point out practical issues and computational considerations, as well as outline the density derivations for the case of 3D images, as well as multiple images in 2D.

2.1 Estimating the Marginal Densities in 2D

Consider the 2D gray-scale image intensity to be a continuous scalar-valued function of the spatial variables, represented as $w = I(x,y)$. Let the total area of the image be denoted by A . Assume a location random variable $Z = (X,Y)$ with a uniform distribution over the image field of view (FOV). Further, assume a new random variable W , which is a transformation of the random variable Z , with the transformation given by the gray-scale image intensity function $W = I(X,Y)$. Then, the cumulative distribution of W at a certain intensity level α is equal to the ratio of the total area of all regions whose intensity is less than or equal to α to the total area of the image:

$$\Pr(W \leq \alpha) = \frac{1}{A} \iint_{I(x,y) \leq \alpha} dx dy. \quad (1)$$

Now, the probability density of W at α is the derivative of the cumulative distribution in (1). This is equal to the difference in the areas enclosed within two level curves that are separated by an intensity difference of $\Delta\alpha$ (or, equivalently, the area enclosed between two level curves of intensity α and $\alpha + \Delta\alpha$) per unit difference, as $\Delta\alpha \rightarrow 0$ (see Fig. 1). The formal expression for this is

$$p(\alpha) = \frac{1}{A} \lim_{\Delta\alpha \rightarrow 0} \frac{\iint_{I(x,y) \leq \alpha + \Delta\alpha} dx dy - \iint_{I(x,y) \leq \alpha} dx dy}{\Delta\alpha}. \quad (2)$$

Hence, we have

$$p(\alpha) = \frac{1}{A} \frac{d}{d\alpha} \iint_{I(x,y) \leq \alpha} dx dy. \quad (3)$$

We can now adopt a change of variables from the spatial coordinates (x,y) to $u(x,y)$ and $I(x,y)$, where u and I are the directions parallel and perpendicular to the level curve of intensity α , respectively. Observe that I points in the direction of the image gradient or the direction of maximum intensity change. Noting this fact, we now obtain the following:

$$p(\alpha) = \frac{1}{A} \int_{I(x,y)=\alpha} \left| \frac{\frac{\partial x}{\partial I} \frac{\partial y}{\partial I}}{\frac{\partial x}{\partial u} \frac{\partial y}{\partial u}} \right| du. \quad (4)$$

Note that in (4), $d\alpha$ and dI have “cancelled” each other out, as they both stand for intensity change. Upon a series of algebraic manipulations, we are now left with the following expression for $p(\alpha)$ (with a more detailed derivation to be found in the Appendix):

$$p(\alpha) = \frac{1}{A} \int_{I(x,y)=\alpha} \frac{du}{\sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}}. \quad (5)$$

From the above expression, one can make some important observations. Each point on a given level curve contributes a certain measure to the density at that intensity which is inversely proportional to the magnitude of the gradient at that point. In other words, in regions of high intensity gradient, the area between two level curves at nearby intensity levels would be small, as compared to that in regions of lower image gradient (see Fig. 1). When the gradient value at a point is zero (owing to the existence of a peak, a valley, a saddle point, or a flat region), the contribution to the density at that point tends to infinity. (The practical repercussions of this situation are discussed later on in the paper.) Last, the density at an intensity level can be estimated by traversing the level curve(s) at that intensity and integrating the reciprocal of the gradient magnitude. One can obtain an estimate of the density at several intensity levels (at intensity spacing of h from each other) across the entire intensity range of the image.

2.2 Estimating the Joint Density

Consider two images represented as continuous scalar-valued functions $w_1 = I_1(x,y)$ and $w_2 = I_2(x,y)$, whose overlap area is A . As before, assume a location random variable $Z = (X,Y)$ with a uniform distribution over the (overlap) FOV. Further, assume two new random variables W_1 and W_2 , which are transformations of the random variable Z , with the transformations given by the gray-scale image intensity functions $W_1 = I_1(X,Y)$ and $W_2 = I_2(X,Y)$. Let the set of all regions whose intensity in I_1 is less than or equal to α_1 and whose intensity in I_2 is less than or equal to α_2 be denoted by L . The cumulative distribution $\Pr(W_1 \leq \alpha_1, W_2 \leq \alpha_2)$ at intensity values (α_1, α_2) is equal to the ratio of the total area of L to the total overlap area A . The probability density $p(\alpha_1, \alpha_2)$ in this case is the second partial derivative of the cumulative distribution with respect to α_1 and α_2 . Consider a pair of level curves from I_1 having intensity values α_1 and $\alpha_1 + \Delta\alpha_1$ and another pair from I_2 having intensity α_2 and $\alpha_2 + \Delta\alpha_2$. Let us denote the region enclosed between the level curves of I_1 at α_1 and $\alpha_1 + \Delta\alpha_1$ as Q_1 and the region enclosed between the level curves of I_2 at α_2 and $\alpha_2 + \Delta\alpha_2$ as Q_2 . Then, $p(\alpha_1, \alpha_2)$ can geometrically be interpreted as the area of $Q_1 \cap Q_2$ divided by $\Delta\alpha_1 \Delta\alpha_2$ in the limit as $\Delta\alpha_1$ and $\Delta\alpha_2$ tend to zero. The regions Q_1 , Q_2 , and $Q_1 \cap Q_2$ (dark black region) are shown in Fig. 2a. Using a technique very similar to that shown in (2)–(4), we obtain the expression for the joint cumulative distribution as follows:

$$\Pr(W_1 \leq \alpha_1, W_2 \leq \alpha_2) = \frac{1}{A} \iint_L dx dy. \quad (6)$$

By doing a change of variables, we arrive at the following formula:

$$\Pr(W_1 \leq \alpha_1, W_2 \leq \alpha_2) = \frac{1}{A} \iint_L \left| \begin{array}{cc} \frac{\partial x}{\partial u_1} & \frac{\partial y}{\partial u_1} \\ \frac{\partial x}{\partial u_2} & \frac{\partial y}{\partial u_2} \end{array} \right| du_1 du_2. \quad (7)$$

Here, u_1 and u_2 represent directions along the corresponding level curves of the two images I_1 and I_2 . Taking the second partial derivative with respect to α_1 and α_2 , we get the expression for the joint density:

$$p(\alpha_1, \alpha_2) = \frac{1}{A} \frac{\partial^2}{\partial \alpha_1 \partial \alpha_2} \iint_L \left| \begin{array}{cc} \frac{\partial x}{\partial u_1} & \frac{\partial y}{\partial u_1} \\ \frac{\partial x}{\partial u_2} & \frac{\partial y}{\partial u_2} \end{array} \right| du_1 du_2. \quad (8)$$

It is important to note here again that the joint density in (8) may not exist because the cumulative may not be differentiable. Geometrically, this occurs if 1) both the images have locally constant intensity, 2) if only one image has locally constant intensity, or 3) if the level sets of the two images are locally parallel. In case 1, we have area measures, and in the other two cases, we have curve measures. These cases are described in detail in the following section, but for the moment, we shall ignore these degeneracies.

To obtain a complete expression for the PDF in terms of gradients, it would be highly intuitive to follow purely geometric reasoning. One can observe that the joint probability density $p(\alpha_1, \alpha_2)$ is the sum total of “contributions” at every intersection between the level curves of I_1 at α_1 and those of I_2 at α_2 . Each contribution is the area of parallelogram $ABCD$ (see Fig. 2b) at the level curve intersection, as the intensity differences $\Delta\alpha_1$ and $\Delta\alpha_2$ shrink to zero. (We consider a parallelogram here, because we are approximating the level curves locally as straight lines.) Let the coordinates of the point B be $(\tilde{x}, \tilde{y})$ and the magnitude of the gradient of I_1 and I_2 at this point be $g_1(\tilde{x}, \tilde{y})$ and $g_2(\tilde{x}, \tilde{y})$. Also, let $\theta(\tilde{x}, \tilde{y})$ be the *acute* angle between the gradients of the two images at B . Observe that the intensity difference between the two level curves of I_1 is $\Delta\alpha_1$. Then, using the definition of gradient, the perpendicular

distance between the two level curves of I_1 is given as $\frac{\Delta\alpha_1}{g_1(\tilde{x}, \tilde{y})}$. Looking at triangle CDE (wherein CE is perpendicular to the level curves), we can now deduce the length of CD (or, equivalently, that of AB). Similarly, we can also find the length of CB . The two expressions are given by

$$|AB| = \frac{\Delta\alpha_1}{g_1(\tilde{x}, \tilde{y}) \sin \theta(\tilde{x}, \tilde{y})}, |CB| = \frac{\Delta\alpha_2}{g_2(\tilde{x}, \tilde{y}) \sin \theta(\tilde{x}, \tilde{y})}. \quad (9)$$

Now, the area of the parallelogram is equal to $|AB||CB|\sin \theta(\tilde{x}, \tilde{y})$, which evaluates to

$\frac{\Delta\alpha_1 \Delta\alpha_2}{g_1(\tilde{x}, \tilde{y}) g_2(\tilde{x}, \tilde{y}) \sin \theta(\tilde{x}, \tilde{y})}$. With this, we finally obtain the following expression for the joint density:

$$p(\alpha_1, \alpha_2) = \frac{1}{A} \sum_C \frac{1}{g_1(x, y) g_2(x, y) \sin \theta(x, y)}, \quad (10)$$

where the set C represents the (countable) locus of all points where $I_1(x, y) = \alpha_1$ and $I_2(x, y) = \alpha_2$. It is easy to show through algebraic manipulations that (8) and (10) are equivalent formulations of the joint probability density $p(\alpha_1, \alpha_2)$. These results could also have been

derived purely by manipulation of Jacobians (as done while deriving marginal densities), and the derivation for the marginals could also have proceeded following geometric intuitions.

The formula derived above tallies beautifully with intuition in the following ways: First, the area of the parallelogram $ABCD$ (i.e., the joint density contribution) in regions of high gradient (in either or both image(s)) is smaller as compared to that in the case of regions with lower gradients. Second, the area of parallelogram $ABCD$ (i.e., the joint density contribution) is the least when the gradients of the two images are orthogonal and maximum when they are parallel or coincident (see Fig. 3a). In fact, the joint density tends to infinity in the case where either (or both) gradient(s) is (are) zero or when the two gradients align so that $\sin \theta$ is zero. The repercussions of this phenomenon are discussed in the following section.

2.3 From Densities to Distributions

In Sections 2.1 and 2.2, we observed the divergence of the marginal density in regions of zero gradient or of the joint density in regions where either (or both) image gradient(s) is (are) zero or when the gradients locally align. The gradient goes to zero in regions of the image that are flat in terms of intensity and also at peaks, valleys, and saddle points on the image surface. We can ignore the latter three cases as they are a finite number of points within a continuum. The probability contribution at a particular intensity in a flat region is proportional to the area of that flat region. Some ad hoc approaches could involve simply “weeding out” the flat regions altogether, but that would require the choice of sensitive thresholds. The key thing is to notice that in these regions, *the density does not exist, but the probability distribution does*. Therefore, we can switch entirely to probability distributions everywhere by introducing a nonzero lower bound on the “values” of $\Delta\alpha_1$ and $\Delta\alpha_2$. Effectively, this means that we always look at parallelograms representing the intersection between pairs of level curves from the two images, separated by a *nonzero* intensity difference, denoted as, say, h . Since these parallelograms have finite areas, we have circumvented the situation of choosing thresholds to prevent the values from becoming unbounded, and the probability at α_1, α_2 , denoted as $\hat{p}(\alpha_1, \alpha_2)$ is obtained from the areas of such parallelograms. We term this area-based method of density estimation as *AreaProb*. Later on in the paper, we shall show that the switch to distributions is principled and does not reduce our technique to standard histogramming in any manner whatsoever.

The notion of an image as a continuous entity is one of the pillars of our approach. We adopt a locally linear formulation in this paper, for the sake of simplicity, though the technical contributions of this paper are in no way tied to any specific interpolant. For each image grid point, we estimate the intensity values at its four neighbors within a horizontal or vertical distance of 0.5 pixels. We then divide each square defined by these neighbors into a pair of *triangles*. The intensities within each triangle can be represented as a planar patch, which is given by the equation $z_1 = A_1x + B_1y + C_1$ in I_1 . Isointensity lines at levels α_1 and $\alpha_1 + h$ within this triangle are represented by the equations $A_1x + B_1y + C_1 = \alpha_1$ and $A_1x + B_1y + C_1 = \alpha_1 + h$ (likewise for the isointensity lines of I_2 at intensities α_2 and $\alpha_2 + h$, within a triangle of corresponding location). The contribution from this triangle to the joint probability at (α_1, α_2) , i.e., $\hat{p}(\alpha_1, \alpha_2)$, is the area bounded by the two pairs of parallel lines, clipped against the body of the triangle itself, as shown in Fig. 7. In the case that the corresponding gradients from the two images are parallel (or coincident), they enclose an infinite area between them, which when clipped against the body of the triangle yields a closed polygon of finite area, as shown in Fig. 7. When both the gradients are zero (which can be considered to be a special case of gradients being parallel), the probability contribution is equal to the area of the entire triangle. In the case where the gradient of only one of the images is zero, the contribution is equal to the area enclosed between the parallel isointensity lines of the *other* image, clipped

against the body of the triangle (see Fig. 7). Observe that though we have to treat pathological regions specially (despite having switched to distributions), we now do not need to select thresholds, nor do we need to deal with a mixture of densities and distributions. The other major advantage is added robustness to noise, as we are now working with probabilities instead of their derivatives, i.e., densities.

The issue that now arises is how the value of h may be chosen. It should be noted that although there is no “optimal” h , our density estimate would convey more and more information as the value of h is reduced (in complete contrast to standard histogramming). In Fig. 5, we have shown plots of our joint density estimate and compared it to standard histograms for P equal to 16, 32, 64, and 128 bins in *each* image (i.e., 32^2 , 64^2 , etc., bins in the joint), which illustrate our point clearly. We found that the standard histograms had a far greater number of empty bins than our density estimator, for the same number of intensity levels. The corresponding marginal discrete distributions for the original retinogram image (Fig. 4) [31] for 16, 32, 64, and 128 bins are shown in Fig. 6.

2.4 Joint Density between Multiple Images in 2D

For the simultaneous registration of multiple ($d > 2$) images, the use of a single d -dimensional joint probability has been advocated in previous literature [2], [35]. Our joint probability derivation can be easily extended to the case of $d > 2$ images by using a similar geometric intuition to obtain the polygonal area between d intersecting pairs of level curves (see Fig. 3b for the case of $d = 3$ images). Note here that the d -dimensional joint distribution lies essentially in a 2D subspace, as we are dealing with 2D images. A naïve implementation of such a scheme has a complexity of $O(NP^d)$, where P is the number of intensity levels chosen for each image, and N is the size of each image. Interestingly, however, this

exponential cost can be sidestepped by first computing the at most $\binom{d(d-1)}{2}P^2$ points of intersection between pairs of level curves from all d images with one another, for every pixel. Second, a graph can be created, each of whose nodes is an intersection point. Nodes are linked by edges labeled with the image number (say, k th image) if they lie along the same isocontour of that image. In most cases, each node of the graph will have a degree of four (and in the unlikely case where level curves from all images are concurrent, the maximal degree of a node will be $2d$). Now, this is clearly a planar graph, and hence, by Euler’s formula, we have the number of (convex polygonal) faces

$\tilde{F} = \frac{d(d-1)}{2} * 4P^2 - \frac{d(d-1)}{2} P^2 + 2 = O(P^2 d^2)$, which is quadratic in the number of images. The areas of the polygonal faces are the contributions to the joint probability distribution. In a practical implementation, there is no requirement to even create the planar graph. Instead, we can implement a simple incremental face-splitting algorithm [7, Section 8.3]. In such an implementation, we create a list of faces $\mathcal{F}$, which is updated incrementally. To start with, $\mathcal{F}$ consists of just the triangular face constituting the three vertices of a chosen half-pixel in the image. Next, we consider a single level line l at a time and split into two any face in $\mathcal{F}$ that l intersects. This procedure is repeated for all level lines (separated by a discrete intensity spacing) of all the d images. The final output is a listing of all polygonal faces $\mathcal{F}$ created by incremental splitting, which can be created in just $O(FPd)$ time. The storage requirement can be made polynomial by observing that for d images, the number of unique intensity tuples will be at most FN in the worst case (as opposed to P^d). Hence, all intensity tuples can be efficiently stored and indexed using a hash table.

2.5 Extensions to 3D

When estimating the probability density from 3D images, the choice of an optimal smoothing parameter is a less critical issue, as a much larger number of samples are

available. However, at a theoretical level, this still remains a problem, which would worsen in the multiple-image case. In 3D, the marginal probability can be interpreted as the total volume sandwiched between two isosurfaces at neighboring intensity levels. The formula for the marginal density $p(\alpha)$ of a 3D image $w = I(x,y,z)$ is given as follows:

$$p(\alpha) = \frac{1}{V} \frac{d}{d\alpha} \iiint_{I(x,y,z) \leq \alpha} dx dy dz. \quad (11)$$

Here, V is the volume of the image $I(x,y,z)$. We can now adopt a change of variables from the spatial coordinates x, y , and z to $u_1(x,y,z)$, $u_2(x,y,z)$, and $I(x,y,z)$, where I is perpendicular to the level surface (i.e., parallel to the gradient), and u_1 and u_2 are mutually perpendicular directions parallel to the level surface. Noting this fact, we now obtain the following:

$$p(\alpha) = \frac{1}{V} \iiint_{I(x,y,z) = \alpha} \begin{vmatrix} \frac{\partial x}{\partial I} & \frac{\partial y}{\partial I} & \frac{\partial z}{\partial I} \\ \frac{\partial x}{\partial u_1} & \frac{\partial y}{\partial u_1} & \frac{\partial z}{\partial u_1} \\ \frac{\partial x}{\partial u_2} & \frac{\partial y}{\partial u_2} & \frac{\partial z}{\partial u_2} \end{vmatrix} du_1 du_2. \quad (12)$$

Upon a series of algebraic manipulations just as before, we are left with the following expression for $p(\alpha)$:

$$p(\alpha) = \frac{1}{V} \iiint_{I(x,y,z) = \alpha} \frac{du_1 du_2}{\sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 + \left(\frac{\partial I}{\partial z}\right)^2}}. \quad (13)$$

For the joint-density case, consider two 3D images represented as $w_1 = I_1(x,y,z)$ and $w_2 = I_2(x,y,z)$, whose overlap volume (the FOV) is V . The cumulative distribution $\Pr(W_1 \leq \alpha_1, W_2 \leq \alpha_2)$ at intensity values (α_1, α_2) is equal to the ratio of the total volume of all regions whose intensity in the first image is less than or equal to α_1 and whose intensity in the second image is less than or equal to α_2 to the total image volume. The probability density $p(\alpha_1, \alpha_2)$ is again the second partial derivative of the cumulative distribution. Consider two regions R_1 and R_2 , where R_1 is the region trapped between level surfaces of the first image at intensities α_1 and $\alpha_1 + \Delta\alpha_1$, and R_2 is defined analogously for the second image. The density is proportional to the volume of the intersection of R_1 and R_2 divided by $\Delta\alpha_1$ and $\Delta\alpha_2$ when the latter two tend to zero. It can be shown through some geometric manipulations that the area

of the base of the parallelepiped formed by the isosurfaces is given as $\frac{\Delta\alpha_1 \Delta\alpha_2}{|\vec{g}_1 \times \vec{g}_2|} = \frac{\Delta\alpha_1 \Delta\alpha_2}{|g_1 g_2 \sin(\theta)|}$, where $\vec{g}_1$ and $\vec{g}_2$ are the gradients of the two images, and θ is the angle between them. Let $\vec{h}$ be a vector that points in the direction of the height of the parallelepiped (parallel to the base normal, i.e., $\vec{g}_1 \times \vec{g}_2$ and $d\vec{h}$ be an infinitesimal step in that direction. Then, the probability density is given as follows:

$$\begin{aligned} p(\alpha_1, \alpha_2) &= \frac{1}{V} \frac{\partial^2}{\partial \alpha_1 \partial \alpha_2} \iiint_{V_s} dx dy dz \\ &= \frac{1}{V} \frac{\partial^2}{\partial \alpha_1 \partial \alpha_2} \iiint_{V_s} \frac{d\vec{u}_1 d\vec{u}_2 d\vec{h}}{|\vec{g}_1 \times \vec{g}_2|} = \frac{1}{V} \int_C \frac{d\vec{h}}{|\vec{g}_1 \times \vec{g}_2|}. \end{aligned} \quad (14)$$

In (14), $\vec{u}_1$ and $\vec{u}_2$ are directions parallel to the isosurfaces of the two images, and $\vec{h}$ is their cross-product (and parallel to the line of intersection of the individual planes), while C is the

3D space curve containing the points where I_1 and I_2 have values α_1 and α_2 , respectively, and $V_s \stackrel{\text{def}}{=} \{(x, y, z): I_1(x, y, z) \leq \alpha_1, I_2(x, y, z) \leq \alpha_2\}$.

2.6 Implementation Details for the 3D Case

The density formulation for the 3D case suffers from the same problem of divergence to infinity, as in the 2D case. Similar techniques can be employed, this time using level *surfaces* that are separated by finite intensity gaps. To trace the level surfaces, each cube-shaped voxel in the 3D image can be divided into 12 tetrahedra. The apex of each tetrahedron is located at the center of the voxel, and the base is formed by dividing one of the six square faces of the cube by one of the diagonals of that face (see Fig. 8a). Within each triangular face of each such tetrahedron, the intensity can be assumed to be a linear function of location. Note that the intensities in different faces of one and the same tetrahedron can thus be expressed by *different* functions, all of them linear. Hence, the isosurfaces at different intensity levels within a single tetrahedron are nonintersecting but not necessarily parallel. These level surfaces at any intensity within a single tetrahedron turn out to be either triangles or quadrilaterals in 3D. This interpolation scheme does have some bias in the choice of the diagonals that divide the individual square faces. A scheme that uses 24 tetrahedra with the apex at the center of the voxel and four tetrahedra based on every single face has no bias of this kind (see Fig. 8b). However, we still used the former (and faster) scheme as it is simpler and does not noticeably affect the results. Level surfaces are again traced at a finite number of intensity values, separated by equal intensity intervals. The marginal density contributions are obtained as the volumes of convex polyhedra trapped in between consecutive level surfaces clipped against the body of individual tetrahedra. The joint distribution contribution from each voxel is obtained by finding the volume of the convex polyhedron resulting from the intersection of corresponding convex polyhedra from the two images, clipped against the tetrahedra inside the voxel. We refer to this scheme of finding joint densities as *VolumeProb*.

2.7 Joint Densities by Counting Points and Measuring Lengths

For the specific case of registration of two images in 2D, we present another method of density estimation. This method, which was presented by us earlier in [23], is a biased estimator that does not assume a uniform distribution on location. In this technique, the total number of cooccurrences of intensities α_1 and α_2 from the two images, respectively, is obtained by counting the total number of intersections of the corresponding level curves. Each half-pixel can be examined to see whether level curves of the two images at intensities α_1 and α_2 can intersect within the half-pixel. This process is repeated for different (discrete) values from the two images (α_1 and α_2), separated by equal intervals and selected a priori (see Fig. 9). The cooccurrence counts are then normalized so as to yield a joint probability mass function (PMF). We denote this method as *2DPoint-Prob*. The marginals are obtained by summing up the joint PMF along the respective directions. This method also avoids the histogramming binning problem as one has the liberty to choose as many level curves as desired. However, it is a biased density estimator because more points are picked from regions with a high image gradient. This is because more level curves (at equispaced intensity levels) are packed together in such areas. It can also be regarded as a weighted version of the joint density estimator presented in Section 2.6, with each point weighted by the gradient magnitudes of the two images at that point, as well as the sine of the angle between them. Thus, the joint PMF by this method is given as

$$p(\alpha_1, \alpha_2) = \frac{\partial^2}{\partial \alpha_1 \partial \alpha_2} \frac{1}{K} \int \int_D g_1(x, y) g_2(x, y) \sin \theta(x, y) dx dy, \quad (15)$$

where D denotes the regions where

$$I_1(x, y) \leq \alpha_1, I_2(x, y) \leq \alpha_2,$$

and K is a normalization constant. This simplifies to the following:

$$p(\alpha_1, \alpha_2) = \frac{1}{K} \sum_C 1. \quad (16)$$

Hence, we have $p(\alpha_1, \alpha_2) = \frac{|C|}{K}$, where C is the (countable) set of points where $I_1(x, y) = \alpha_1$ and $I_2(x, y) = \alpha_2$. The marginal (biased) density estimates can be regarded as the lengths of the individual isocontours. With this notion in mind, the marginal density estimates are seen to have a close relation with the total variation of an image, which is given by $TV = \int_{I=\alpha} |\nabla I(x, y)| dx dy$ [26]. We clearly have $TV = \int_{I=\alpha} du$, by doing the same change of variables (from x, y to u, I) as in (4) and (5), thus giving us the length of the isocontours at any given intensity level.

In 3D, we consider the segments of intersection of two isosurfaces and calculate their lengths, which become the PMF contributions. We refer to this as *LengthProb* (see Fig. 10a).

Both *2DPointProb* and *LengthProb*, however, require us to ignore those regions in which level sets do not exist because the intensity function is flat or those regions where level sets from the two images are parallel. The case of flat regions in one or both images can be fixed to some extent by slight blurring of the image. The case of aligned gradients is trickier, especially if the two images are in complete registration. However, in the multimodality case or if the images are noisy/blurred, perfect registration is a rare occurrence, and hence, perfect alignment of level surfaces will rarely occur.

To summarize, in both these techniques, *location is treated as a random variable with a distribution that is not uniform but instead peaked at (biased toward) locations where specific features of the image itself (such as gradients) have large magnitudes or where gradient vectors from the two images are closer toward being perpendicular than parallel*. Such a bias toward high gradients is principled, as these are the more salient regions of the two images. Empirically, we have observed that both these density estimators work quite well on affine registration and that *LengthProb* is more than 10 times faster than *VolumeProb*. This is because the computation of segments of intersection of planar isosurfaces is much faster than computing polyhedron intersections. Joint PMF plots for histograms and *LengthProb* for 128 bins and 256 bins are shown in Fig. 11.

There exists one more major difference between *AreaProb* and *VolumeProb* on one hand and *LengthProb* or *2DPointProb* on the other. The former two can be easily extended to compute joint density between multiple images (needed for coregistration of multiple images using measures such as modified MI (MMI) [2]). All that is required is the intersection of multiple convex polyhedra in 3D or multiple convex polygons in 2D (see Section 2.4). However, *2DPointProb* is strictly applicable to the case of the joint PMF between exactly two images in 2D, as the problem of intersection of three or more level curves at *specific* (discrete) intensity levels is overconstrained. In 3D, *LengthProb* also deals with strictly two images only, but one can extend the *LengthProb* scheme to also compute the joint PMF between exactly three images. This can be done by making use of the fact that

three planar isosurfaces intersect in a point (excepting degenerate cases) (see Fig. 10b). The joint PMFs between the three images are then computed by counting point intersections. We shall name this method as *3DPoint Prob*. The differences between all the aforementioned methods, *AreaProb*, *2DPointProb*, *VolumeProb*, *LengthProb*, and *3DPointProb*, are summarized in Table 1 for quick reference. It should be noted that *2DPointProb*, *LengthProb*, and *3DPointProb* compute PMFs, whereas *AreaProb* and *VolumeProb* compute cumulative measures over finite intervals.

2.8 Image Entropy and Mutual Information

We are ultimately interested in using the estimated values of $p(a_1, a_2)$ to calculate (Shannon) JE and MI. A major concern is that in the limit as the binwidth $h \rightarrow 0$, the Shannon entropy does *not* approach the continuous entropy but becomes unbounded [6]. There are two ways to deal with this. First, a normalized version of the joint entropy (NJE) obtained by dividing the Shannon joint entropy (JE) by $\log P$ (where P is the number of bins) could be employed instead of the Shannon joint entropy. As $h \rightarrow 0$ and the Shannon entropy tends toward $+\infty$, NJE would still remain stable, owing to the division by $\log P$, which would also tend toward

$+\infty$ (in fact, NJE will have a maximal upper bound of $\frac{\log P^2}{\log P} = 2$, for a uniform joint distribution). Alternatively (and this is the more principled strategy), we observe that unlike the case with Shannon entropy, the continuous MI is indeed the limit of the discrete MI as $h \rightarrow 0$ (see [6] for the proof). Now, as P increases, we effectively obtain an increasingly better approximation to the continuous MI.

In the multiple-image case ($d > 2$), we avoid using a pairwise sum of MI values between different image pairs, because such a sum ignores the simultaneous joint overlap between multiple images. Instead, we can employ measures such as modified mutual information (MMI) [2], which is defined as the KL divergence between the d -way joint distribution and the product of the marginal distributions, or its normalized version (MNMI) obtained by dividing MMI by the joint entropy. The expressions for MI between two images and MMI for three images are given as follows:

$$MI(I_1, I_2) = H_1(I_1) + H_2(I_2) - H_{12}(I_1, I_2), \quad (17)$$

which can be explicitly written as

$$MI(I_1, I_2) = \sum_{j_1} \sum_{j_2} p(j_1, j_2) \log \frac{p(j_1, j_2)}{p(j_1)p(j_2)}, \quad (18)$$

where the summation indices j_1 and j_2 range over the sets of possibilities of I_1 and I_2 , respectively. For three images, we have

$$MMI(I_1, I_2, I_3) = H_1(I_1) + H_2(I_2) + H_3(I_3) - H_{123}(I_1, I_2, I_3), \quad (19)$$

which has the explicit form

$$MMI(I_1, I_2, I_3) = \sum_{j_1} \sum_{j_2} \sum_{j_3} p(j_1, j_2, j_3) \log \frac{p(j_1, j_2, j_3)}{p(j_1)p(j_2)p(j_3)}, \quad (20)$$

where the summation indices j_1, j_2 , and j_3 range over the sets of possibilities of I_1, I_2 , and I_3 , respectively. Though normalized MI (NMI) and MNMI are not compatible in the discrete and continuous formulations (unlike MI and MMI), in our experiments, we ignored this fact as we chose very specific intensity levels.

3 EXPERIMENTAL RESULTS

In this section, we describe our experimental results on the estimation of PDFs and a comparison between our area-based method for 2D images versus standard histogramming with subpixel sampling. Further, we present results for 1) the case of registration of two images in 2D, 2) the case of registration of multiple images in 2D, and 3) the case of registration of two images in 3D.

3.1 Area-Based PDFs versus Histograms with Several Subpixel Samples

The accuracy of the histogram estimate will no doubt approach the true PDF as the number of samples N_s (drawn from subpixel locations) tends to infinity. However, we wish to point out that our method implicitly and efficiently considers every point as a sample, thereby constructing the PDF directly, i.e., the accuracy of what we calculate with the area-based method will always be an upper bound on the accuracy yielded by *any* sample-based approach, *under the assumption that the true interpolant is known to us*. We show here an anecdotal example for the same, in which the number of histogram samples N_s is varied from 5,000 to 2×10^9 . The L_1 and L_2 norms of the difference between the joint PDF of two 90×109 images (downsampled MR-T1 and MR-T2 slices obtained from Brainweb [4]) as computed by our method and that obtained by the histogram method, as well as the Jensen-Shannon divergence (JSD) between the two joint PDFs, are plotted in Fig. 12 versus $\log N_s$. The number of bins used was 128×128 (i.e. $h = 128$). Visually, it was observed that the joint density surfaces begin to appear ever more similar as N_s increases. The timing values for the joint PDF computation are shown in Table 2, clearly showing the greater efficiency of our method.

3.2 Registration of Two Images in 2D

For this case, we took preregistered MR-T1 and MR-T2 slices from Brainweb [4], downsampled to size 122×146 (see Fig. 12), and created a 20-degree rotated version of the MR-T2 slice. To this rotated version, zero-mean Gaussian noise of different variances was added using the *imnoise* function of Matlab. The chosen variances were 0.01, 0.05, 0.1, 0.2, 0.5, 1, and 2. All of these variances are chosen for an intensity range between zero and one. To create the probability distributions, we chose bin counts of 16, 32, 64, and 128. For each combination of bin count and noise, a brute-force search was performed so as to optimally align the synthetically rotated noisy image with the original one, as determined by finding the maximum of MI or NMI between the two images. Six different techniques were used for MI estimation:

1. simple histograms with bilinear interpolation for image warping (referred to as “Simple Hist”),
2. our proposed method using isocontours (referred to as “Isocontours”),
3. histogramming with PVI (referred to as “PVI”),
4. histogramming with cubic spline interpolation (referred to as “Cubic”),
5. the method *2DPointProb* proposed in [23], and
6. simple histogramming with 10^6 samples taken from subpixel locations uniformly randomly followed by usual binning (referred to as “Hist Samples”).

These experiments were repeated for 30 noise trials at each noise standard deviation. For each method, the mean and the variance of the error (absolute difference between the predicted alignment and the ground-truth alignment) was measured (Fig. 13). The same experiments were also performed using a Parzen-window-based density estimator using a Gaussian kernel and $\sigma = 5$ (referred to as “Parzen”) over 30 trials. In each trial, 10,000 samples were chosen. Out of these, 5,000 were chosen as centers for the Gaussian kernel, and the rest were used for the sake of entropy computation. The error mean and variance was recorded (see Table 3).

The adjoining error plots (Fig. 13) show results for all these methods for all bins counts for noise levels of 0.05, 0.2, and 1. The accompanying trajectories (for all methods except histogramming with multiple subpixel samples) with MI for bin counts of 32 and 128 and noise levels 0.05, 0.2, and 1.00 are shown as well, for the sake of comparison, for one arbitrarily chosen noise trial (Fig. 14). From these figures, one can appreciate the superior resistance to noise shown by both our methods, even at very high noise levels, as evidenced both by the shape of the MI and NMI trajectories, as well as the height of the peaks in these trajectories. Amongst the other methods, we noticed that PVI is more stable than simple histogramming with either bilinear or cubic spline-based image warping. In general, the other methods perform better when the number of histogram bins is small, but even there, our method yields a smoother MI curve. However, as expected, noise does significantly lower the peak in the MI and NMI trajectories in the case of all methods including ours, due to the increase in joint entropy. Though histogramming with 10^6 subpixel samples performs well (as seen in Fig. 13, our method efficiently and directly (rather than asymptotically) approaches the true PDF, and hence, the true MI value, under the assumption that we have access to the true interpolant. Parzen windows with the chosen σ value of five gave good performance, comparable to our technique, but we wish to reemphasize that the choice of the parameter was arbitrary and the computation time was much more for Parzen windows.

All the aforementioned techniques were also tested on affine image registration (except for histogramming with multiple subpixel samples and Parzen windowing, which were found to be too slow). For the same image as in the previous experiment, an affine-warped version was created using the parameters $\theta = 30$ degrees, $t = -0.3$, $s = -0.3$, and $\phi = 0$. During our experiments, we performed a brute-force search on the 3D parameter space so as to find the transformation that optimally aligned the second image with the first one. The exact parameterization for the affine transformation is given in [34]. Results were collected for a total of 20 noise trials, and the average predicted parameters were recorded, as well as the variance of the predictions. For a low noise level of 0.01 or 0.05, we observed that all methods performed well for a quantization up to 64 bins. With 128 bins, all methods except the two we have proposed broke down, i.e., yielded a false optimum of θ around 38 degrees and s and t around 0.4. For higher noise levels, all methods except ours broke down at a quantization of just 64 bins. The *2DPointProb* technique retained its robustness until a noise level of 1, whereas the area-based technique still produced an optimum of $\theta = 28$ degrees, $s = -0.3$, $t = -0.4$ (which is very close to the ideal value). The area-based technique broke down only at an incredibly high noise level of 1.5 or 2. The average and standard deviation of the *estimate* of the parameters θ , s , and t , for 32 and 64 bins for all five methods and for noise levels 0.2 and 1.00 are presented in Tables 4 and 5. We also performed two-sided Kolmogorov-Smirnov tests [11] for the statistical significance on the absolute error between the true and estimated affine transformation parameters for 64 bins and a noise of variance 1. We found that the difference in the error values for MI, as computed using standard histogramming and our isocontour technique, was statistically significant, as ascertained at a level of 0.01.

We also performed experiments on determining the angle of rotation using larger images with varying levels of noise ($\sigma = 0.05, 0.2, 1$). The same Brainweb images, as mentioned before, were used, except that their original size of 183×219 was retained. For a bin count up to 128, all/most methods performed quite well (using a brute-force search) even under high noise. However with a large bin count (256 bins), the noise resistance of our method stood out. The results of this experiment with different methods and under varying noise are presented in Tables 6, 7, and 8.

3.3 Registration of Multiple Images in 2D

The images used were preregistered MR-PD, MR-T1, and MR-T2 slices (from Brainweb) of sizes 90×109 . The latter two were rotated by $\theta_1 = 20$ degrees and by $\theta_2 = 30$ degrees, respectively (see Fig. 15). For different noise levels and intensity quantizations, a set of experiments was performed to optimally align the latter two images with the former using MMI and MNMI as criteria. These criteria were calculated using our area-based method as well as simple histogramming with bilinear interpolation. The range of angles was from 1 to 40 degrees in steps of 1 degree. The estimated values of θ_1 and θ_2 are presented in Table 9.

3.4 Registration of Volume Data Sets

Experiments were performed on subvolumes of size $41 \times 41 \times 41$ from MR-PD and MR-T2 data sets from the Brainweb simulator [4]. The MR-PD portion was warped by 20 degrees about the Y and Z axes. A brute-force search (from 5 to 35 degrees in steps of 1 degree, with a joint PMF of 64×64 bins) was performed so as to optimally register the MR-T2 volume with the prewarped MR-PD volume. The PMF was computed both using *LengthProb* and using simple histogramming and used to compute the MI/NMI just as before. The computed values were also plotted against the two angles as indicated in Figs. 16a and 16b. As the plots indicate, both the techniques yielded the MI peak at the correct point in the θ_Y, θ_Z plane, i.e. at 20 degrees, 20 degrees. When the same experiments were run using *VolumeProb*, we observed that the joint PMF computation for the same intensity quantization was more than 10 times slower. Similar experiments were performed for registration of three volume data sets in 3D, namely, $41 \times 41 \times 41$ subvolumes of MR-PD, MR-T1, and MR-T2 data sets from Brainweb. The three data sets were warped through -2 , -21 , and -30 degrees around the x -axis. A brute-force search was performed so as to optimally register the latter two data sets with the former using MMI as the registration criterion. Joint PMFs of size $64 \times 64 \times 64$ were computed, and these were used to compute the MMI between the three images. The MMI peak occurred when the second data set was warped through $\theta_2 = 19$ degrees, and the third was warped through $\theta_3 = 28$ degrees, which is the correct optimum. The plots of the MI values calculated by simple histogramming and *3DPointProb* versus the two angles are shown in Figs. 16a and 16b, respectively.

The next experiment was designed to check the effect of zero-mean Gaussian noise on the accuracy of affine registration of the same data sets used in the first experiment, using histogramming and *LengthProb*. Additive Gaussian noise of variance σ^2 was added to the MR-PD volume. Then, the MR-PD volume was warped by a 4×4 affine transformation matrix (expressed in homogeneous coordinate notation) given as $A = SHR_zR_yR_xT$, where R_z , R_y , and R_x represent rotation matrices about the Z, y- and x-axes, respectively, H is a shear matrix, and S represents a diagonal scaling matrix whose diagonal elements are given by 2^{S_x} , 2^{S_y} , and 2^{S_z} . (A translation matrix T is included as well. For more information on this parameterization, please see [28].) The MR-T1 volume was then registered with the MR-PD volume using a coordinate descent on all parameters. The actual transformation parameters were chosen to be 7 degrees for all angles of rotation and shearing and 0.04 for s_x , s_y , and s_z . For a smaller number of bins (32), it was observed that both the methods gave good results under low noise, and histogramming occasionally performed better. Table 10 shows the

performance of histograms and *LengthProb* for 128 bins over 10 different noise trials. In summary, we observed that our method produced superior noise resistance as compared to histogramming when the number of bins was larger. To evaluate the performance on real data, we chose volumes from the Visible Human Data Set² (Male). We took subvolumes of MR-PD and MR-T1 volumes of size $101 \times 101 \times 41$ (slices 1,110-1,151). The two volumes were almost in complete registration, so we warped the former using an affine transformation matrix with 5 degrees for all angles of rotation and shearing and a value of 0.04 for s_x , s_y , and s_z , resulting in a matrix with a sum of absolute values of 3.6686. A coordinate descent algorithm for 12 parameters was executed on MI calculated using *LengthProb* so as to register the MR-T1 data set with the MR-PD data set, producing a registration error of 0.319 (see Fig. 17).

4 Discussion

We have presented a new density estimator that is essentially geometric in nature, using continuous image representations and treating the probability density as the area sandwiched between isocontours at intensity levels that are infinitesimally apart. We extended the idea to the case of joint density between two images, both in 2D and 3D, as also the case of multiple images in 2D. Empirically, we showed superior noise resistance on registration experiments involving rotations and affine transformations. Furthermore, we also suggested a faster biased alternative based on counting pixel intersections that performs well and extended the method to handle volume data sets. The relationship between our techniques and histogramming with multiple subpixel samples was also discussed. Here are a few clarifications about our technique in question/answer format:

1. **How does our method compare to histogramming on an upsampled image?** If an image is upsampled several times and histogramming is performed on it, there will be more samples for the histogram. At a theoretical level though, there is still the issue of not being able to relate the number of bins to the available number of samples. Furthermore, it is recommended that the rate of increase in the number of bins be less than the square root of the number of samples for computing the joint density between two images [35], [8]. If there are d images in all, the number of bins ought to be less than $\frac{1}{N^{d/2}}$, where N is the total number of pixels or samples to be taken [35], [8]. Consider that this criterion suggested that N samples were enough for a joint density between two images with χ bins. Suppose that we now wished to compute a joint density with χ bins for d images of the same size. This would require the images to be upsampled by a factor of at least $N^{d/2}$, which is exponential in the number of images. Our simple area-based method clearly avoids this problem.
2. **How does one choose the optimal number of bins or the optimal interpolant for our method?** One can choose as many bins as needed for the application, constrained only by availability of appropriate hardware (processor speed or memory). We chose a (piecewise) linear interpolant for the sake of simplicity, though in principle any other interpolant could be used. It is true that we are making an assumption on the continuity of the intensity function that may be violated in natural images. However, given a good enough resolution of the input image, interpolation across a discontinuity will have a negligible impact on the density as those discontinuities are essentially a measure-zero set. One could even incorporate an edge-preserving interpolant [20] by running an anisotropic diffusion

²This data set was obtained from the Visible Human Project (http://www.nlm.nih.gov/research/visible/getting_data.html).

to detect the discontinuities and then taking care not to interpolate across the two sides of an edge.

3. **What are the limitations of our technique?** As of now, our technique is not differentiable, which is important for nonrigid registration. Differentiability could be achieved by fitting (say) a spline to the obtained probability tables. However, this again requires smoothing the density estimate in a manner that is not tied to the image geometry. Hence, this goes against the philosophy of our approach. For practical or empirical reasons, however, there is no reason why one should not experiment with this.
4. **Other avenues for future research:** Currently, we do not have a closed-form expression for our density estimate. Expressing the marginal and joint densities solely in terms of the parameters of the chosen image interpolant is a challenging theoretical problem. We could also apply our density estimation scheme to images whose pixel values belong to a non-euclidean manifold, such as unit vectors or covariance matrices, or to data fields that are defined on non-euclidean surfaces.

APPENDIX

In this section, we derive the expression for the marginal density of the intensity of a single 2D image. We begin with (4) derived in Section 2.1:

$$p(\alpha) = \frac{1}{A} \int_{I(x,y)=\alpha} \left| \begin{array}{cc} \frac{\partial x}{\partial I} & \frac{\partial y}{\partial I} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \end{array} \right| du. \quad (21)$$

Consider the following two expressions that appear while performing a change of variables and applying the chain rule:

$$\begin{bmatrix} dx & dy \end{bmatrix} = \begin{bmatrix} dI & du \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial I} & \frac{\partial y}{\partial I} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \end{bmatrix}, \quad (22)$$

$$\begin{bmatrix} dI & du \end{bmatrix} = \begin{bmatrix} dx & dy \end{bmatrix} \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \\ \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{bmatrix} = \begin{bmatrix} dx & dy \end{bmatrix} \begin{bmatrix} I_x & u_x \\ I_y & u_y \end{bmatrix}. \quad (23)$$

Taking the inverse in the latter, we have

$$\begin{bmatrix} dx & dy \end{bmatrix} = \frac{1}{I_x u_y - I_y u_x} \begin{bmatrix} u_y & -u_x \\ -I_y & I_x \end{bmatrix} \begin{bmatrix} dI & du \end{bmatrix}. \quad (24)$$

Comparing the individual matrix coefficients, we obtain

$$\left| \begin{array}{cc} \frac{\partial x}{\partial I} & \frac{\partial y}{\partial I} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \end{array} \right| = \frac{I_x u_y - u_x I_y}{(I_x u_y - I_y u_x)^2} = \frac{1}{I_x u_y - I_y u_x}. \quad (25)$$

Now, clearly the unit vector $\vec{u}$ is perpendicular to $\vec{I}$, i.e., we have the following:

$$u_y = \frac{I_x}{\sqrt{I_x^2 + I_y^2}} \text{ and} \quad (26)$$

$$u_x = \frac{-I_y}{\sqrt{I_x^2 + I_y^2}}. \quad (27)$$

This finally gives us

$$\left| \begin{array}{cc} \frac{\partial x}{\partial I} & \frac{\partial y}{\partial I} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \end{array} \right| = \frac{1}{\sqrt{I_x^2 + I_y^2}}. \quad (28)$$

Hence, the expression in (5) for the marginal density (i.e., $p(\alpha) = \frac{1}{A} \int_{I(x,y)=\alpha} \frac{du}{\sqrt{I_x^2 + I_y^2}}$) follows.

Acknowledgments

This work is partially supported by US National Science Foundation grant IIS-0307712 and US National Institutes of Health grant R01NS046812. The authors would like to thank the anonymous reviewers for their thoughtful comments on this paper. The 2D isocontour-based density estimator reported in this paper is available under the terms of the GNU General Public License (GPL version 2) at <http://www.cise.ufl.edu/~anand/isomatch.html>.

REFERENCES

1. Beirlant J, Dudewicz E, Györfi L, van der Meulen EC. Nonparametric Entropy Estimation: An Overview. *Int'l J. Math. and Statistical Sciences* 1997 June;vol. 6(no. 1):17–39.
2. Boes, J.; Meyer, C. Multi-Variate Mutual Information for Registration; *Proc. Second Int'l Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI '99)*; 1999. p. 606-612.
3. Chen H, Arora M, Varshney P. Mutual Information-Based Image Registration for Remote Sensing Data. *J Remote Sensing* 2003;vol. 24(no. 18):3701–3706.
4. Collins DL, et al. Design and Construction of a Realistic Digital Brain Phantom. *IEEE Trans. Medical Imaging* 1998;vol. 17(no. 3):463–468.
5. Costa, J.; Hero, A. Entropic Graphs for Manifold Learning; *Proc. 37th Asilomar Conf. Signals, Systems and Computers*; 2003. p. 316-320.
6. Cover, T.; Thomas, J. *Elements of Information Theory*. Wiley-Interscience; 1991.
7. de Berg, M.; van Kreveld, M.; Overmars, M.; Schwarzkopf, O. *Computational Geometry: Algorithms and Applications*. Springer; 1997.
8. Devroye, L.; Györfi, L.; Lugosi, G. *A Probabilistic Theory of Pattern Recognition*. Springer; 1996.
9. Downie T, Silverman B. A Wavelet Mixture Approach to the Estimation of Image Deformation Functions. *Sankhya Series B* 2001;vol. 63:181–198.
10. Dowson, N.; Bowden, R.; Kadir, T. Image Template Matching Using Mutual Information and NP-Windows; *Proc. 18th Int'l Conf. Pattern Recognition (ICPR '06)*; 2006. p. 1186-1191.
11. Feller W. On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *The Annals of Math. Statistics* 1948;vol. 19(no. 2):177–189.
12. Hadjidemetriou E, Grossberg M, Nayar S. Histogram Preserving Image Transformations. *Int'l J. Computer Vision* 2001;vol. 45(no. 1):5–23.

13. Kadir, T.; Brady, M. Estimating Statistics in Arbitrary Regions of Interest; Proc. British Machine Vision Conf. (BMVC '05); 2005. p. 589-598.
14. Karaçali B. Information Theoretic Deformable Registration Using Local Image Information. Int'l J. Computer Vision 2007;vol. 72(no. 3):219-237.
15. Leventon, M.; Grimson, WEL. Multi-Modal Volume Registration Using Joint Intensity Distributions; Proc. First Int'l Conf. Medical Image Computing and Computer-Assisted Intervention (MIC-CAI '98); 1998. p. 1057-1066.
16. Ma, B.; Hero, A.; Gorman, J.; Michel, O. Image Registration with Minimum Spanning Tree Algorithm; Proc. IEEE Int'l Conf. Image Processing (ICIP '00); 2000. p. 481-484.
17. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality Image Registration by Maximization of Mutual Information. IEEE Trans. Medical Imaging 1997;vol. 16(no. 2):187-198.
18. Maes, F.; Vandermeulen, D.; Suetens, P. Medical Image Registration Using Mutual Information; Proc. IEEE; 2003. p. 1699-1722.
19. Parzen E. On Estimation of a Probability Density Function and Mode. Annals of Math. Statistics 1962;vol. 33:1065-1076.
20. Perona P, Malik J. Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Trans. Pattern Analysis and Machine Intelligence 1990 July;vol. 12(no. 7):629-639.
21. Peter A, Rangarajan A. Maximum Likelihood Wavelet Density Estimation with Applications to Image and Shape Matching. IEEE Trans. Image Processing 2008 Apr;vol. 17(no. 4):458-468.
22. Pluim J, Maintz J, Viergever M. Mutual Information Based Registration of Medical Images: A Survey. IEEE Trans. Medical Imaging 2003;vol. 22(no. 8):986-1004.
23. Rajwade, A.; Banerjee, A.; Rangarajan, A. Continuous Image Representations Avoid the Histogram Binning Problem in Mutual Information Based Image Registration; Proc. Third IEEE Int'l Symp. Biomedical Imaging (ISBI '06); 2006. p. 840-843.
24. Rajwade, A.; Banerjee, A.; Rangarajan, A. New Method of Probability Density Estimation with Application to Mutual Information Based Image Registration; Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '06); 2006. p. 1769-1776.
25. Rao M, Chen Y, Vemuri B, Wang F. Cumulative Residual Entropy: A New Measure of Information. IEEE Trans. Information Theory 2004;vol. 50(no. 6):1220-1228.
26. Rudin L, Osher S, Fatemi E. Nonlinear Total Variation Based Noise Removal Algorithms. Physica D 1992;vol. 60:259-268.
27. Sabuncu, M.; Ramadge, P. Gradient Based Optimization of an EMST Image Registration Function; Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP '05); 2005. p. 253-256.
28. Shekhar R, Zagrodsky V. Mutual Information-Based Rigid and Nonrigid Registration of Ultrasound Volumes. IEEE Trans. Medical Imaging 2002;vol. 21(no. 1):9-22.
29. Silverman, B. Density Estimation for Statistics and Data Analysis. Chapman and Hall; 1986.
30. Thévenaz P, Unser M. Optimization of Mutual Information for Multiresolution Image Registration. IEEE Trans. Image Processing 2000;vol. 9(no. 12):2083-2099.
31. Tina Is No Acronym (TINA) Image Database. UK: Univ. of Manchester and Univ. of Sheffield; 2008. <http://www.tina-vision.net/ilib.php>
32. Viola P, Wells WM. Alignment by Maximization of Mutual Information. Int'l J. Computer Vision 1997;vol. 24(no. 2):137-154.
33. Yang, C.; Duraiswami, R.; Gumerov, N.; Davis, L. Improved Fast Gauss Transform and Efficient Kernel Density Estimation; Proc. Ninth IEEE Int'l Conf. Computer Vision (ICCV '03); 2003. p. 464-471.
34. Zhang, J.; Rangarajan, A. Affine Image Registration Using a New Information Metric; Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '04); 2004. p. 848-855.
35. Zhang, J.; Rangarajan, A. Multimodality Image Registration Using an Extensible Information Metric; Proc. 19th Int'l Conf. Information Processing in Medical Imaging (IPMI '05); 2005. p. 725-737.

Biographies

Ajit Rajwade received the bachelor's degree in computer engineering from the University of Pune, India, and the master's degree in computer science from McGill University, Montreal. He is currently a PhD student in the Department of Computer and Information Science and Engineering, University of Florida, Gainesville. His research interests include computational geometry, image processing, computer vision, and medical imaging. He is a student member of the IEEE.

Arunava Banerjee received the MS and PhD degrees from the Department of Computer Science, Rutgers, the State University of New Jersey, in 1995 and 2001, respectively. He is currently an assistant professor in the Department of Computer and Information Science and Engineering, University of Florida, Gainesville. His research interests include computational neuroscience, computer vision, and algorithms. He is a member of the IEEE.

Anand Rangarajan received the BTech degree in electronics engineering from the Indian Institute of Technology, Madras, India, in 1984 and the PhD degree in electrical engineering from the University of Southern California in 1991. From 1990 to 1992, he was a postdoctoral associate in the Departments of Diagnostic Radiology and Computer Science, Yale University. From 1992 to 1995, he held a joint research faculty position in both departments. From 1995 to 2000, he was an assistant professor in the Image Processing and Analysis Group (IPAG), Departments of Diagnostic Radiology and Electrical Engineering, Yale University. He is now an associate professor in the Department of Computer and Information Science and Engineering (CISE), University of Florida. In 1992, he chaired a Neural Information Processing Systems (NIPS) Workshop entitled "Deterministic Annealing and Combinatorial Optimization," and in 1995, he cochaired a NIPS Workshop entitled "Statistical and Structural Models in Network Vision." He has served on the program committees of the International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 1997, 1999, 2001, and 2007), the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00, '01, '03, and '06), and the International Conference on Information Processing in Medical Imaging (IPMI '05 and '07). He has also cochaired the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA) in 2001 and EMMCVPR in 2003 and 2005 and was an area chair for the IEEE International Conference on Computer Vision (ICCV) in 2007. His current research interests are best summarized as the application of machine learning to

image analysis. He is also interested in the scientific study of consciousness and has been active in the Toward a Science of Consciousness Conferences (1996, 1998, 2000, 2002, and 2008) and the International Congress of Vedanta Conferences (2002, 2005, and 2007). He is a member of the IEEE.

Fig. 1.

$p(\alpha) \propto$ area between level curves at α and $\alpha + \Delta\alpha$ (i.e., region with red dots).

(a)

(b)

Fig. 2.
 (a) Intersection of level curves of I_1 and I_2 : $p(\alpha_1, \alpha_2) \propto$ area of dark black regions. (b)
 Parallelogram approximation: PDF contribution = area (ABCD)

Fig. 3.

(a) Area of parallelogram increases as angle between level curves decreases (left to right). Level curves of I_1 and I_2 are shown in red and blue lines, respectively. (b) Joint probability contribution in the case of three images.

(a)

(b)

Fig. 4.
A retinogram [31] and its rotated negative.

Fig. 5. Joint densities of the retinogram images computed by histograms (top two rows) and by our area-based method (bottom two rows) using 16, 32, 64, and 128 bins.

Fig. 6.

Marginal densities of the retinogram image computed by histograms (top two rows) and our area-based method (bottom two rows) using 16, 32, 64, and 128 bins (rowwise order).

Fig. 7.

(a) Probability contribution equal to area of parallelogram between level curves clipped against the triangle, i.e., half-pixel. (b) Case of parallel gradients. (c) Case when the gradient of one image is zero (blue level lines) and that of the other is nonzero (red level lines). In each case, probability contribution equals area of the dark black

(a)

(b)

Fig. 8.

(a) Splitting a voxel into 12 tetrahedra, two on each of the six faces of the voxel; (b) Splitting a voxel into 24 tetrahedra, four on each of the six faces of the voxel.

Fig. 9.
Counting level curve intersections within a given half-pixel.

Fig. 10. (a) Segment of intersection of planar isosurfaces from the two images. (b) Point of intersection of planar isosurfaces (each shown in a different color) from the three images.

Fig. 11. Joint probability plots. (a) Histograms, 128 bins. (b) Histograms, 256 bins. (c) *LengthProb*, 128 bins. (d) *LengthProb*, 256 bins.

Fig. 12. Plots of the difference between the true joint PDF as computed by the area-based method and the PDF computed by histogramming with N_s subpixel samples versus $\log N_s$ using (b) L_1 norm, (c) L_2 norm, and (d) JSD. The relevant images are in (a).

Fig. 13.

Graphs showing the average error A (i.e., absolute difference between the estimated and the true angle of rotation) and error standard deviation S with MI as the criterion for 16, 32, 64, and 128 bins (rowwise) (a), (b), (c), and (d) with a noise of 0.05, (e), (f), (g), and (h) with a noise of 0.2, and (i), (j), (k), and (l) with a noise of 1. Inside each subfigure, error bars are plotted for six different methods, in the following order: Simple Histogramming, Isocontours, PVI, Cubic, $2DPointProb$, and Histogramming with 10^6 samples. Error bars show the values of $A - S$, A , and $A + S$. If S is small, only the value of A is shown.

Fig. 14.

MI for 32 and 128 bins. (a) and (b) With a noise level of 0.05. (c) and (d) With a noise level of 0.2. (e) and (f) With a noise level of 1.0. In all plots, dark blue: isocontours, cyan: *2DPointProb*, black: cubic, red: simple histogramming, and green: PVI. (Note: These plots should be viewed in color.).

Fig. 15.

(a) MR-PD slice. (b) MR-T1 slice rotated by 20 degrees. (c) MR-T2 slice rotated by 30 degrees.

Fig. 16.

MI computed using (a) histogramming and (b) *LengthProb* (plotted versus θ_Y and θ_Z); MMI computed using (c) histogramming and (d) *3DPointProb* (plotted versus θ_2 and θ_3).

Fig. 17.

(a) Original PD image (left), warped T1 image (middle), and image overlap before registration (right). (b) PD image warped using predicted matrix (left), warped T1 image (middle), and image overlap after registration (right). (c) PD image warped using ideal matrix (left), warped T1 image (middle), and image overlap after registration in the ideal case (right).

TABLE 1

Comparison between Different Methods of Density Estimation with Respect to the Nature of Domain, Bias, Speed, and Geometric Nature of Density Contributions

Method	2D/3D	Density Contr.	Bias	No. of images
AreaProb	2D	Area	No	Any
VolumeProb	3D	Volume	No	Any
LengthProb	3D	Length	Yes	2 only
2DPointProb	2D	Point count	Yes	2 only
3DPointProb	3D	Point count	Yes	3 only

TABLE 2

Timing Values for Computation of Joint PDFs and L_1 Norm of Difference between PDF Computed by Sampling with That Computed Using Isocontours

Method	Time (secs.)	Diff. with isocontour PDF
Isocontours	5.1	0
Hist. 10^6 samples	1	0.0393
Hist. 10^7 samples	11	0.01265
Hist. 10^8 samples	106	0.0039
Hist. 5×10^8 samples	450	0.00176
Hist. 2×10^9 samples	1927	8.58×10^{-4}

Number of bins is 128×128 , size of images 122×146 .

TABLE 3

Average and Standard Deviation of Error in Degrees (Absolute Difference between True and Estimated Angle of Rotation) for MI Using Parzen Windows

Noise Variance	Avg. Error	Std. Dev. of Error
0.05	0.0667	0.44
0.2	0.33	0.8
1	3.6	3
2	4.7	12.51

The isocontour method consistently gave better results than Parzen windowing under higher noise.

TABLE 4

Average Value and Variance of Parameters θ , s , and t Predicted by Various Methods (32 and 64 Bins, Noise $\sigma = 0.2$)

Method	Bins	θ	s	t
MI Hist	32	30, 0	-0.3, 0	-0.3, 0
NMI Hist	32	30, 0	-0.3, 0	-0.3, 0
MI Iso	32	30, 0	-0.3, 0	-0.3, 0
NMI Iso	32	30, 0	-0.3, 0	-0.3, 0
MI PVI	32	30, 0	-0.3, 0	-0.3, 0
NMI PVI	32	30, 0	-0.3, 0	-0.3, 0
MI Spline	32	30.8,0.2	-0.3, 0	-0.3, 0
NMI Spline	32	30.6,0.7	-0.3, 0	-0.3, 0
MI 2DPt.	32	30, 0	-0.3, 0	-0.3, 0
NMI 2DPt.	32	30, 0	-0.3, 0	-0.3, 0
MI Hist	64	29.2,49.7	0.4,0	0.27, 0.07
NMI Hist	64	28.8,44.9	0.4,0	0.33, 0.04
MI Iso	64	30, 0	-0.3, 0	-0.3, 0
NMI Iso	64	30, 0	-0.3, 0	-0.3, 0
MIP VI	64	30, 0	-0.3, 0	-0.3, 0
NMI PVI	64	30, 0	-0.3, 0	-0.3, 0
MI Spline	64	24,21.5	0.4,0	0.33, 0.04
NMI Spline	64	24.3,20.9	0.4,0	0.33, 0.04
MI 2DPt.	64	30, 0	-0.3, 0	-0.3, 0
NMI 2DPt.	64	30,0	-0.3, 0	-0.3, 0

Ground truth: $\theta = 30$, $s = t = -0.3$.

TABLE 5

Average Value and Variance of Parameters θ , s and t Predicted by Various Methods (32 and 64 Bins, Noise $\sigma = 1$)

Method	Bins	θ	s	t
MI Hist	32	33.7, 18.1	0.4,0	0.13,0.08
NMI Hist	32	34.3, 15.9	0.4,0	0.13, 0.08
MI Iso	32	30,0.06	-0.3, 0	-0.3, 0
NMI Iso	32	30,0.06	-0.3, 0	-0.3, 0
MI PVI	32	28.1, 36.25	0.26, 0.08	0.19,0.1
NMI PVI	32	28.1, 36.25	0.3, 0.05	0.21,0.08
MI Spline	32	30.3,49.39	0.4,0	0.09,0.1
NMI Spline	32	31.2,48.02	0.4,0	0.05,0.1
MI 2DPt.	32	30.3,0.22	-0.3, 0	-0.3, 0
NMI 2DPt.	32	30.3,0.22	-0.3, 0	-0.3, 0
MI Hist	64	27.5, 44.65	0.4,0	0.25,0.08
NMI Hist	64	27,43.86	0.4,0	0.246, 0.08
MI Iso	64	30.5, 0.12	-0.27, 0.035	-0.28, 0.02
NMI Iso	64	31.2,0.1	-0.27, 0.058	-0.28, 0.02
MI PVI	64	26.2,36.96	0.4,0	0.038,0
NMI PVI	64	26.8,41.8	0.4,0	0.038,0
MI Spline	64	25.9,40.24	0.4,0	0.3, 0.06
NMI Spline	64	25.7,26.7	0.4,0	0.3, 0.06
MI 2DPt.	64	30.5, 0.25	-0.24, 0.0197	-0.23, 0.01
NMI 2DPt.	64	30.5, 0.25	-0.26, 0.0077	-0.22, 0.02

Ground truth: $\theta = 30$, $s = t = -0.3$.

TABLE 6

Average Error (Absolute Difference) and Variance in Measuring Angle of Rotation Using MI and NMI Calculated with Different Methods, Noise $\sigma = 0.05$

Method	128 bins	256 bins
MI Hist.	0,0	0.13,0.115
NMI Hist.	0,0	0.067,0.062
MI Iso.	0,0	0,0
NMI Iso.	0,0	0,0
MI PVI	0,0	0,0
NMI PVI	0,0	0,0
MI Spline	0,0	0.33,0.22
NMI Spline	0,0	0.33,0.22
MI 2DPt.	0,0	0,0
NMI 2DPt.	0,0	0,0

TABLE 7

Average Error (Absolute Difference) and Variance in Measuring Angle of Rotation Using MI and NMI Calculated with Different Methods, Noise $\sigma = 0.2$

Method	128 bins	256 bins
MI Hist.	0.07,0.196	0.2,0.293
NMI Hist.	0.07,0.196	0.13,0.25
MI Iso.	0,0	0,0
NMI Iso.	0,0	0,0
MI PVI	0,0	0,0
NMI PVI	0,0	0,0
MI Spline	2.77,10	4.77,10
NMI Spline	2.77,10	18,0.06
MI 2DPt.	0,0	0,0
NMI 2DPt.	0,0	0,0

TABLE 8

Average Error (Absolute Difference) and Variance in Measuring Angle of Rotation Using MI and NMI Calculated with Different Methods, Noise $\sigma = 1$

Method	128 bins	256 bins
MI Hist.	1.26,31	27.9,3.1
NMI Hist.	1.2,30	28,3.3
MI Iso.	0,0	0,0
NMI Iso.	0,0	0,0
MI PVI	0,0.26	26.9,14.3
NMI PVI	0,0.26	26.8,14.5
MI Spline	10,0.2	18,0.33
NMI Spline	9.8,0.15	18,0.06
MI 2DPt.	0.07,0.06	0.07,0.06
NMI 2DPt.	0.267,0.32	0.07,0.06

TABLE 9

Three-Image Case: Angles of Rotation Using MMI and MNMI Calculated with the Isocontour Method and Simple Histograms for Noise Variance $\sigma = 0.05, 0.1, 1$ (Ground Truth: 20 and 30 degrees)

Noise Variance	Method	32 bins	64 bins
0.05	MMI Hist.	21,30	22,31
0.05	MNMI Hist.	21,30	22,31
0.05	MMI Iso.	20,30	20,30
0.05	MNMI Iso.	20,30	20,30
0.2	MMI Hist.	15,31	40,8
0.2	MNMI Hist.	15,31	40,8
0.2	MMI Iso.	22,29	20,30
0.2	MNMI Iso.	22,29	20,30
1	MMI Hist.	40,9	38,4
1	MNMI Hist.	40,9	34,4
1	MMI Iso.	22,30	35,23
1	MNMI Iso.	22,30	40,3

TABLE 10Error (Average, Standard Deviation) Validated over 10 Trials with *LengthProb* and Histograms for 128 Bins

Noise Level	Error with <i>LengthProb</i>	Error with histograms
0	0.09, 0.02	0.088, 0.009
$\sqrt{50R}$	0.135, 0.029	0.306, 0.08
$\sqrt{100R}$	0.5, 0.36	1.47, 0.646
$\sqrt{150R}$	0.56, 0.402	1.945, 0.56

R refers to the intensity range of the image.