A Contextual Maximum Likelihood Framework for Modeling Image Registration

Christian Wachinger[1,2] and Nassir Navab[1]

[1]Computer Aided Medical Procedures, Technische Universität München, Germany
[2]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, USA

Abstract

We introduce a novel probabilistic framework for image registration. This framework considers, in contrast to previous ones, local neighborhood information. We integrate the neighborhood information into the framework by adding layers of latent random variables, characterizing the descriptive information of each image. This extension has multiple advantages. It allows for a unified description of geometric and iconic registration, with the consequential analysis of similarities. It enables to arrange registration techniques in a continuum, limited by pure intensity- and feature-based registration. With this wide spectrum of techniques combined, we can model hybrid registration approaches. The probabilistic coupling allows further to deduce optimal descriptors and to model the adaptation of description layers during the process, as it is done for joint registration/segmentation. Finally, we deduce a new registration algorithm that allows for a dynamic adaptation of the description layers during the registration. Excellent results confirm the advantages of the new registration method, the major contribution of this article lies, however, in the theoretical analysis.

1. Introduction

Registration is a fundamental process in computer vision. A common classification is to distinguish between geometry- and intensity-based approaches. Geometric approaches establish the spatial relationship between images based on extracted features, landmarks, surfaces, or point clouds. Intensity-based or iconic approaches directly operate on the images by comparing their pixel intensities or photometric properties. For intensity-based registration, unifying probabilistic frameworks [20, 28, 38] were proposed. These frameworks are essential in better understanding and categorizing different types of registration approaches. With a strict deduction from a mathematical framework, it is possible to detect implicitly incorporated assumptions. Discovering such assumptions allows for better adapting registration to specific applications and to justify the adequacy of an approach in a specific scenario. Concepts from probability theory, such as maximum likelihood or a posteriori estimation, were in this context shown to be very useful to reason about image registration. The limitation of currently existing probabilistic frameworks is, however, their focus on modeling the similarity measure.

Looking at registration in practice, we observe that processing steps such as gradient calculation, multi-scale analysis, and noise reduction are applied to the images, before performing the alignment. Further, the comparison of single pixel information is prone to noise, leading to the introduction of context and spatial information in registration. With the presented contextual, probabilistic framework we are able to model these approaches. Moreover, we can model geometric approaches through the introduction of layers of latent random variables. Dealing with these representations allows for differentiating between pure image processing steps, such as smoothing and gradient calculation, and the estimation of the similarity between images. This helps to classify registration techniques and identify commonalities.

2. Probabilistic Registration Framework

In order to describe image registration from a probabilistic point of view, we consider each image to be a random variable U. The probability of the appearance of a concrete sample image u is $p(U = u)$, with the simplified notation $p(u)$. Considering further that an image is defined on a grid Ω, each spatial location $U(x)$ with $x \in \Omega$ is a random variable. Taking the set of intensity values $\mathcal{I}$, *e.g.* $\mathcal{I} = \{0, 1, \dots, 255\}$, the probability of a location having a certain intensity is $p(U(x) = i)$ with $i \in \mathcal{I}$. The goal of registration is to find the transformation T that expresses the spatial relationship between the two images u and v

$$u(x) = v(T(x)). \tag{1}$$

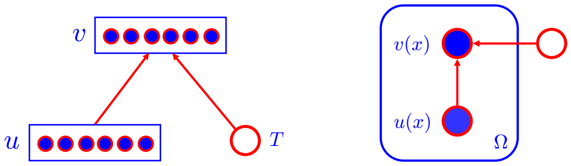

Figure 1. Left: Probabilistic graphical model showing the probability $p(v|u,T)$, where each of the images consists of a random variable for each location $x \in \Omega$, in this case 6. Right: Assumption of i.i.d. coordinate samples, $\prod p(v(x)|u(x),T)$, illustrated as plate.

This is the underlying model of the generative, joint probability

$$\hat{T} = \arg\max_{T \in \mathcal{T}} p(u,v,T), \tag{2}$$

with $\mathcal{T}$ being the space of transformations and $\hat{T}$ the optimal transformation with respect to the model. The model in equation (1) is commonly augmented with noise and an intensity mapping [20]. Other noise distributions than the standard Gaussian can be used to adapt the registration to specific applications, such as speckle noise in ultrasound images [29]. The intensity mapping accounts for multi-modal registration and leads, *e.g.*, to sum of squared differences (SSD), correlation ratio (CR), and mutual information (MI), by assuming an identical, functional, or statistical intensity relationship [28, 20], respectively.

2.1. I.I.D. Coordinate Samples

A general assumption of the unifying approaches [20, 28, 39] are independent and identically distributed (i.i.d.) coordinate samples. With this assumption, equation (2) simplifies tox

$$p(u,v,T) = \prod_{x \in \Omega} p(u(x),v(x),T). \tag{3}$$

This is illustrated in figure 1 for the probability $p(v|u,T)$, which is the likelihood term of the generative probability, obtained with Bayes' theorem [20]. Since each of the spatial locations in the images corresponds to a random variable, we use the *plate* visualization [6], permitting a compact representation of the graph.

The i.i.d. assumption splits the general problem of similarity estimation between images up into several subproblems of similarity estimation between pixels. This simplification is necessary for the deduction of similarity measures such as SSD, CR, and MI [20], but does not accurately model the real world. Objects in the image have a certain size, which is rather rarely limited to the extent of one pixel, so that the i.i.d. assumption is not justified.

Recent registration approaches show the increasing importance of explicitly integrating *context information*, such as shape context [5], local self-similarity [25], and contextual flow [33], into registration. Moreover, the addition of *spatial information* into the similarity estimation, as it is *e.g.* done with higher-order densities [22], neighborhood patches [23], conditional mutual information [13], local volume densities [37], Markov random fields [36], and spatial-context mutual information [35], leads to improvements. The i.i.d. assumption of current frameworks, however, prohibits their consideration.

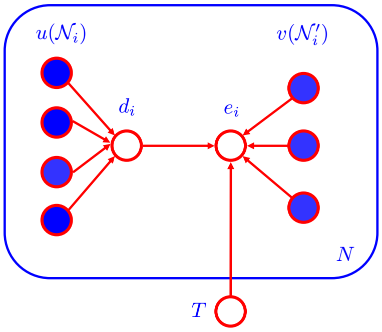

Figure 2. Contextual graphical model as plate. Descriptors d_i and e_i are dependent on a local neighborhood $\mathcal{N}_i$ of the original images. Further, e_i is dependent on d_i and the transformation T. Observed random variables are filled blue, links indicate dependency.

2.2. Contextual Probabilistic Graphical Model

The key component of the novel probabilistic framework is to replace the assumptions of independence of coordinate samples in equation (3), by the Markov property. This leads to a dependency of a pixel position on a local pixel neighborhood. One could think of a variety of possibilities for modeling the local neighborhood in a maximum likelihood framework. We decided to introduce two additional layers d and e, because it facilitates the representation of the neighborhood dependency. Each of the layers, we refer to as *description layers*, consists of latent random variables d_i and e_i, respectively, with $1 \leq i \leq N$ and $N = |\Omega|$. The layers d and e are lying on the same grid as the images do, so we have a dense set of descriptors. In our model, we let each descriptor d_i be dependent on a local neighborhood $\mathcal{N}_i$ of the image $u(\mathcal{N}_i)$, analogously, e_i is dependent on $v(\mathcal{N}'_i)$. The relationship between descriptors d_i and e_i is one-to-one. The creation of the layer e is dependent on the transformation T.

We utilize *probabilistic graphical models* [6] for establishing the relationship between random variables because they are advantageous in representing the structure and dependency for a multitude of variables. Further, we choose a *directed* graphical model, where nodes represent random variables and directed edges express probabilistic dependency between them. The graphical model for our framework is shown in figure 2 as plate with an exemplary 4- and 3-neighborhood, for u and v, respectively. Another illustration, not as plate, is shown in figure 3 without the consideration of T due to clarity of presentation. The presented

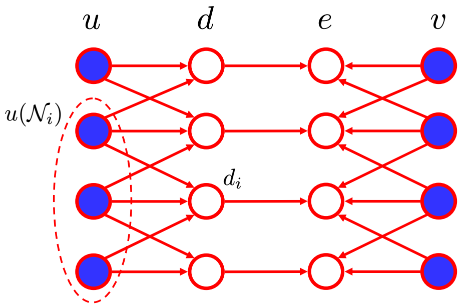

Figure 3. Graphical model of the registration framework. Observed 1D images u and v. Latent description layers d and e. $u(\mathcal{N}_i)$ neighborhood system of descriptor d_i.

graphical model factorizes to

$$\begin{aligned} p(u,v,d,e,T) =& p(T) \cdot \prod_{i=1}^{N} p(u(x_i)) \cdot p(v(T(x_i))) \\ & \cdot p(d_i|u(\mathcal{N}_i)) \cdot p(e_i|d_i, v(\mathcal{N}'_i), T). \end{aligned} \tag{4}$$

We deduce the term $p(e_i|d_i, v(\mathcal{N}'_i), T)$ further by applying the product rule and Bayes' theorem. Moreover, we assume the conditional independence of d_i and $v(\mathcal{N}'_i)$ with respect to e_i, and the independence of d_i and $v(\mathcal{N}'_i)$). Since the description layer e is separating the two layers d and v, the assumption is justified, leading to

$$\begin{aligned} p(e_i|d_i, v(\mathcal{N}'_i), T) &= \frac{p(d_i|e_i) \cdot p(v(\mathcal{N}'_i), T|e_i) \cdot p(e_i)}{p(d_i) \cdot p(v(\mathcal{N}'_i), T)} \\ &= \frac{p(e_i|d_i) \cdot p(e_i|v(\mathcal{N}'_i), T)}{p(e_i)}. \end{aligned} \tag{5}$$

Setting this result in equation (4) leads to

$$\begin{aligned} p(u,v,d,e,T) =& p(T) \cdot \prod_{i=1}^{N} p(u(x_i)) \cdot p(v(T(x_i))) \\ & \cdot p(d_i|u(\mathcal{N}_i)) \cdot \frac{p(e_i|d_i) \cdot p(e_i|v(\mathcal{N}'_i), T)}{p(e_i)}. \end{aligned} \tag{6}$$

Therein, the marginal terms $p(T)$, $p(u(x_i))$, $p(v(T(x_i)))$, and $p(e_i)^{-1}$ represent the probabilities for the transformation, the images, and the description layer. The reason that only the descriptor layer e appears in the formulation is rooted in the asymmetric formulation of the registration by only transforming the image v. This can be changed with a symmetric formulation, by transforming both images, which is shown in the supplementary material for the general case of groupwise registration.

The marginal terms are used to incorporate prior information into the registration, with the purpose of improving the robustness and capture range [38]. In most cases, we do not have any a priori knowledge about the probability distribution of these terms, so that we presume a uniform distribution, leading to

$$p(u,v,d,e,T) = \prod_{i=1}^{N} \underbrace{p(e_i|d_i)}_{\text{similarity}} \cdot \underbrace{p(d_i|u(\mathcal{N}_i)) \cdot p(e_i|v(\mathcal{N}'_i), T)}_{\text{coupling}}. \tag{7}$$

It is mainly the interplay of these three probabilities that determines the functionality of our model. The *similarity term* $p(e_i|d_i)$ is the standard likelihood function as used in previous unifying frameworks [20, 28, 39]. However, instead of comparing the images u and v, it compares the description layers d and e. If we could arbitrarily modify the description layers and just optimize $p(e|d)$, we would simply change the layers to be totally exact. The *coupling terms* $p(d_i|u(\mathcal{N}_i))$ and $p(e_i|v(\mathcal{N}'_i), T)$ prevent this simplistic solution by expressing how well the description layers fit the original images. In the optimization, they are counterbalancing the influence of the similarity term like a regularizer.

The joint distribution, $p(u,v,e,d,T)$, we finally end up with, is different from the one used in maximum likelihood (ML) frameworks, $p(u,v,T)$. This can, however, be obtained by marginalizing with respect to the descriptors

$$p(u,v,T) = \sum_{d,e} p(u,v,d,e,T). \tag{8}$$

Practically, it is not possible to sum over all possible descriptors. Thus, the alignment is only optimal with respect to a specific descriptor or a small set of descriptors, which is discussed further in section 3.3 on hybrid approaches.

3. A Continuum of Registration Approaches

In this section, we discuss several approaches for geometry- and intensity-based registration and show how they fit into the proposed framework. These methods form, in fact, a continuum of registration approaches, going all the way from pure geometric to pure iconic registration. On the one end, we identify landmark-based registration, where users manually pick salient points in the image. The description is optimal because we exactly know about the correspondence of points. On the other end, we identify intensity-based registration, with single intensity values as minimalistic descriptors. The number of approaches in between can be arranged by the uniqueness of their descriptors, as illustrates figure 4.

On the right-hand side of the spectrum, we consider SIFT and GLOH with comparatively high uniqueness of the descriptors. SIFT/GLOH correspondence hypotheses are created without location information, therefore descriptors must uniquely characterize the position they are extracted from. For DAISY [26], the dense arrangement of descriptors relaxes this requirement, equally for self-similarity [25]. Entropy images [31] extract structural information of images for multi-modal registration and re-

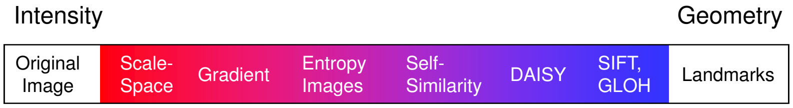

Figure 4. Continuum between feature- and intensity-based registration, augmented with exemplary approaches. Arranged by the uniqueness of descriptors.

semble gradient images. Scale-space images are close to the original ones, because locally weighted averages are created with an emphasis on the center location.

3.1. Intensity-Based Registration

Existing probabilistic frameworks for intensity-based registration focus on similarity measures and do not model common processing steps on the images. We demonstrate in the following how these steps can be integrated in the new framework. The proposed framework is a true extension of previous maximum likelihood frameworks, which can be obtained by setting $\mathcal{N}_i = (x_i)$, $d_i = u(x_i)$, $\mathcal{N}'_i = (T(x_i))$, and $e_i = v(T(x_i))$.

3.1.1 Image Filtering

Image filtering is a common pre-processing step for image registration. One application of filtering is image enhancement through operations such as sharpening, noise reduction, and contrast adjustment. Another application is the creation of a scale-space [12]. Although these processing steps are very popular, it has not yet been described under which conditions they are optimal choices.

With the proposed framework, it is possible to deduce optimal filters under the incorporation of certain assumptions, similar to the derivation of similarity measures. For this, we focus on the maximization of the coupling term $p(d|u)$ with all considerations analogously for $p(e|v,T)$. The MAP estimation reads as

$$\hat{d} = \max_d p(d|u) = \max_d \prod_{i=1}^{N} p(d_i|u(\mathcal{N}_i)) \tag{9}$$

$$= \max_d \prod_{i=1}^{N} \frac{p(u(\mathcal{N}_i)|d_i) \cdot p(d_i)}{p(u(\mathcal{N}_i))}, \tag{10}$$

where we can maximize for each d_i separately in the following. Further, applying the logarithm leads to

$$\begin{aligned}\hat{d}_i &= \max_{d_i} \{\log p(u(\mathcal{N}_i)|d_i) + \log p(d_i) - \log p(u(\mathcal{N}_i))\}\\ &= \max_{d_i} \left\{ \sum_{j\in\mathcal{N}_i} \log p(u_j|d_i) + \log p(d_i) \right\} - \sum_{j\in\mathcal{N}_i} \log p(u_j)\\ &\approx \max_{d_i} \left\{ \sum_{j\in\mathcal{N}_i} \log p(u_j|d_i) \right\} + \mathrm{H}[u(\mathcal{N}_i)], \end{aligned}\tag{11}$$

where we obtain the entropy H from the asymptotic equipartition property, which results from the application of the weak law of large numbers [8]. The entropy has no influence on the maximization. It is, however, interesting to notice that setting $d_i = \mathrm{H}[u(\mathcal{N}_i)]$ corresponds to the recently proposed entropy images for multi-modal registration [31].

Incorporating the assumption of a Gaussian noise into the maximization

$$\max_{d_i} \sum_{j\in\mathcal{N}_i} \log p(u_j|d_i) = \max_{d_i} \sum_{j\in\mathcal{N}_i} -\omega_j (d_i - u_j)^2, \tag{12}$$

with weights ω_j. Following standard maximum likelihood estimation [6] leads to the optimal solution for d. This estimation was extended to the usage of various norms, considering for instance least absolute values, instead of least squares. Further extensions resulted in M-estimators, and later, generalized M-estimators [10]. We consider in the following the minimization problem

$$\min_{d_i} \|\Lambda.(d_i\mathbf{1} - u(\mathcal{N}_i))\| \quad \text{with} \quad \Lambda = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_{|\mathcal{N}_i|} \end{pmatrix}, \tag{13}$$

a vector norm $\|.\|$, the one vector $\mathbf{1}$ of length $|\mathcal{N}_i|$, and weights Λ. Calculating the derivative with respect to d_i and setting it to zero leads to optimal descriptors. For different norms and $\Lambda = \mathbf{1}$, this results in the following descriptors:

Norm	Descriptor d_i
$\|.\|_2^2$	$\mathbb{E}[u(\mathcal{N}_i)]$
$\|.\|_1$	median$[u(\mathcal{N}_i)]$
$\|.\|_\infty$	$\frac{\max[u(\mathcal{N}_i)] - \min[u(\mathcal{N}_i)]}{2}$

As an example for least squares $\|.\|_2^2$ and arbitrary weights ω_j, we obtain

$$d_i = \frac{1}{\Pi} \sum_{j\in\mathcal{N}_i} \omega_j u_j \quad \text{with} \quad \Pi = \sum_{j\in\mathcal{N}_i} \omega_j. \tag{14}$$

Modifying the weights in this case allows for modeling arbitrary linear filters. For creating linear scale-spaces, the weights correspond the entries of a Gaussian filter mask.

3.1.2 Gradient-Based Similarity Measures

Gradient-based similarity measures are, for instance, of interest in 2D-3D registration [17]. Example metrics are gradient correlation and gradient difference. The gradients are calculated with the Sobel operator represented as 3×3 filter mask. Subsequently, the correlation coefficient or difference is evaluated between the gradients of the images. For modeling the Sobel operator in the maximum likelihood framework, as described in section 3.1.1, we have to adapt equation (14), because the weights for differential operators sum up to zero. Consequently, we do not consider the normalization factor Π and set the weights ω_j according to the Sobel mask. The description layers of our framework represent the gradient images, which are successively matched. In a more recent article, Shams *et al.* [24] propose gradient intensity-based registration, where mutual information between the gradient images is calculated. The description layers for both registration approaches [17, 24] are the same, it is only the metric that is changing. This shows the increased modularity provided by our framework due to the explicit consideration of description layers.

3.2. Geometry-Based Registration

The integration of geometric registration in our framework corresponds to embedding the feature points on a dense grid. Once the descriptors are calculated for each image, the next step is the comparison between the images. Looking at the approaches for geometry-based registration, we observe that typically SSD is evaluated between the descriptors, which is derived from the similarity term $p(e|d)$. The difference to intensity-based registration is, however, the focus on certain keypoint locations. To account for this, we introduce indicator variables $k_i, l_i \in \{0, 1\}$, where $k_i = 1$ signifies that descriptor d_i is located at a keypoint location, analogously for l_i and e_i. Location i is only considered if it corresponds to keypoint locations in both images, leading the the approximation of the probability

$$p(.) \propto \prod_{i=1}^{N} \left(1 + \underbrace{p(e_i|d_i)}_{\text{similarity}} \cdot \underbrace{p(d_i|u(\mathcal{N}_i))p(e_i|v(\mathcal{N}'_i))}_{\text{coupling}} \right)^{\overbrace{k_i \cdot l_i}^{}\underbrace{}_{\text{keypoint}}} . \tag{15}$$

In the following, we describe landmark- and feature-based registration in more details.

3.2.1 Landmark-Based Registration

The term landmark-based registration is ambiguously used in the literature, where we consider it in the sense that experts identify the location of the keyoint and also provide a distinctive description. Most important is the probability $p(e_i \mid d_i)$, which evaluates the similarity that locations with the same labels overlap. The terms $p(d_i \mid u(\mathcal{N}_i))$ and $p(e_i \mid v(\mathcal{N}'_i))$ can be used to model the confidence in the assignment of the label to the keypoint location. For the keypoint locations the values of k_i and l_i are set to one.

3.2.2 Feature-Based Registration

While in landmark-based registration, the localization and description of the keypoints takes place manually, and for point clouds, the localization is automatic but no description is provided, feature-based registration performs the extraction as well as the description automatically. The first task, the *keypoint localization*, is to identify locations that can repeatedly be assigned under different views of the same object. Popular methods include the difference-of-Gaussian (DoG) [14], Harris detector, Harris-affine, and Hessian-affine detector [15]. Depending on the output of these detectors the keypoint variables k_i and l_i are set. The second step, the *feature description*, has to represent the characteristics of the point within its local neighborhood. Frequently used image descriptors are *e.g.* Scale-Invariant Feature Transform (SIFT) [14], Speeded-Up Robust Features (SURF) [4], and Gradient Location and Orientation Histogram (GLOH) [16]. The descriptors are assigned to the corresponding locations on the description layers. The last step is the *feature matching*, where descriptors of both images at the corresponding locations are compared.

In our framework, the terms $p(d_i|u(\mathcal{N}_i))$ and $p(e_i|v(\mathcal{N}'_i))$ are applied for the deduction and calculation of the descriptors from the images. They ensure that the descriptors well characterize the local image context. $p(e_i|d_i)$ expresses the similarity of descriptors. k_i and l_i restrict the evaluation to keypoint locations.

Looking at the feature-based approaches, we clearly see the local nature of these techniques. Considering SIFT as an example, the keypoint localization with DoG is achieved by searching for the local maximum in scale-space. The DoG can be modeled by setting the appropriate weights in the linear filtering in equation (14). The maximum search only considers the direct neighbors. The SIFT descriptor uses 4×4 blocks around the keypoint, where each block consists of 4×4 pixels of the corresponding scale-space level. So in total, a 16×16 neighborhood of each keypoint is considered for building the descriptor. This shows that the descriptors are built using the local context. In our ML framework we are able to describe them due to the extension with neighborhood information and the integration of latent layers.

3.3. Hybrid Methods

Hybrid registration approaches combine multiple alignment techniques to achieve an improved registration result. So far, it has not been possible to describe hybrid approaches that combine techniques from geometric and iconic registration in a common framework, because there was no framework that enabled the modeling of both registration approaches. As seen in sections 3.1 and 3.2, the proposed probabilistic framework enables the description of a multitude of registration techniques by choosing different descriptors. A possible differentiation of hybrid approaches is to distinguish between the consecutive application of registration [2, 11, 21], or the coupling to a joint energy formulation [7, 18, 30, 32]. For the joint formulation, we consider the sets of descriptors $\mathcal{D}$ and $\mathcal{E}$, which can contain descriptors from geometric registration, such as SIFT, and from iconic registration, such as entropy and gradient information. The final marginalization is similar to equation (8)

$$p(u, v, T) \approx \sum_{d \in \mathcal{D}, e \in \mathcal{E}} p(u, v, d, e, T). \tag{16}$$

Since we marginalize only over finite sets and not all possible descriptors, we only achieve an approximation.

4. Dynamic Adaption of Description Layers

In the last section, we showed how registration techniques can be modeled with the proposed framework. Further, we illustrated a continuum of registration approaches, classified by the uniqueness of their descriptors. We achieve this increased flexibility by introducing layers of latent random variables. For the approaches in the last section, these layers were calculated with various deterministic algorithms and did not change during the registration. In this section, we illustrate the second advantage of the our model, the dynamic adaption of the description layers. Instead of reducing the optimization to the similarity term $p(e|d)$, we now rely on the interaction of coupling and similarity terms.

4.1. Joint Registration and Segmentation

Fundamental operations in image analysis include the segmentation and registration of images. Although they are most times solved separately, there are applications where they can mutually benefit from each other and accordingly a joint formulation is useful [1, 19, 27, 34]. The performance of any segmentation approach is primarily dependent on the discriminative power of the underlying likelihood model for the data [34]. Multiple measurements with different imaging modalities or viewpoints could therefore improve segmentation. On the other hand, the alignment of segmented images, instead of the original ones, significantly reduces the influence of noise and consequently facilitates the registration. In our framework, the description layers represent the segmented images. The similarity term $p(e|d)$ drives the correct global alignment and also provides the combination of both image segmentations. The coupling terms $p(d|u)$ and $p(e|v, T)$ counterbalance the effect of letting both segmentations looking as similar as possible, by ensuring the segmentations to be close to the underlying data.

We show how the MAP MRF approach in [34] naturally integrates into our framework. The MAP problem is stated using Bayes

$$p(d, e, T|u, v) = \frac{p(u, v|d, e, T)p(d, e, T)}{p(u, v)} \tag{17}$$

with the images u, v given and the transformation T and segmentations d, e to calculate. The likelihood term $p(u, v|d, e, T)$ is represented with a Gaussian mixture model and the prior $p(d, e, T)$ with an MRF using the Ising model. At the beginning of the registration, when the images are far from being correctly aligned, the joint modeling of both images is not meaningful. Therefore, the independence of the images and consequently the labels is assumed, leading to

$$p(d, e, T|u, v) = \frac{p(u|d)p(d)p(v|e)p(e)p(T)}{p(u)p(v)}. \tag{18}$$

For the joint optimization, an alternation is done between solving for the optimal labeling with iterated conditional modes and the alignment with the Powell method.

4.2. Registration with Dynamic Adaptation

Next to the analysis of existing registration algorithms, we also want to illustrate the deduction of new methods with the proposed framework. We create an instantiation of the framework by assigning specific distributions to the involved probabilities. More precisely, we consider the problem of images being distorted by severe artifacts. An application where the registration of such images is required is endovascular stent graft placement, where the bright stent is only present in the intra-operative radiograph [3]. Instead of pre-processing images as in [3], we deduce a registration algorithm that is robust to such artifacts. This is advantageous because we retain all the information during the registration and the algorithm identifies artifacts automatically.

We incorporate a Gaussian distribution for the similarity term $p(e|d)$ leading to $p(e|d) \propto \exp\left(-\frac{\|e-d\|_2^2}{\sigma^2}\right)$, with variance σ^2. We allow the description layers to change during the registration, integrating the local neighborhood information, by setting a uniform distribution

$$p(d_i|u(\mathcal{N}_i)) \propto \begin{cases} c & d_i \in u(\mathcal{N}_i) \\ 0 & d_i \notin u(\mathcal{N}_i) \end{cases} \tag{19}$$

with constant c. Since we are interested in MAP estimation, the partition function plays no role in the optimization. Instead of assigning a constant likelihood to all patch locations, one could also choose weights that favor the selection of locations close to the patch center.

For simplicity of presentation, we allow the dynamic adaptation only on layer d. The log-likelihood of equation (7) reads as

$$\log p(u, v, d, e) \propto \sum_{i=1}^{N} -\frac{(e_i - d_i)^2}{\sigma^2} + \begin{cases} \log c & d_i \in u(\mathcal{N}_i) \\ -\infty & d_i \notin u(\mathcal{N}_i) \end{cases} \tag{20}$$

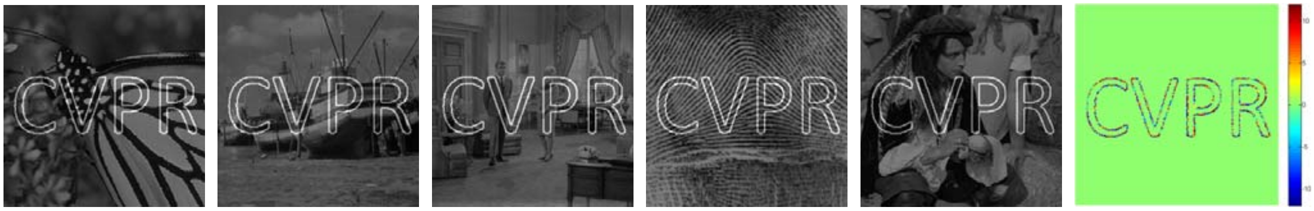

Figure 5. Five images with artifacts. Right: Illustration of selected locations in each local neighborhood. 0 corresponds to center location.

Due to the $-\infty$ cost, this is equivalent to the restriction of d_i to values in $u(\mathcal{N}_i)$. The optimization formulates as

$$\max_{T,d} -\sum_{i=1}^{N}(e_i - d_i)^2 \qquad \text{subject to} \quad d_i \in u(\mathcal{N}_i), \quad (21)$$

with e depending on T. We optimize simultaneously over the layer d, which can select values in the local neighborhoods, and the transformation T, which affects the layer e. At each step of the Nelder-Mead simplex optimizer, we set that value in the local neighborhood $u(\mathcal{N}_i)$ to d_i that minimizes the squared difference $(d_i - e_i)^2$. In our experiments, we consider a 5 × 5 neighborhood. For a 1-neighborhood, $u(\mathcal{N}_i) = \{u(x_i)\}$, the algorithm reduces to the standard SSD registration.

We perform rigid registration experiments on five images, see figure 5. We set as the moving image v the original, noise-free image. For u, we add the CVPR artifact and white Gaussian noise to the image. Further, we displace u, in our experiments by 10 pixel along the vertical direction. This is the true transformation that we want to recover. We start the registrations from random initial translations along x and y axes, guaranteeing a root mean squared (RMS) distance from the true transformation of 30 pixel. For each image, we run the registration 100 times from the random positions and calculate the RMS error between the registration result and the true transformation. A statistical analysis of the errors is presented in a box-and-whisker diagram in figure 6. We compare the approach to the registration with SSD, NCC, and MI [1]. Our results show that the addition of artifacts significantly influences the performance of SSD, NCC, and MI. In contrast, the proposed algorithm with the adaptation of description layers leads to excellent results. Further, we illustrate in figure 5 the location of the local neighborhood $\mathcal{N}_i$ that is assigned to d_i in the final step of the optimization. Illustrated is the assignment for the registration with the butterfly image, with similar results for the other images. The values range from -12 to 12, because it corresponds to the vector indexing of the 5 × 5 patch. 0 is the center location. We observe that across the image mainly the central location is selected. For the artifact region, however, locations in the neighborhood are selected to maximize the cost function, as expected. This information can be of value for further processing steps.

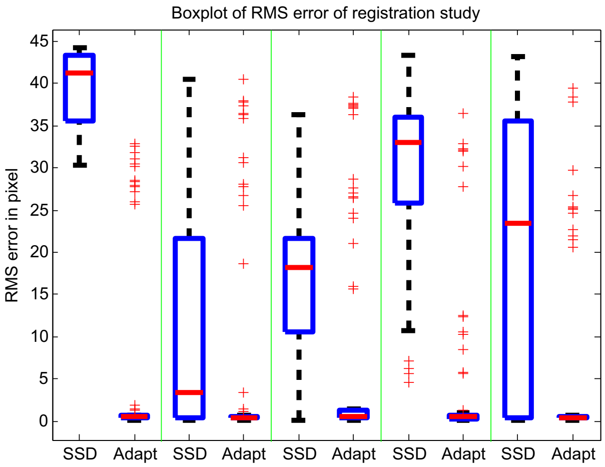

Figure 6. Results of random registration study. The order of the results is corresponding to the order of the images in figure 5.

5. Conclusion

We presented a novel probabilistic framework for image registration, which is general enough to describe intensity-based, as well as, geometry-based registration. The proposed framework allows us to move from just modeling the similarity function towards modeling larger parts of the registration process. The key extension, with respect to previous frameworks, is the consideration of local neighborhood information, so replacing the assumption of independent coordinate samples by the Markov property. We reviewed various registration approaches and showed their deduction within our framework. We further introduced a continuum of registration approaches, limited by pure geometric and iconic registration, arranged by the uniqueness of their descriptors. We used the coupling terms in the proposed framework to derive optimal descriptors, as well as, to integrate the dynamic adaption of descriptors during the registration. Finally, we instantiated the framework with specific distributions to deduce a novel registration algorithm. The proposed framework provides further insights about the relationship of various registration techniques, and moreover, helps to understand and classify them.

[1] See supplementary material for results on NCC and MI. We plot SSD because it performed best.

In the supplementary material, we present the extension of the proposed framework to groupwise registration and additional experimental results.

Acknowledgment: The work was partly supported by the Humboldt foundation and European Commission. We thank Sandy Wells for discussions.

References

[1] J. Ashburner and K. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005. 6

[2] A. Azar, C. Xu, X. Pennec, and N. Ayache. An interactive hybrid non-rigid registration framework for 3d medical images. In *ISBI*, pages 824–827, April 2006. 6

[3] M. Baust, S. Demirci, and N. Navab. Stent graft removal for improving 2d-3d registration. In *ISBI*, 2009. 6

[4] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008. 5

[5] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 2002. 2

[6] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006. 2, 4

[7] T. Brox, C. Bregler, and J. Malik. Large Displacement Optical Flow . In *CVPR*, 2009. 6

[8] T. Cover, J. Thomas, J. Wiley, et al. *Elements of information theory*, volume 6. Wiley Online Library, 1991. 4

[9] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. A unified framework for atlas based brain image segmentation and registration. In *WBIR*, pages 136–143, 2006.

[10] D. Hoaglin, F. Mosteller, and J. Tukey. *Understanding robust and exploratory data analysis*, volume 3. Wiley New York, 1983. 4

[11] H. Johnson and G. Christensen. Consistent landmark and intensity-based image registration. *IEEE TMI*, 21(5):450–461, 2002. 6

[12] J. J. Koenderink. The structure of images. *Biological Cybernetics*, V50(5):363–370, 1984. 4

[13] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens. Nonrigid image registration using conditional mutual information. *IEEE TMI*, 29(1):19 –29, 2010. 2

[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 5

[15] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004. 5

[16] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *PAMI*, 27(10):1615–1630, 2005. 5

[17] G. Penney, J. Weese, J. Little, P. Desmedt, D. Hill, and D. Hawkes. A comparison of similarity measures for use in 2-d-3-d medical image registration. *IEEE TMI*, 17(4):586–595, Aug. 1998. 4, 5

[18] J. Pluim, J. Maintz, and M. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE TMI*, 19(8):809–814, 2000. 6

[19] K. Pohl, J. Fisher, W. Grimson, R. Kikinis, and W. Wells. A Bayesian model for joint segmentation and registration. *Neuroimage*, 31(1):228–239, 2006. 6

[20] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80, 2000. 1, 2, 3

[21] K. Rohr, P. Cathier, and S. Wörz. Elastic registration of electrophoresis images using intensity information and point landmarks. *Pattern recognition*, 37(5):1035–1048, 2004. 6

[22] D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes. Non-rigid registration using higher-order mutual information. In *SPIE*, volume 3979, pages 438–447, 2000. 2

[23] D. B. Russakoff, C. Tomasi, T. Rohlfing, C. R. Maurer, and Jr. Image similarity using mutual information of regions. In *ECCV*, pages 596–607. Springer, 2004. 2

[24] R. Shams, R. A. Kennedy, P. Sadeghi, and R. Hartley. Gradient intensity-based registration of multi-modal images of the brain. In *ICCV*, Oct. 2007. 5

[25] E. Shechtman and M. Irani. Matching Local Self-Similarities across Images and Videos. *CVPR*, 2007. 2, 3

[26] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *CVPR*, 2008. 3

[27] C. Vidal and B. Jedynak. Learning to match: Deriving optimal template-matching algorithms from probabilistic image models. *IJCV*, 88(2):189–213, 2010. 6

[28] P. A. Viola. *Alignment by Maximization of Mutual Information*. Ph.d. thesis, Massachusetts Institute of Technology, 1995. 1, 2, 3

[29] C. Wachinger, T. Klein, and N. Navab. Locally adaptive nakagami-based ultrasound similarity measures. *Ultrasonics*, 52(4):547 – 554, 2012. 2

[30] C. Wachinger and N. Navab. Alignment of viewing-angle dependent ultrasound images. In *MICCAI*, 2009. 6

[31] C. Wachinger and N. Navab. Entropy and laplacian images: Structural representations for multi-modal registration. *Medical Image Analysis*, 16(1):1 – 17, 2012. 3, 4

[32] M. Wacker and F. Deinzer. Automatic robust medical image registration using a new democratic vector optimization approach with multiple measures. In *MICCAI*, pages 590–597, 2009. 6

[33] Y. Wu and J. Fan. Contextual Flow. In *CVPR*, 2009. 2

[34] P. Wyatt and J. Noble. MAP MRF joint segmentation and registration of medical images. *Medical Image Analysis*, 7(4):539–552, 2003. 6

[35] Z. Yi and S. Soatto. Multimodal registration via spatial-context mutual information. In *IPMI*, 2011. 2

[36] G. Zheng. Effective incorporation of spatial information in a mutual information based 3d-2d registration of a ct volume to x-ray images. In *MICCAI*, 2008. 2

[37] X. Zhuang, D. Hawkes, and S. Ourselin. Unifying encoding of spatial information in mutual information for nonrigid registration. In *IPMI*, volume 5636, pages 491–502. 2009. 2

[38] L. Zöllei. *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images*. Ph.d. thesis, MIT; MIT-CSAIL, 2006. 1, 3

[39] L. Zöllei, J. Fisher III, and W. Wells III. A Unified Statistical and Information Theoretic Framework for Multi-modal Image Registration. In *IPMI*, 2003. 2, 3