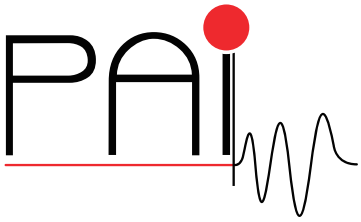
Research Network FWF S105

Photoacoustic Imaging in Medicine and Biology

http://pai.uibk.ac.at



Distance Measures and Applications to Multi-Modal Variational Imaging

Christiane Pöschl, Otmar Scherzer

March 2010

PAI Report No. 23

Der Wissenschaftsfonds.

Distance Measures and Applications to Multi-Modal Variational Imaging

Christiane Pöschl * Otmar Scherzer †‡

March 23, 2010

Contents

1 Introduction

Today **imaging** is rapidly improving by increased specificity and sensitivity of measurement devices. However, even more diagnostic information can be gained by combination of data recorded with different imaging systems.

In particular in medicine information of different modalities is used for diagnosis. From the various imaging technologies used in medicine we mention exemplary **positron emission tomography** (PET), **single photon emission computed tomography** (SPECT), **magnetic resonance imaging** (MRI), **magnetic resonance spectroscopy** (MRS), and **ultrasound**. Soft tissue can be well visualized in magnetic resonance scans, while bone structures is more easily discernible by X-ray imaging.

Image registration is an appropriate tool to align the information gained from different modalities. Thereby it is necessary to use similarity measures that are able to compare images

*Dept. of Information & Communication Technologies, Universitat Pompeu Fabra, C/ Tánger 122-140, 08018 Barcelona, Spain

†Computational Science Center, University of Vienna, Nordbergstr. 15, 1090 Vienna, Austria

‡Radon Institute of Computational and Applied Mathematics, Austrian Academy of Sciences, Altenberger Str. 69, 4040 Linz, Austria

of different modalities such that in a post processing step the data can be fused and relevant information can be aligned.

The main challenge for computer assisted comparison of images from different modalities is to define an appropriate distance measure between the images from different modalities.

Similarity measures of images can be categorized as follows:

1. Pixel wise comparison of intensities.
2. A morphological measure defines the distance between images by the distance between their level sets.
3. Measures based on the images gray value distributions.

In the following we review distance measures for images according to the above catalog.

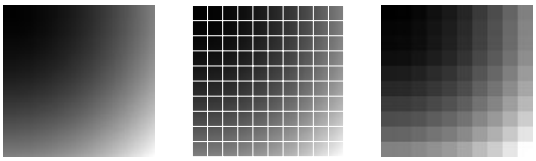
We use the notation Ω for the squared domain $(0,1)^2$. Images are simultaneously considered as matrices or functions on Ω: A **discrete image** is an $N \times N$-matrix $U \in \{0,\dots,255\}^{N\times N}$. Each of the entries of the matrix represents an intensity value at a pixel. Therewith is associated a piecewise constant function

$$u_N(x) = \sum_{i=1}^{N}\sum_{j=1}^{N} U^{ij}\chi_{\Omega_{ij}}(x)\,, \tag{1}$$

where

$$\Omega_{ij} := \left(\frac{i-1}{N},\frac{i}{N}\right)\times\left(\frac{j-1}{N},\frac{j}{N}\right) \text{ for } 1\le i,j\le N\,,$$

and $\chi_{\Omega_{ij}}$ is the characteristic function of Ω_{ij}. In the context of image processing U^{ij} denotes the **pixel intensity** at the **pixel** $\chi_{\Omega_{ij}}$. A **continuous** image is a function $u:\Omega\to\mathbb{R}$.



We emphasize that the measures for comparing images, presented below, can be applied in a straight forward way to higher dimensional domains, e.g. voxel data. However, here, for the sake of simplicity of notation and readability we restrict attention to a two dimensional squared domain Ω. Even more, we restrict attention to intensity data, and do not consider vector valued data, such as color images or tensor data.

2 Distance Measures

In the following we review distance measures for comparing discrete and continuous images. We review the standard and a morphological distance measure, both of them are deterministic. Moreover, based on the idea to consider images as random variable, we consider in the last two subsections two statistical approaches.

2.1 Deterministic Pixel Measure

The most widely used distance measures for discrete and continuous images are the l^p, L^p distance measures, consequently. There, two discrete images U_1 and U_2 are similar, if

$$\|U_1 - U_2\|_p := \frac{1}{p} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| U_1^{ij} - U_2^{ij} \right|^p \;, \quad 1 \le p < \infty \,,$$
$$\|U_2 - U_2\|_\infty := \sup_{i,j=1,\dots,N} \left| U_1^{ij} - U_2^{ij} \right| \;, \quad p = \infty \,,$$

respectively, is small. Two continuous images $u_1, u_2 : \Omega \to \mathbb{R}$ are similar if

$$\|u_1 - u_2\|_p := \frac{1}{p} \int_\Omega |u_1(x) - u_2(x)|^p \, dx \quad 1 \le p < \infty \,,$$
$$\|u_2 - u_2\|_\infty := \text{ess sup}_{x,y} \, |u_1(x) - u_2(x)| \;, \quad p = \infty \,,$$

is small. Here ess sup denotes the essential supremum.

2.2 Morphological Measures

In this subsection we consider continuous images $u_i : \Omega \to [0, 255]$, $i = 1, 2$. u_1 and u_2 are **morphologically equivalent**, if there exists a one-to-one gray value transformation $\beta : [0, 255] \to [0, 255]$, such that

$$\beta \circ u_1 = u_2 \;.$$

Level sets of a continuous function u are defined as

$$\Omega_t(u) := \{x \in \Omega : u(x) = t\} \;.$$

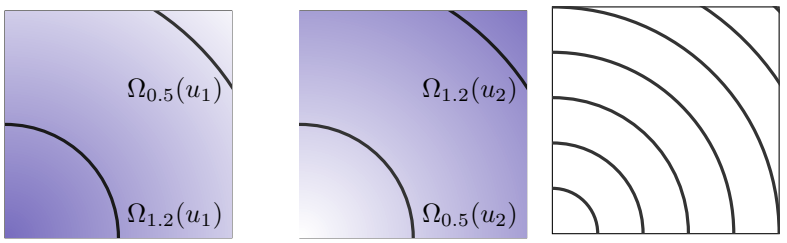


Figure 1: The gray values of the images are completely different, but the images u_1, u_2 have the same morphology.

The level sets $\Omega_{\mathbb{R}}(u) := \{\Omega_t(u) : t \in [0, 255]\}$ form the objects of an image that remain invariant under gray value transforms. The **normal field (Gauss map)** is given by the normals to the level lines, and can be written as

$$\begin{array}{rccl} \mathbf{n}(u): & \Omega & \to & \mathbb{R}^d \\ & x & \mapsto & \begin{cases} 0 & \text{if } \nabla u(x) = 0 \\ \frac{\nabla u(x)}{\|\nabla u(x)\|} & \text{else.} \end{cases} \end{array}$$

Droske & Rumpf [7] consider images as similar, if intensity changes occur at the same locations. Therefore the compare the normal fields of the images with the similarity measure

$$\mathcal{S}_g(u_1, u_2) = \int_\Omega g(\mathbf{n}(u_1)(x), \mathbf{n}(u_2)(x))\, dx\ , \tag{2}$$

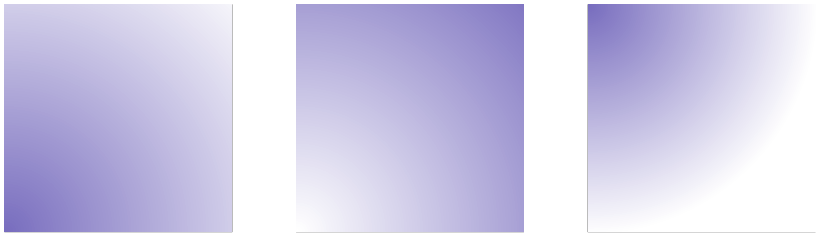
where they choose the function $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ appropriately. The vectors $\mathbf{n}(u_1)(x)$, $\mathbf{n}(u_2)(x)$ form an angle which is minimal if the images are morphologically equivalent. Therefore an appropriate choice of the function g is an increasing function of the minimal angle between v_1, v_2 and $v_1, -v_2$. For instance setting g to be the cross or the negative dot product, we obtain:

$$\mathcal{S}_\times(u_1, u_2) = \frac{1}{2}\int_\Omega |\mathbf{n}(u_1)(x) \times \mathbf{n}(u_2)(x)|^2\ dx$$
$$\mathcal{S}_\circ(u_1, u_2) = \frac{1}{2}\int_\Omega (1 - \mathbf{n}(u_1)(x) \cdot \mathbf{n}(u_2)(x))^2\ dx\ .$$

(The vectors $\mathbf{n}$ have to be embedded in $\mathbb{R}^3$ in order to calculate the crossproduct).

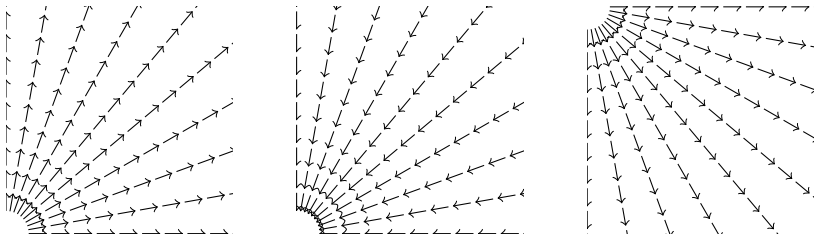
Example 1. *Consider the following scaled images $u_i : [0,1]^2 \to [0,1]$,*

$$u_1(x) = x_1x_2, \qquad u_2(x) = 1 - x_1x_2, \qquad u_3(x) = (1 - x_1)x_2,$$



with gradients

$$\nabla u_1(x) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} \qquad \nabla u_2(x) = \begin{pmatrix} -x_2 \\ -x_1 \end{pmatrix} \qquad \nabla u_3(x) = \begin{pmatrix} -x_2 \\ 1 - x_1 \end{pmatrix}\ .$$



With $g(u, v) := \frac{1}{2}\,|u_1v_2 - u_2v_1|$, the functional $\mathcal{S}_g$ defined in (2) attains the following values for the particular images:

$$\mathcal{S}_g(u_1, u_2) = \frac{1}{2}\int_\Omega |-x_2x_1 + x_2x_1|\, dx = 0$$
$$\mathcal{S}_g(u_2, u_3) = \frac{1}{2}\int_\Omega |x_2x_1 + x_2x_1|\, dx = \frac{1}{4}$$
$$\mathcal{S}_g(u_3, u_1) = \frac{1}{2}\int_\Omega |-x_2x_1 - (1 - x_1)x_2|\, dx = \frac{1}{4}\ .$$

The similarity measure indicates that u_1 and u_2 are morphological identical.

The normalized gradient field is set valued in regions where the function is constant. Therefore, the numerical evaluation of the gradient field is highly unstable. To overcome this drawback Haber & Modersitzki [15] suggested to use regularized normal gradient fields:

$$\begin{array}{rccc} \mathbf{n}_\varepsilon(u): & \Omega & \to & \mathbb{R}^d \\ & x & \mapsto & \frac{\nabla u(x)}{\|\nabla u(x)\|_\varepsilon} \end{array}$$

where, $\|v\|_\varepsilon := \sqrt{v^T v + \varepsilon^2}$ for every $v \in \mathbb{R}^d$. The parameter ε is connected to the estimated noise level in the image. In regions where ε is much larger than the gradient, the regularized normalized fields $\mathbf{n}_\varepsilon(u)$ are almost zero and therefore do not have a significant effect of the measures $\mathcal{S}_\times$ or $\mathcal{S}_\circ$, respectively. However, in regions where ε is much smaller than the gradients, the regularized normal fields are close to the non-regularized ones.

2.3 Statistical Distance Measures

Several distance measures for pairs of images can be motivated from statistics by considering the images as random variables. In the following we analyze discrete images from a statistical point of view. For this purpose we need some elementary statistical definitions. Applications of the following measures are mentioned in Section 4.

Correlation Coefficient:

$$\overline{U} := \frac{1}{N^2} \sum_{i,j=1}^{N} U^{ij} \text{ and } \mathrm{Var}(U) = \sum_{i,j=1}^{N} (U^{ij} - \overline{U})^2$$

denote the **mean intensity** and **variance** of the discrete image U.

$$\mathrm{Cov}(U_1, U_2) = \sum_{i=1}^{N} \sum_{j=1}^{N} (U_1^{ij} - \overline{U}_1)(U_2^{ij} - \overline{U}_2)$$

denotes the **covariance** of two images U_1 and U_2, and the **correlation coefficient** is defined by

$$\rho(U_1, U_2) = \frac{\mathrm{Cov}(U_1, U_2)}{\sqrt{\mathrm{Var}(U_1)\mathrm{Var}(U_2)}} .$$

The correlation coefficient is a measure of **linear dependence** of two images. The range of the correlation coefficient is $[-1, 1]$, and if $|\rho(U_1, U_2)|$ is close to one then it indicates that U_1 and U_2 are linearly dependent.

Correlation Ratio: In statistics, the correlation ratio is used to measure the relationship between the statistical dispersion within individual categories and the dispersion across the whole population. The **correlation ratio** is defined by

$$\eta(U_2 \mid U_1) = \frac{\mathrm{Var}\big(E(U_2 \mid U_1)\big)}{\mathrm{Var}(U_2)} ,$$

where $E(U_2 \mid U_1)$ is the conditional expectation of U_2 subject to U_1.

To put this into the context of image comparison let

$$\Omega_t(U_1) := \Big\{ (i,j) \mid U_1^{ij} = t \Big\}$$

be the discrete level set of intensity $t \in \{0, \ldots, 255\}$. Then the expected value of U_2 on the t-th level set of U_1 is given by

$$E(U_2 \mid U_1 = t) := \frac{1}{\#\left(\Omega_t(U_1)\right)} \sum_{\Omega_t(U_1)} U_2^{ij} \,,$$

where $\#\left(\Omega_t\left(U_1\right)\right)$ denotes the number of pixels in U_1 with gray-value t. Moreover, the according conditional variance is defined by

$$V(U_2 \mid U_1 = t) = \frac{1}{\#\left(\Omega_t\left(U_1\right)\right))} \sum_{\Omega_t(U_1)} \left(U_2^{ij} - E(U_2 \mid U_1 = t)\right)^2 \,.$$

The function

$$\begin{array}{rccc} H(U_1): & \{0, \ldots, 255\} & \to & \mathbb{N} \\ & t & \mapsto & \#\left(\Omega_t\left(U_1\right)\right) \end{array}$$

is called the **discrete histogram** of U_1.

The correlation ratio is non symmetric, that is $\eta(Y \mid X) \neq \eta(X \mid Y)$, and takes values between $[0, 1]$. It is a measure of **(non)-linear dependence** between two images. If $U_1 = U_2$, then the correlation ratio is maximal.

Variance of Intensity Ratio, Ratio Image Uniformity: This measure is based on the definition of similarity that two images are similar, if the factor $R^{ij}(U_1, U_2) = U_1^{ij}/U_2^{ij}$ has a small variance. The **ratio image uniformity** (or normalized variance of the intensity ratio) can be calculated by

$$RIU(U_1, U_2) = \frac{Var(R)}{\overline{R}} \,.$$

It is not symmetric.

Example 2. *Consider the discrete images U_1, U_2, U_3 in Figure 2. Table 2 shows a comparison of the different similarity measures. The variance of the intensity ratio is insignificant and therefore cannot be used to determine similarities. The correlation ratio is maximal for the pairing U_1, U_2 and in fact there is a functional dependence of the intensity values of U_1 and U_2. However, the dependence of the intensity values of U_1 and U_2 is nonlinear, hence the absolute value of the correlation coefficient (measure of linear dependence) is close to one, but not identical to one.*

4	4	5	6	6	7	7	7	5	3	3	1	1	9	6	4	4	9
3	3	4	4	5	6	8	8	7	7	5	3	2	10	8	9	3	5
2	2	3	4	4	6	9	9	8	7	7	3	1	8	9	3	3	10
1	2	2	3	4	5	10	9	9	8	7	5	8	3	5	2	10	4
1	1	2	2	3	4	10	10	9	9	8	7	9	10	3	8	6	10
1	1	1	2	3	4	10	10	10	9	8	7	9	8	10	6	1	6
U_1						U_2						U_3					

Figure 2: Images for Examples 2 and 16. Note that there is a dependence between U_2 and U_2: $U_2 \sim 11 - (U_1)^3$.

	U_1, U_2	U_2, U_1	U_2, U_3	U_3, U_2	U_3, U_1	U_1, U_3
Correlation Coefficient	**-0.98**	**-0.98**	0.10	0.10	-0.14	-0.14
Correlation Ratio	**1.00**	**1.00**	0.28	0.32	0.29	0.64
Variance of Intensity Ratio	1.91	2.87	2.25	1.92	**3.06**	0.83

Table 1: Comparison of the different pixel based similarity measures. The images U_1, U_2 are related in a nonlinear way, this is reflected in a correlation ratio of 1. Since this dependence is nonlinear, the absolute value of the correlation coefficient is close to, but smaller than one. We see that the variance of intensity ratio, is not symmetric and not significant to make a statement on a correlation between the images.

2.4 Statistical Distance Measures (Density Based)

In general, if two images are of different modality, the l^p, L^p-distances, respectively, of the intensities are weak.

In order to use statistical measures, we consider an image as a random variable. The basic terminology of random variables is as follows:

Definition 3. *A **continuous random variable** is a real valued function $X : \Omega^S \to \mathbb{R}$ defined on the **sample space** Ω^s. For a **sample** x, $X(x)$ is called **observation**.*

Remark 4 (Images as Random Variables)**.** *When we consider an image $u : \Omega \to \mathbb{R}$ a continuous random variable, the sample space is Ω. For a sample $x \in \Omega$ the observation $u(x)$ is the intensity of u at x.*

By regarding the intensity values of an image as an observation of a random processes allows us to compare images via their intrinsic probability densities. Since the density **cannot** be calculated directly, it has to be estimated. This is outlined in Subsection 2.4.1, below. There exists a variety of distance measures for probability densities (see for instance [30]). In particular, we review f-divergences in Subsection 2.4.2 and explain how to use the f-information as a image similarity measure in Subsection 2.4.3.

2.4.1 Density Estimation

This section reviews the problem of **density estimation**, which is the construction of an estimate of the density function from the observed data.

Definition 5. *Let $X : \Omega^S \to \mathbb{R}$ be a random variable, i.e. a function mapping the (measurable) sample space Ω^S of a random process to the real numbers.*

*The **cumulated probability density function** of X is defined by*

$$P(t) := \frac{1}{\text{meas}(\Omega^S)} meas\{x : X(x) \leq t\} \quad t \in \mathbb{R} \ .$$

*The **probability density** function p is the derivative of P.*

*The **joint cumulated probability density function** of two random variables X_1, X_2 is defined by*

$$\hat{P}(t_1, t_2) := \frac{1}{\text{meas}(\Omega^S)^2} meas\{(x_1, x_2) : X_1(x_1) \leq t_1, X_2(x_2) \leq t_2\} \quad t_1, t_2 \in \mathbb{R} \ .$$

*The **joint probability density** function $\hat{p}$ satisfies*

$$\hat{P}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \hat{p}(s_1, s_2) ds_1 \, ds_2 \ .$$

Remark 6. *When we consider an image $u : \Omega \to \mathbb{R}$ a random variable with sample space Ω, we write $p(u)(t)$ for the probability density function of the image u. For the joint probability of two images u_1 and u_2 we write $\hat{p}(u_1, u_2)(t_1, t_2)$, to emphasize, as above, that the images are considered as random variables.*

The terminology of Definition 5 is clarified by the following one dimensional example:

Example 7. *Let*

$$u : \Omega := [0,1] \to [0,255] \,.$$
$$x \to 255x^2$$

The cumulated probability density function $P : [0,255] \to [0,1]$ is obtained by integration:

$$P(t) := meas\left\{x : 255\, x^2 \leq t\right\} = \sqrt{\frac{t}{255}} \,.$$

The probability density function of u is given by the derivative of P, which is

$$p(u)(t) = \frac{1}{2\sqrt{255}} \frac{1}{\sqrt{t}} \,.$$

In image processing it is common to view the discrete image U (or u_N as in (1)) as an approximation of an image u. We aim for the probability density function of u, which is approximated via kernel density estimation using the available information of u, which is U. A kernel histogram is the normalized probability density function according to the discretized image U, where for each pixel a **kernel function** (see (3) below) is superimposed. Kernel functions depend on a parameter, which can be used to control the smoothness of the kernel histogram.

We first give a general definition of kernel density estimation:

Definition 8 (Kernel Density Estimation)**.** *Let $t_1, t_2, \ldots, t_M$ be a sample of M independent observations from a measurable real random variable X with probability density function p. The **kernel density approximation** at t is given by*

$$p_\sigma(t) = \frac{1}{M} \sum_{i=1}^{M} k_\sigma\left(t - t_i\right) \,, \quad t \in [0,255]$$

*where k_σ is the kernel function with bandwidth σ. p_σ is called **kernel density approximation** with parameter σ.*

*Let $t_1, t_2, \ldots, t_M$ and $s_1, s_2, \ldots, s_M$ be samples of M independent observations from measurable real random variables X_1, X_2 with with joint probability density function $\hat{p}$, then the **joint kernel density approximation** of $\hat{p}$ is given by*

$$\hat{p}_\sigma(s,t) = \frac{1}{M} \sum_{i=1}^{M} K_\sigma(s - s_i, t - t_i) \,,$$

where $K_\sigma(s,t) = k_\sigma(s) k_\sigma(t)$ is the two-dimensional kernel function.

Remark 9 (Kernel Density Estimation of an Image)**.** *Let u be a continuous image, which is identified with a random variable. Moreover, let U be $N \times N$ samples of u. In analogy to Definition 8 we denote the **kernel density estimation** based on the discrete image U, by*

$$p_\sigma(t) = \frac{1}{N^2} \sum_{i,j=1}^{N} k_\sigma\left(t - U^{ij}\right) \,.$$

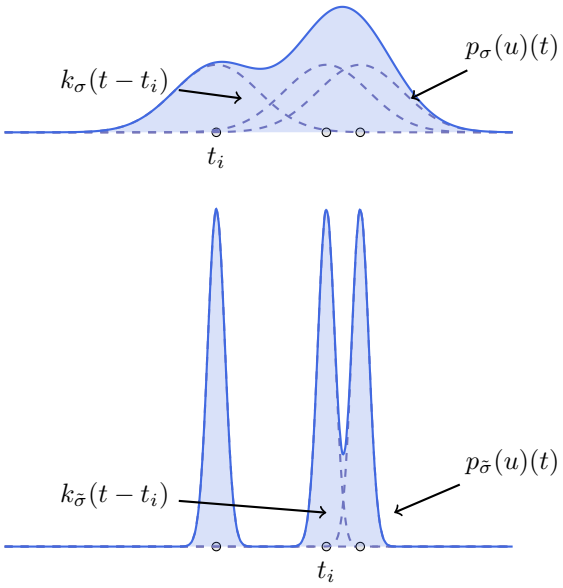


Figure 3: Density estimate for different parameters σ.

and remark that for u_N *as in* (1)

$$p_\sigma(u_N)(t) := \int_\Omega k_\sigma\left(t - u_N(x)\right) dx = \frac{1}{N^2} \sum_{i,j=1}^{N} k_\sigma\left(t - U^{ij}\right) . \tag{3}$$

The joint kernel density of two images u_1, u_2 *with observations* U_1 *and* U_2 *is given by*

$$\hat{p}_\sigma(s,t) = \frac{1}{N^2} \sum_{i,j=1}^{N} K_\sigma(s - U_1^{ij}, t - U_2^{ij}) ,$$

where $K_\sigma(s,t) = k_\sigma(s)k_\sigma(t)$ *is the two-dimensional kernel function. Moreover, we remark that for* $u_{1,N}, u_{2,N}$

$$\hat{p}_\sigma(u_{1,N}, u_{2,N})(s,t) := \int_\Omega K_\sigma\left(s - u_{1,N}(x), t - u_{2,N}(x)\right) dx = \frac{1}{N^2} \sum_{i,j=1}^{N} K_\sigma(s - U_1^{ij}, t - U_2^{ij}) .$$

In the following we review particular kernel functions and show that Parzen windowing and standard histograms are kernel density estimations.

Example 10. *Assume that* $u_i : \Omega \to [0, 255]$, $i = 1, 2$ *are continuous images, with discrete approximations* $u_{i,N}$ *as in* (1).

- ***Parzen windowing*** *density estimation uses the joint density kernel* $K_\sigma(s,t) := k_\sigma(s)k_\sigma(t)$, *where* k_σ *is the* ***normalized Gaussian kernel*** *of variance* σ. *Then for* $i = 1, 2$, *the estimates for the marginal densities are given by*

$$p_\sigma(u_{i,N})(t) := \int_\Omega k_\sigma(u_{i,N}(x) - t)\, dx = \frac{1}{\sqrt{2\pi}\sigma} \int_\Omega \exp\left(\frac{-\left(u_{i,N}(x) - t\right)^2}{2\sigma^2}\right) dx ,$$

and the joint density approximation reads as follows

$$\begin{aligned}\hat{p}_\sigma(s,t) :&= \int_\Omega K_\sigma((u_1(x),u_2(x))-(s,t))\,dx \\ &= \frac{1}{2\pi\sigma^2}\int_\Omega \exp\left(\frac{-\left(u_{1,N}(x)-s\right)^2}{2\sigma^2}\right)\exp\left(\frac{-\left(u_{2,N}(x)-t\right)^2}{2\sigma^2}\right)dx\,.\end{aligned}$$

- ***Histograms:*** *Assume that U only takes values in $\{0,1,\dots,255\}$. When we choose the characteristic function function $\chi_{[-\sigma,\sigma)}$, with $\sigma=\frac{1}{2}$ as kernel function, we obtain the density estimate*

$$\begin{aligned}p_{\chi,\sigma}(t) &= \int_\Omega \chi_{[-\sigma,\sigma[}(u(x)-t)\,dx \\ &= \operatorname{meas}\{x:\ t-\sigma\le u(x)<t+\sigma\} \\ &= \textit{size of pixel}\ \times \textit{number of pixels with value}\ \lfloor t+\sigma\rfloor\,.\end{aligned}$$

Hence $p_{\chi,\sigma}$ corresponds with the histogram of the discrete image.

Example 11. *We return to Example 7. The domain $\Omega=[0,1]$ is partitioned into N equidistant pieces. Let*

$$u_N := \sum_{i=1}^{N}\left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} u(x)dx\right)\chi_{[\frac{i-1}{N},\frac{i}{N}[}\,.$$

Moreover, we consider the piecewise function u_N^T represented in Figure 4. The density according

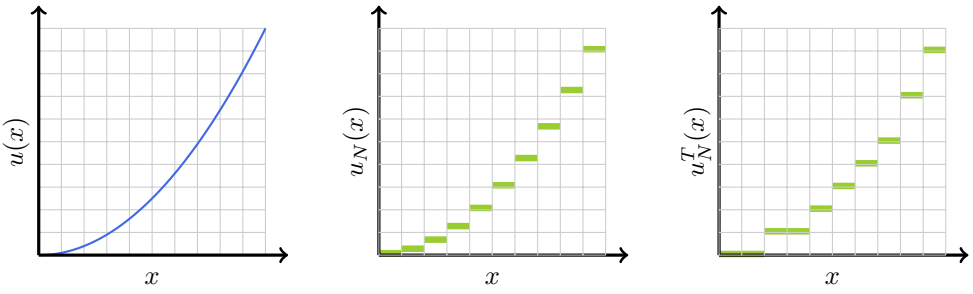


Figure 4: Original u, and discretized versions u_N and u_N^T.

to u, denoted by $p(u)$ and the kernel density estimates of u_N and u_N^T are represented in Figure 5. The resemble the actual density very well.

2.4.2 Csiszár-Divergences (f-Divergences)

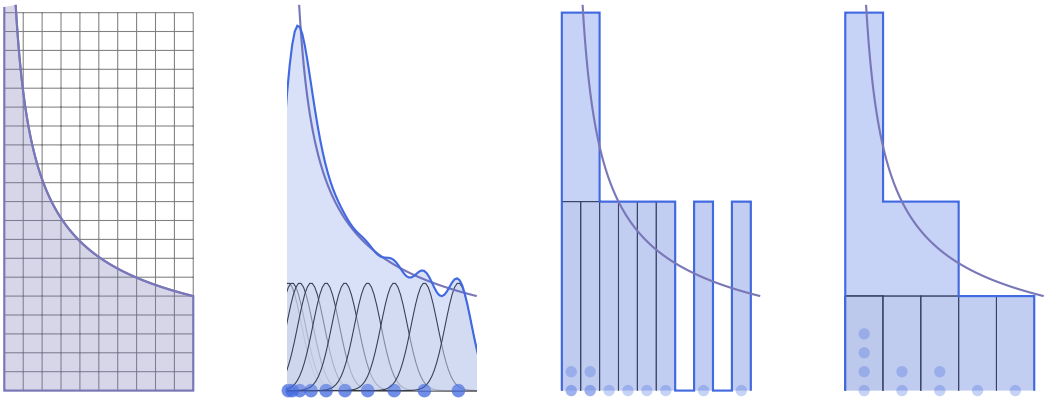
The concept of f-divergences has been introduced by Csiszár in [5] as a generalization of Kullback's I-divergence and Rényi's I-divergence, and at the same time by Ali & Silvey [1]. In probability calculus f-divergences are used to measure the distances between probability densities.

Definition 12. *Set*

$$\mathcal{F}_0 := \{f:[0,\infty)\to\mathbb{R}\cup\{\infty\} : f\ \textit{is convex in}\ [0,\infty),\ \textit{continuous at}\ 0,\ \textit{and satisfies}\ \ f(1)=0\}$$

and

$$V_{pdf} := \left\{p\in L^1(\mathbb{R}) : p\ge 0, \int_{\mathbb{R}} p(t)\,dt\ =\ 1\right\}\,.$$



(a) (b) (c) (d)

Figure 5: (a) Density from the original image u, (b): Parzen window estimation based on u_N ($N = 10$), with $\sigma = 0.07$, (c,d): normalized histogram, based on u_N^T, with $\sigma = 0.05, 0.1$.

Let $g_1, g_2 \in V_{pdf}$ be probability densities functions. The f-divergence between g_1, g_2 is given by

$$\begin{array}{rccl} \mathcal{D}_f: & V_{pdf} \times V_{pdf} & \to & [0,\infty) \\ & (g_1, g_2) & \to & \int_{\mathbb{R}} g_2(t)\, f\left(\frac{g_1(t)}{g_2(t)}\right) dt\,. \end{array} \tag{4}$$

Remark 13.

- *In (4) the integrand at positions t where $g_2(t) = 0$ is understood in the following sense:*

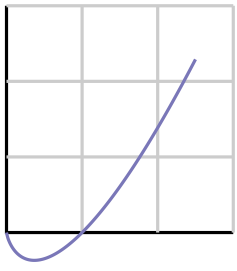
$$0 f\left(\frac{g_1(t)}{0}\right) := \lim_{\bar{t} \searrow 0} \left(\bar{t} f\left(\frac{g_1(t)}{\bar{t}}\right)\right), \quad t \in \mathbb{R}\,.$$

- *In general f-divergences are not symmetric, unless there exists some number c such that the generating f satisfies $f(x) = x f\left(\frac{1}{x}\right) + c(x-1)$.*

Examples for f-Divergences We list several f-divergences that have been used in literature (see [12],[6] and references therein).

Kullback-Leibler Divergenceis the f-divergence with $f(x) = x \log(x)$

$$\mathcal{D}_f(g_1, g_2) = \int_{\mathbb{R}} g_1(t) \log\left(\frac{g_1(t)}{g_2(t)}\right) dt\,.$$

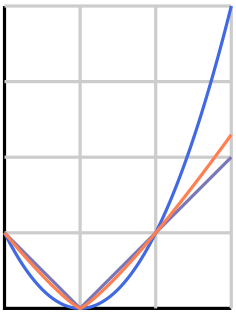


Jensen-Shannon Divergence is the symmetric Kullback-Leibler divergence:

$$\begin{aligned} &\mathcal{D}_f(g_1, g_2) \\ &= \int_{\mathbb{R}} \left(g_1(t) \log\left(\frac{g_1(t)}{g_2(t)}\right) + g_2(t) \log\left(\frac{g_2(t)}{g_1(t)}\right)\right) dt\,. \end{aligned}$$

χ^s-Divergences: These divergences are generated by

$$f^s(x) = |x-1|^s\,, \quad s \in [1,\infty)$$

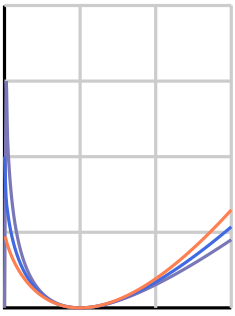


and have the form

$$\begin{aligned}&\mathcal{D}_f(g_1,g_2)\\ &= \int_{\mathbb{R}} g_2(t)\left|\frac{g_1(t)}{g_2(t)}-1\right|^s = \int g_2^{1-s}(t)\,|g_1(t)-g_2(t)|^s\;dt\,.\end{aligned}$$

The χ^1-divergence is a metric. The most widely used out of this family of χ^s divergences is the χ^2-divergence (Pearson).

Dichotomy Class Divergences: The generating function of this class is given by



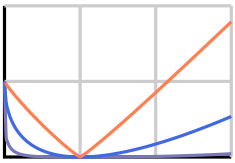
$$f(x) = \begin{cases} x-1-\ln(x) & \text{for } s=0,\\ \frac{1}{s(1-s)}\left(sx+1-s-x^s\right) & \text{for } s\in\mathbb{R}\setminus\{0,1\},\\ 1-x+x\ln(x) & \text{for } s=1\,.\end{cases}$$

The parameter $s=\frac{1}{2}$ provides a distance namely the Hellinger metric

$$\mathcal{D}_f(g_1,g_2) = 2\int_{\mathbb{R}}\left(\sqrt{g_1(t)}-\sqrt{g_2(t)}\right)^2\;dt\,.$$

Matsushita's Divergences: The elements of this class, which is generated by the function

$$f(x) = |1-x^s|^{\frac{1}{s}}\,, \quad 0<s\le 1\,,$$



are prototypes of metric divergences. The distance is given by

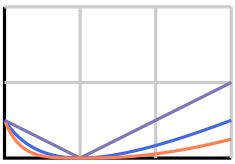
$$d(g_1,g_2) = \left(\mathcal{D}_f(g_1,g_2)\right)^s$$

where

$$\mathcal{D}_f(g_1,g_2) = \int_{\mathbb{R}} g_1(t)\left|1-\left(\frac{g_2(t)}{g_1(t)}\right)^s\right|^{\frac{1}{s}}\;dt\,.$$

Puri-Vincze Divergences: This class is generated by the functions

$$f(x) = \frac{|1-x|^s}{2\,(x+1)^{s-1}}, \qquad s\in[1,\infty)\,.$$



For $s=2$ we obtain the triangular divergence

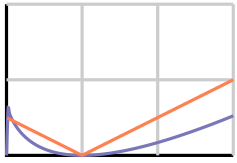
$$\mathcal{D}_f(g_1,g_2) = \int_{\mathbb{R}}\frac{(g_2(t)-g_1(t))^2}{g_2(t)+g_1(t)}\;dt\,.$$

Divergences of Arimoto Type: Generated by the functions

$$f(x) = \begin{cases} \frac{s}{s-1}\left((1+x^s)^{\frac{1}{s}} - 2^{\frac{1}{s}-1}(1+x)\right) & \text{for } s \in (0,\infty)\setminus\{1\} \\ (1+x)\ln(2) + x\ln(x) - (1+x)\ln(1+x) & \text{for } s = 1 \\ \frac{1}{2}|1-x| & \text{for } s = \infty\ . \end{cases}$$

For $s = \infty$ the divergence is proportional to the χ^1 divergence. For $s \in (0,\infty)\setminus\{1\}$ we obtain

$$\mathcal{D}_f(g_1,g_2) = \int_{\mathbb{R}} \frac{s}{s-1}\left(\sqrt[s]{g_1^s(t)+g_2^s(t)} - 2^{\frac{1-s}{s}}(g_1(t)+g_2(t))\right)dt$$



Moreover, this class provides the distances

$$d(g_1,g_2) = (\mathcal{D}_f(g_1,g_2))^{\min\left\{s,\frac{1}{s}\right\}} \quad \text{for } s \in (0,\infty)\ .$$

2.4.3 f-Information

In the following we review the f-information for measuring the distance between probability densities. The most important f-information measure is the **mutual information**.

The notion of **information gain** induced by simultaneously observing two probability measures compared to their separate observations is tightly related to divergences. It results from quantifying the information content of the joint measure in comparison with the product measure.

This motivation leads to the following definition.

Definition 14 (f-information for images)**.** *For $f \in \mathcal{F}_0$ (see Definition 12) we define the f-**information** of $u_1, u_2 \in L^\infty(\Omega)$ by*

$$I_f(u_1,u_2) := \mathcal{D}_f\left(p(u_1)\,p(u_2), p(u_1,u_2)\right)\ ,$$

where the $p(u_i)$ is the probability density of u_i, as introduced in the Section 2.4.1.

*Additionally we define the f-**entropy** of an image u_1 by*

$$H_f(u_1) := I_f(u_1,u_1)\ .$$

In analogy to independent probability densities we call two images u_1, u_2 **independent** if there is no information gain, that is

$$p(u_1,u_2) = p(u_1)p(u_2)\ .$$

Remark 15. *The f-information has the following properties*

- ***Symmetry:*** $I_f(u_1,u_2) = I_f(u_2,u_1)$.
- ***Bounds:*** $0 \le I_f(u_1,u_2) \le \min\{H_f(u_1), H_f(u_2)\}$.
- $I_f(u_1,u_2) = 0$ *if and only if u_1, u_2 are mutually independent.*

The definition of f-information does not make assumptions on the relationship between the image intensities (see [37] for discussion). It does neither assume a linear, nor a functional correlation but only a predictable relationship. For more information on f-information see [35].

Example 16. *The most famous examples of f-informations are the following*

Mutual/Shannon Information: *For $f(x) = x \ln x$ we obtain*

$$I_f(u_1, u_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} p(u_1, u_2)(t_1, t_2) \ln \left(\frac{p(u_1, u_2)(t_1, t_2)}{p(u_1)(t_1)p(u_2)(t_2)} \right) \, dt_1 \, dt_2 \, ,$$

with Shannon entropy

$$H_f(u) = \int_{\mathbb{R}} p(u)(t) \ln \left(\frac{1}{p(u)(t)} \right) dt$$

joint entropy

$$H_f(u_1, u_2) = - \int_{\mathbb{R}} \int_{\mathbb{R}} p(u_1, u_2)(t_1, t_2) \ln \left(p(u_1, u_2)(t_1, t_2) \right) \, dt_1 \, dt_2 \, ,$$

conditional entropy

$$H_f(u_2 \mid u_1) = \int_{\mathbb{R}} p(u_1)(t) H_f(u_2 \mid u_1 = t) \, dt$$

and relative entropy (Kullback Leibler divergence)

$$H_f(u_1 \mid u_2) = \int_{\mathbb{R}} p(u_1)(t) \ln \left(\frac{p(u_1)}{p(u_2)} \right) \, .$$

The relative entropy is not symmetric. Maes et al. [25] and Studholme et al [34] both suggested the use of joint entropy for multi modal image registration. Maes et al. demonstrate the robustness of registration, using mutual information with respect to partial overlap and image degradation, such as noise and intensity inhomogeneities.

Hellinger Information: *For $f(x) = 2x - 2 - 4\sqrt{x}$ (see also Dichotomy Class in Section 2.4.2) we obtain*

$$I_f(u_1, u_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\sqrt{p(u_1, u_2)(t_1, t_2)} - \sqrt{p(u_1)(t_1)p(u_2)(t_2)} \right)^2 \, dt_1 \, dt_2 \, ,$$

with Hellinger entropy

$$H_f(u) = 2 \left(1 - \int_{\mathbb{R}} \left(p(u)(t) \right)^{\frac{3}{2}} dt \right) \, .$$

Both are bounded from above by 2.

For measuring the distance between discrete images U_1 and U_2 it is common to map the images via kernel estimation to continuous estimates of their intensity value densities $p_\sigma(u_{i,N})$, where $p_\sigma(u_{i,N})$ is as defined in (3). The difference between images is then measured via the distance between the associated estimated probability densities.

Example 17. *For $U_i, i = 1, \ldots, 3$ as in Figure 2, let $u_{i,N}$ be the corresponding piecewise constant functions. Note that U_1 and U_2 are somehow related. In other words, they highly dependent on each other, so we can expect a low information value. Comparing the images point wise with least squares, shows a higher similarity value for U_2 and U_3 than for U_1 and U_2.*

For the ease of presentation we work with histograms, instead of Parzen windowing or any other kernel density estimation. Recall that the estimated probability function $p_\sigma(u_{i,N})$ is equal to the normalized histogram of U_i. The histograms (connected to the marginal density densities) are given by

	1	2	3	4	5	6	7	8	9	10
$H(U_1)$	6	7	6	9	3	4	1	0	0	0
$H(U_2)$	1	0	4	0	3	0	9	6	7	6
$H(U_3)$	3	2	5	3	2	4	0	5	6	6

$JH(U_1,U_2)$	1	2	3	4	5	6	7	8	9	10	$H(U_2)$
1							1				1
2											0
3						4					4
4											0
5					3						3
6											0
7				9							9
8			6								6
9		7									7
10	6										6
$H(U_1)$	6	7	6	9	3	4	1	0	0	0	

Table 2: Joint histogram of U_1 and U_2. Note that the entries are put in an order, due to the dependence between U_1 and U_2. This will be reflected in a low f-information value.

$JH(U_2,U_3)$	1	2	3	4	5	6	7	8	9	10	$H(U_3)$
1							1	1	1		3
2								2			2
3					1		2		2		5
4			2		1						3
5			1						1		2
6					1		1	1	1		4
7											0
8							1		2	2	5
9	1						2	1		2	6
10			1				2	1		2	6
$H(U_2)$	1	0	4	0	3	0	9	6	7	6	

$JH(U_3,U_1)$	1	2	3	4	5	6	7	8	9	10	$H(U_1)$
1								2	2	2	6
2	1		2		1	1		2			7
3	1	2				1			1	1	6
4	1		2			1		1	2	2	9
5			1	1		1					3
6				2	1					1	4
7									1		1
8											0
9											0
10											0
$H(U_3)$	3	2	5	3	2	4	0	5	6	6	

Table 3: Joint histograms of U_2, U_3 and U_3, U_1. The entries are disperse, this will be reflected in a lower f-information as in the case for U_1, U_2.

In order to calculate the information measures, we calculate the joint histograms of U_1, U_2, U_3, that is $JH(U_1,U_2) : (s,t) \to$ number of pixels such that $U_1^{ij} = s$ and $U_2^{ij} = t$ (see Tables 2, 3).

The entries in the joint histogram of U_1, U_2 are located along a diagonal, whereas the entries of the other two joint histograms are spread all over. Hence we can observe the dependence of $p_\sigma(u_{1,N}), p_\sigma(u_{2,N})$ already by looking at the joint histogram. Next we calculate the Hellinger and the Mutual information.

For the f-entropies we obtain

	U_1	U_2	U_3
Mutual entropy	1.91	1.91	2.13
Hellinger entropy	1.17	1.17	1.30

and for the f-information measures:

	(U_1,U_2)	(U_2,U_3)	(U_3,U_1)
Mutual information	**1.91**	0.74	0.74
Hellinger information	**1.17**	0.57	0.57
Sum of least squares	31.44	**14.56**	21.56

Indeed, in both cases (Hellinger and Mutual information), U_1, U_2 (high f-information value) can be considered as more similar than U_1 and U_3. Whereas the least squares value between U_1, U_2 is the highest, meaning that they differ at most.

We can observe in Example 17 that the values of f-information differ a lot by different choices of the function f. Moreover, it's not easy to interpret the values, hence one is interested in calculating normalized values.

Normalized Mutual Information Studholme [34] proposed a normalized measure of mutual information. Normalized f-information is defined by

$$NI_f(u_1, u_2) := \frac{H_f(u_1) + H_f(u_2)}{I_f(u_1, u_2)} .$$

If $u_1 = u_2$, then the normalized f-information is minimal with value 2.

Entropy Correlation Coefficient Collignon and Maes [25] suggested the use of the entropy correlation coefficient, another form of normalized f-information:

$$H_f CC(u_1, u_2) = \frac{2\, I_f(u_1, u_2)}{H_f(u_1) + H_f(u_2)} = 2 - \frac{2}{NI_f(u_1, u_2)} .$$

The entropy correlation coefficient is one if $u_1 = u_2$ and zero if u_1 and u_2 are completely independent.

Exclusive f-Information is defined by

$$EI_f(u_1, u_2) := H_f(u_1) + H_f(u_2) - 2I_f(u_1, u_2)$$

Note that the exclusive f-information is **minimal** for $u_1 = u_2$.

Mutual Information	(u_1, u_2)	(u_2, u_3)	(u_3, u_1)
Normalized*	**2.00**	5.32	5.32
Entropy Correlation Coefficient	**1.00**	0.38	0.38
Exclusive*	**0.00**	2.46	2.46
Hellinger Information	(u_1, u_2)	(u_2, u_3)	(u_3, u_1)
Normalized*	**2.00**	4.36	4.36
Entropy Correlation Coefficient	**1.00**	0.46	0.46
Exclusive *	**0.00**	1.34	1.34

Table 4: Comparison of measures composed by f-information and f-entropies. * Normalized and exclusive informations are minimal, if the images are equal, whereas the entropy correlation coefficient is maximal.

2.5 Distance Measures Including Statistical Prior Information

Most multi modal measures used in literature do not consider the underlying image context or other statistical prior information on the image modalities. Recently several groups developed similarity measures that incooperate such information:

- Leventon & Grimson [23] proposed to learn **prior information** from training data (registered multi modal images) by estimating the joint intensity distributions of the training images. Based on this paper Chung et al. [4] proposed a to use the Kullback-Leiber distance

to compare the learned joint intensity distribution, with the joint intensity distribution of the images, in order to compare multimodal images. This idea was extended by Guetter et al. [14], who combines mutual information with the incorporation of learned prior knowledge with a Kullback-Leibler term.

As a follow up of their ideas we suggest the following type of generalized similarity measures: Let p_σ^l be the learned joint intensity density (learned from the training data set) and $\alpha \in [0,1]$. For $f \in \mathcal{F}_0$ define

$$\mathcal{S}_\alpha(u_1,u_2) := \alpha \mathcal{D}_f(p_\sigma^l, p_\sigma(u_1,u_2)) + (1-\alpha) \underbrace{\mathcal{D}_f\left(p_\sigma(u_1,u_2), p_\sigma(u_1)p_\sigma(u_2)\right)}_{I_f(u_1,u_2)} .$$

- Instead of using a universal, but a-priori fixed similarity measure one can **learn a similarity** measure in a discriminative manner. The methodology proposed by Lee et al. [22] uses a learning algorithm, that constructs a similarity measure based on a set of pre-registered images.

3 Mathematical Models for Variational Imaging

In the following we proceed with an abstract setting. We are given a physical model F, which in mathematical terms is an operator between spaces U and V. For given data $v \in V$ we aim for solving the operator equation

$$F(\Phi) = v .$$

In general the solution is not unique and we aim for finding the solution with **minimal energy**, that is we aim for a minimizer of the constraint optimization problem

$$\mathcal{R}(\Phi) \to \min \text{ subject to } F(\Phi) = v .$$

In practice a complication of this problem is that only approximate (noisy) data $v^\delta \in V$ of v is available. To take into account uncertainty of the data it is then intuitive to consider the following constrained optimization problem instead

$$\mathcal{R}(\Phi) \to \min \text{ subject to } \left\|F(\Phi) - v^\delta\right\|^2 \le \delta , \tag{5}$$

where δ is an upper bound for the approximation error $\left\|v - v^\delta\right\|$. It is known that solving (5) is equivalent to minimizing the Tikhonov functional,

$$\Phi \to \frac{1}{2}\left\|F(\Phi) - v^\delta\right\|^2 + \alpha \mathcal{R}(\Phi), \tag{6}$$

where $\alpha > 0$ is chosen according to Morozov's discrepancy principle [19].

For the formulation of the constrained optimization problem, Tikhonov method, respectively, it is essential that $F(\Phi)$ and v, v^δ, respectively, represent data of the same modality. If $F(\Phi)$ and the data, which we denote now by w, are of different kind, then it is intuitive to use a multi modal similarity measure $\mathcal{S}^r$, instead of the least squares distance, which allows for comparison of $F(\Phi)$ and w. Consequently, we consider the multi modal variational method, which consists in minimization of

$$\Phi \to \mathcal{T}_{\alpha,w^\delta}(\Phi) := \mathcal{S}^r(F(\Phi), w^\delta) + \alpha \mathcal{R}(\Phi), \quad \alpha > 0 .$$

In the limiting case, i.e., for $\delta \to 0$, one aims for recovering an $\mathcal{RS}^r$-minimizing solution $\Phi^\dagger$ if

$$\mathcal{R}(\Phi^\dagger) = \min\left\{\mathcal{R}(\Phi) : \Phi \in \mathcal{A}\right\} \text{ where } \mathcal{A} = \left\{\Phi : \Phi = \operatorname{argmin}\left\{\mathcal{S}^r\left(F(\cdot), w\right)\right\}\right\} .$$

To take into account priors in Tikhonov regularization the standard way is again by a least squares approach. In this case, for regularization the least squares functional

$$\Phi \to \mathcal{R}_1(\Phi) = \frac{1}{2} \|\Phi - \Phi_0\|^2$$

is added to $\frac{1}{2} \left\|F(\Phi) - v^\delta\right\|^2$ (see e.g. [8]). In analogy, we consider the regularization functional (6) and incorporate priors by adding generalization of the functional $\mathcal{R}_1(\Phi)$. Taking into account prior information Ψ_0, that might come, for instance, from another modality, this leads to the following class of generalized Tikhonov functionals

$$\mathcal{T}_{\alpha,\beta}^{w^\delta,\Psi_0}(\Phi) := \mathcal{S}^r(F(\Phi), w^\delta) + \alpha\mathcal{R}(\Phi) + \beta\mathcal{S}^p(\Phi, \Psi_0) \ .$$

Here $\mathcal{S}^p$ is an appropriate multi modal similarity measure. In the limiting case, i.e., for $\delta \to 0$, one aims for recovering an $\gamma - \mathcal{R}\mathcal{S}^r\mathcal{S}^p$-minimizing solution $\Phi^\dagger$ if

$$\begin{aligned} \mathcal{R}(\Phi^\dagger) + \gamma\mathcal{S}^p(\Phi^\dagger, \Psi_0) = \min\left\{\mathcal{R}(\Phi) + \gamma\mathcal{S}^p(\Phi^\dagger, \Psi_0) : \Phi \in \mathcal{A}\right\} \text{ where} \\ \mathcal{A} = \left\{\Phi : \Phi = \operatorname{argmin}\left\{\mathcal{S}^r\left(F(\cdot), w\right)\right\}\right\} \ . \end{aligned}$$

The γ-parameter balances between the amount of prior information and regularization and satisfies $\gamma = \lim_{\alpha,\beta\to 0} \frac{\beta}{\alpha}$. For theoretical results on existence of minimizing elements of the functionals and convergence we refer to [11, 29].

4 Registration

In this section we review variational methods for **image registration**. This problem consists in determining a spatial transformation (vector field) Φ that aligns pixels of two images u_R and u_T in an optimal way. We use the terminology **reference image** for u_R and **template image** for u_T, which both are assumed to be compactly supported functions in Ω. That is, we consider the problem of determining the optimal transformation, which minimizes the functional

$$u \to \mathcal{S}(u_T \circ (id + \Phi), u_R) \ . \tag{7}$$

To establish the context to the inverse problems setting we use the setting $F(\Phi) = u_T(id + \Phi)$

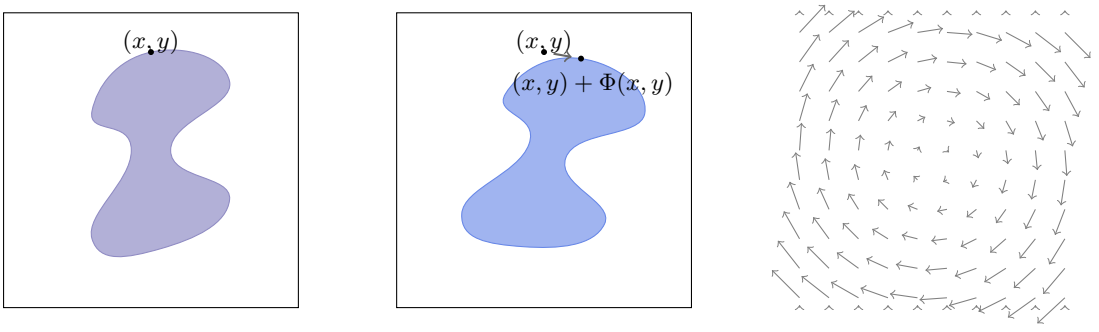


Figure 6: left: images u_R, u_T, right: deformation field Φ.

and $w^\delta = u_R$. In general the problem of minimizing (7) is ill-posed. Tikhonov type variational regularization for registration then consists in minimization of the functional

$$\Phi \to \mathcal{S}^r\left(u_T \circ (id + \Phi), u_R\right) + \alpha\mathcal{R}(\Phi) \ . \tag{8}$$

(we do not consider constained registration here but concentrate on Tikhonov regularization)

Image registration (also of voxel (3D) data) is widely used in medical imaging, for instance for monitoring and evaluating tumor growth, disease development, and therapy evaluation.

Variational methods for registration differ by the choice of the regularization functional $\mathcal{R}$ and the similarity measure $\mathcal{S}^r$. There exists a variety of similarity measures that are used in practice. For some surveys we refer to [31, 17, 27].

The regularization functional $\mathcal{R}$ typically involves differential operators. In particular, for nonrigid registration energy functionals from elasticity theory and fluid dynamics are used for regularization.

The optimality condition for a minimizer of (8) reads as follows:

$$\alpha D_\Phi \left(\mathcal{R}(\Phi), \Psi\right) + D_\Phi \left(\mathcal{S}^r \left(u_T \circ (id + \Phi), u_R\right), \Psi\right) = 0 \qquad \text{for all } \Psi \in U \ , \tag{9}$$

where $D_\Phi(\mathcal{T}, \Psi)$ denotes the directional derivative of a functional $\mathcal{T}$ in direction Ψ. The left hand side of the equation is the steepest descent functional of the energy functional (8). In the following we highlight some steepest descent functionals according to variational registration methods.

Example 18 (Elastic Registration)**.** *Set $\alpha = 1$, $\mathcal{S}^r(v_1, v_2) = \frac{1}{2}\|v_1 - v_2\|^2_{L^2}$. We consider an elastic regularization functional of the form*

$$\mathcal{R}(\Phi) = \int_\Omega \sum_{i=1}^{2}\sum_{j=1}^{2} \left(\frac{\lambda}{2} \frac{\partial}{\partial x_i}\Phi^i \frac{\partial}{\partial x_j}\Phi^j + \frac{\mu}{4}\left(\frac{\partial}{\partial x_j}\Phi^i + \frac{\partial}{\partial x_i}\Phi^j\right)^2 \right) \ ,$$

where $\lambda, \mu \geq 0$ are Lamé parameters and $\Phi = (\Phi^1, \Phi^2)$. λ is adjusted to control the rate of growth or shrinkage of local regions within the deforming template and μ is adjusted to control shearing between adjacent regions of the template [3]. In this case the optimality condition for minimizing $\alpha\mathcal{R}(\Phi) + \mathcal{S}^r(u_T \circ (id + \Phi))$, given by (9), is satisfied if Φ solves the following PDE

$$\mu\Delta\Phi(x) + (\mu + \lambda)\nabla\left(\nabla \cdot \Phi(x)\right) = \underbrace{-\frac{1}{\alpha}\left(u_T\left(x + \Phi(x)\right) - u_R(x)\right)\nabla u_T\left(x + \Phi(x)\right)}_{\frac{\partial}{\partial\Phi}\mathcal{S}^r} \ .$$

Here $\Delta\Phi = (\Delta\Phi^1, \Delta\Phi^2)$ and $D_\Phi\left(\mathcal{S}^r\left(u_T \circ (id + \Phi), u_R\right), \Psi\right) = \int_\Omega \frac{\partial}{\partial\Phi}\mathcal{S}^r \cdot \Psi$. This partial differential equation is known as linear elastic equation and is derived assuming small angles of rotation and small linear deformations. When large displacements are inherent it is not applicable [2, 13, 24, 18, 28, 39].

Example 19 (Elastic Registration with f-Information:)**.** *Assume that $k_\sigma \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ is some kernel density function. Moreover, let $K_\sigma(s,t) = k_\sigma(s)k_\sigma(t)$. We pose the similarity measure as the f-information between the template and the reference image:*

$$\mathcal{S}^r\left(u_T \circ (id + \Phi), u_R\right) = H_f(u_R) - I_f\left(u_T \circ (id + \Phi), u_R\right) \ ,$$

and set α and $\mathcal{R}$ as in the previous example. In order to write the derivative of I_f in a compact way, we use the abbreviations $\tilde{\Phi} := id + \Phi$. The derivative of $p_\sigma(u_T \circ \tilde{\Phi})$ with respect to $\tilde{\Phi}$, in direction Ψ is given by

$$D_{\tilde{\Phi}}\left(p_\sigma(u_T \circ \tilde{\Phi}), \Psi\right)(t) = \int_\Omega k'_\sigma\left(t - u_T(\tilde{\Phi}(x))\right)\ \nabla u_T(\tilde{\Phi}(x)) \cdot \Psi(x)\, dx$$

and

$$D_{\tilde{\Phi}}\left(\hat{p}_\sigma(u_T \circ \tilde{\Phi}, u_R), \Psi\right)(s,t) = \int_\Omega k'_\sigma\left(s - u_T(\tilde{\Phi}(x))\right) k_\sigma(t - u_R(x))(\nabla u_T(\tilde{\Phi}(x)) \cdot \Psi(x))\, dx \ .$$

We use the following abbreviations:

$$g_1(s,t) := \frac{p_\sigma(u_T \circ \tilde{\Phi})(s) p_\sigma(u_R)(t)}{p_\sigma(u_T \circ \tilde{\Phi}, u_R)(s,t)}\,, \quad g_2(s,t) := \frac{p_\sigma(u_R)(t)}{\left(\hat{p}_\sigma(u_T \circ \tilde{\Phi}, u_R)(s,t)\right)^2}\,,$$

and

$$\begin{aligned} g_3(s,t) :=& D_{\tilde{\Phi}}(p_\sigma(u_T \circ \tilde{\Phi}), \Psi)(s) \hat{p}_\sigma(u_T \circ \tilde{\Phi}, u_R)(s,t) + \\ & p_\sigma(u_T \circ \tilde{\Phi})(s) D_{\tilde{\Phi}} \left(\hat{p}_\sigma(u_T \circ \tilde{\Phi}, u_R), \Psi\right)(s,t)\,. \end{aligned}$$

With this we can calculate the derivative of the f-information to be

$$\begin{aligned} D_{\tilde{\Phi}} \left(I_f(u_T \circ \tilde{\Phi}, u_R), \Psi\right) = \int_{\mathbb{R}} \int_{\mathbb{R}} & D_{\tilde{\Phi}} \left(p_\sigma(u_T \circ \tilde{\Phi}), \Psi\right)(s) p_\sigma(u_R)(t) f\left(g_1(s,t)\right) \\ & + p_\sigma(u_T \circ \tilde{\Phi})(s) p_\sigma(u_R)(t) f'\left(g_1(s,t)\right) g_2(s,t) g_3(s,t)\, dt\, ds\,. \end{aligned}$$

For mutual information this simplifies to

$$\begin{aligned} D_{\tilde{\Phi}} \left(MI(u_T \circ \tilde{\Phi}, u_R), \Psi\right) = \int_{\mathbb{R}} \int_{\mathbb{R}} & \left(D_{\tilde{\Phi}} \left(\hat{p}_\sigma(u_T \circ \tilde{\Phi}, u_R), \Psi\right)(s,t) \ln\left(\frac{1}{g_1(s,t)}\right) \right. \\ & \left. + \frac{D_{\tilde{\Phi}} \left(\hat{p}_\sigma(u_T \circ \tilde{\Phi}, u_R), \Psi\right)(s,t)}{p_\sigma(u_T \circ \tilde{\Phi})(s) p_\sigma(u_R)(t)} \right) ds\, dt\,. \end{aligned}$$

A detailed exposition on elastic registration with mutual information can be found in [9, 16, 11].

In this section we have presented a general framework on multi modal image registration. Below we give a short overview on relevant literature on this topic.

Kim and Fessler [21] describe an intensity-based image registration technique that uses a robust correlation coefficient as a similarity measure for images. It is less sensitive to outliers, that are present in one image, but not in the other. Kaneko [20] proposed the selective correlation coefficient, as an extension of the correlation coefficient. Van Elsen et al. investigated similarity measures for MR and CT images. She proposed to calculate the correlation coefficient of geometrical features [36]. Alternatively to the correlation coefficient, one could calculate Spearman's rank correlation coefficient (also known as Spearman's ρ), which is a non parametric measure of correlation [10], but not very popular in multi modal imaging. Roche et al. [33, 32] tested the correlation ratio to align MR, CT and PET images. Woods et al [41] developed an algorithm based on this measure for automated aligning and re-slicing PET images. Independently, several groups realized that the problem of registering two different images modalities, can be cast in an information theoretic framework. Collignon et al. [25] and Studholme et al. [34] both suggested using the joint entropy of the combined images as a registration potential. For MR-CT registrations, the learned similarity measure by Lee et al. outperforms all standard measures. Experimental results for learning similarity measures for multi modal images can be found in [22].

5 Recommended Reading

For recent results on divergences and information measures, we refer to **Computational Information Geometry**. Website: `http://blog.informationgeometry.org/`.

Comparison and evaluation of different similarity measures for CT,MR,PET brain images can be found in [40].

It's worth to mention two databases:

- **The Retrospective Image Registration Evaluation Project** is designed to compare different multi modal registration techniques. It involves the use of a database of image volumes, commonly known as the "Vanderbilt Database", on which the registrations are to be performed. Moreover it provides a training data set for multi modal image registration. Link: `http://www.insight-journal.org/RIRE/index.php`

- **Validation of Medical Image Registration.** This is a data base with references (international publications) on medical image registration including a validation study of different similarity measures. Link: `http://idm.univ-rennes1.fr/VMIP/model/index.html`

A number of image registration software tools have been developed in the last decade. The following support multi modal image comparison:

ITK is an open-source, cross-platform system that provides developers with an extensive suite of software tools for image analysis. It supports the following similarity measures: mean squares metric, normalized cross correlation metric, mean reciprocal square differences, mutual information (different implementations [38, 26]), Kullback-Leibler distance, normalized mutual information, correlation coefficient, kappa statistics (for binary images), gradient difference metric. Website: `http://www.itk.org/`.

FLIRT stands for fast and flexible image registration toolbox, and has been developed by the SAFIR-research group in Lübeck. It includes sum of squared differences, mutual information, and normalized gradient fields. Website: `http://www.math.uni-luebeck.de/safir/software.shtml`.

AIR stands for automated image registration. It supports standard deviation of ratio images, least squares and least squares with global intensity rescaling. Website: `http://bishopw.loni.ucla.edu/AIR5/`.

RView This software integrates a number of 3D/4D data display and fusion routines together with 3D rigid volume registration using normalized Mutual information. It also contains many interactive volume segmentation and painting functions for structural data analysis. Website: `http://www.colin-studholme.net/software/software.html`.

Acknowledgement

The work of OS has been supported by the Austrian Science Fund (FWF) within the national research networks Industrial Geometry, project 9203-N12, and Photoacoustic Imaging in Biology and Medicine, project S10505-N20.

The work of CP has been supported by the Austrian Science Fund (FWF) via the Erwin Schrödinger scholarship J2970.

References

[1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B*, 28:131–142, 1966.

[2] R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46:1–21, 1989.

[3] G. Christensen. *Deformable shape models for anatomy.* PhD thesis, Washington University, Dept. of Electrical Engineering., 1994.

[4] A.C.S. Chung, W.M. Wells, A. Norbash, and W.E.L. Grimson. Multi-modal image registration by minimizing Kullback-Leibler distance. In *MICCAI '02: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II*, pages 525–532, London, UK, 2002. Springer-Verlag.

[5] I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:85–108, 1963.

[6] S. S. Dragomir. Some general divergence measures for probability distributions. *Acta Math. Hungar.*, 109(4):331–345, 2005.

[7] M. Droske and M. Rumpf. A variational approach to nonrigid morphological image registration. *SIAM J. Appl. Math.*, 64(2):668–687 (electronic), 2003/04.

[8] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[9] K. Ens, H. Schumacher, A. Franz, and B. Fischer. Improved elastic medical image registration using Mutual Information. In J. P. W. Pluim and J. M. Reinhardt, editors, *Medical Imaging 2007: Image Processing*, volume 6512, page 65122C. SPIE, 2007.

[10] L. Fahrmeir, R. Künstler, I. Pigeot, and G. Tutz. *Statistik*. Springer, Berlin, 5 edition, 2004.

[11] O. Faugeras and G. Hermosillo. Well-posedness of two nonrigid multimodal image registrational methods. *SIAM J. Appl. Math.*, 64(5):1550–1587 (electronic), 2004.

[12] D. Feldman and F. Österreicher. A note on f-divergences. *Studia Sci. Math. Hungar.*, 24(2-3):191–200, 1989.

[13] J. Gee, D. Haynor, L. Le Briquer, and R. Bajcsyand. Advances in elastic matching theory and its implementation. In *CVRMed*, pages 63–72, 1997.

[14] C. Guetter, C. Xu, F. Sauer, and J. Hornegger. Learning based non-rigid multi-modal image registration using Kullback-Leibler divergence. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, volume 3750 of *Lecture Notes in Computer Science*, pages 255–262. Springer Verlag, 2005.

[15] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. In *Methods of Information in Medicine, Schattauer Verlag*, pages 726–733, 2006.

[16] S. Henn and K. Witsch. Multimodal image registration using a variational approach. *SIAM J. Sci. Comput.*, 25(4):1429–1447, 2003.

[17] G. Hermosillo. *Variational methods for multimodal image matching*. PhD thesis, Université de Nice, France, 2002.

[18] L. Hömke. *A multigrid method for elastic image registration with additional structural constraints*. PhD thesis, Heinrich-Heine Universität Düsseldorf, 2006.

[19] V. K. Ivanov, V. V. Vasin, and V. P. Tanana. *Theory of linear ill-posed problems and its applications*. Inverse and Ill-posed Problems Series. VSP, Utrecht, second edition, 2002. Translated and revised from the 1978 Russian original.

[20] S. Kaneko, Y. Satoh, and S. Igarashi. Using selective correlation coefficient for robust image registration. *Pattern Recognition*, 36(5):1165–1173, 2003.

[21] J. Kim and J. A. Fessler. Intensity-based image registration using robust correlation coefficients. *Medical Imaging, IEEE Transactions on*, 23(11):1430–1444, 2004.

[22] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N.D. Cahill, and B. Scholkopf. Learning similarity measure for multi-modal 3d image registration. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 0:186–193, 2009.

[23] M.E. Leventon and W.E.L. Grimson. Multi-modal volume registration using joint intensity distributions. In *MICCAI '98: Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1057–1066, London, UK, 1998. Springer-Verlag.

[24] B. Likar and F. Pernus. A hierarchical approach to elastic registration based on Mutual Information. *Image and Vision Computing*, 19:33–44, 2001.

[25] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of Mutual Information. *Medical Imaging, IEEE Transactions on*, 16(2):187–198, April 1997.

[26] D. Mattes, D.R. Haynor, H. Vesselle, T.K. Lewellen, and W. Eubank. PET-CT image registration in the chest using free-form deformations. *Medical Imaging, IEEE Transactions on*, 22(1):120–128, Jan. 2003.

[27] J. Modersitzki. *Numerical Methods for Image Registration*. Numerical Mathematics and Scientific Computation. Oxford University Press, 2004.

[28] W. Peckar, C. Schnörr, K. Rohr, and H.S. Stiehl. Non-rigid image registration using a parameter-free elastic model. In *British Machine Vision Conference*, 1998.

[29] C. Pöschl. *Tikhonov Regularization with Gerneral Residual Term*. PhD thesis, Leopold Franzens Universität, Innsbruck, Austria, November 2008.

[30] S. T. Rachev. *Probability metrics and the stability of stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1991.

[31] A. Roche. *Recalage d'images médicales par inférence statistique*. PhD thesis, Université de Nice, Sophia-Antipolis, France, 2001.

[32] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Lecture Notes in Computer Science*, volume 1496, pages 1115–1124. Springer Verlag, 1998.

[33] A. Roche, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Medical Image Computing and Computer-Assisted Interventation MICCAI98*, pages 1115–1124. Springer Verlag, 1998. LNCS 1496.

[34] C. Studholme. *Measures of 3D Medical Image Alignment*. PhD thesis, University of London, August 1997.

[35] I. Vajda. *Theory of statistical inference and information*. Kluwer Academic Publishers, 1989.

[36] P. van den Elsen, J. B. A. Maintz, and M. A. Viergever. Automatic registration of CT and MR brain images using correlation of geometrical features. *IEEE Trans. Med. Imaging*, 14(2):384–396, 1995.

[37] P.A. Viola. *Alignment by Maximization of Mutual Information*. PhD thesis, Massachusetts Institute of Technology, 1995.

[38] P.A. Viola and W.M. Wells. Alignment by maximization of Mutual Information. In *ICCV '95: Proceedings of the Fifth International Conference on Computer Vision*, page 16. IEEE Computer Society, 1995.

[39] X.Y. Wang and D.D. Feng. Automatic elastic medical image registration based on image intensity. *International Journal of Image and Graphics*, 5(2):351–369, April 2005.

[40] J. West, J.M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer, R.M. Kessler, R.J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. van den Elsen, S. Napel, T.S. Sumanaweera, B. Harkness, P.F. Hemler, D.L.G. Hill, D.J. Hawkes, C. Studholme, J.B.A. Maintz, M.A. Viergever, G. Malandain, X. Pennec, M.E. Noz, G.Q. Maguire, M. Pollack, C.A. Pelizzari, R.A. Robb, D. Hanson, and R.P. Woods. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21:554–566, 1997.

[41] R.P. Woods, S.R. Cherry, and J.C. Mazziotta. Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assist Tomogr*, 16:620–633, 1992.