Nonparametric Priors on the Space of Joint Intensity Distributions for Non-Rigid Multi-Modal Image Registration

Daniel Cremers
Department of Computer Science
University of Bonn, Germany

Christoph Guetter, Chenyang Xu
Department of Imaging and Visualization
Siemens Corporate Research, Princeton, NJ

Abstract

The introduction of prior knowledge has greatly enhanced numerous purely low-level driven image processing algorithms. In this work, we focus on the problem of non-rigid image registration. A number of powerful registration criteria have been developed in the last decade, most prominently the criterion of maximum mutual information. Although this criterion provides for good registration results in many applications, it remains a purely low-level criterion. As a consequence, registration results will deteriorate once this low-level information is corrupted, due to noise, partial occlusions or missing image structure. In this paper, we will develop a Bayesian framework that allows to impose statistically learned prior knowledge about the joint intensity distribution into image registration methods. The prior is given by a kernel density estimate on the space of joint intensity distributions computed from a representative set of pre-registered image pairs. This nonparametric prior accurately models previously learned intensity relations between various image modalities and slice locations. Experimental results demonstrate that the resulting registration process is more robust to missing low-level information as it favors intensity correspondences statistically consistent with the learned intensity distributions.

1. Introduction

Image registration is one of the fundamental problems of computer vision, with applications ranging from motion estimation, superresolution, and shape matching to medical image analysis. Classical motion estimation approaches [8] are based on the assumption that corresponding pixels have similar intensity values. In practice, this assumption is often violated: In the context of motion estimation, lighting changes from one frame to the next may cause intensity changes. In the field of medical image analysis, the registration of images obtained with different modalities (e.g. CT and MRI) may require far more sophisticated matching constraints, since the two modalities may assign different intensities to the same medical structure.

In recent years, the concept of *Maximum Mutual Information (MI)* has become established as a powerful criterion for image registration [3, 17, 15, 11]. The key idea is to find a displacement field $\hat{u}$ that maximizes the statistical dependency between the intensity distributions of the two images:

$$\hat{u} = \arg\max_u \mathcal{I}_{MI}\big(f_1(x), f_2(x+u(x))\big), \tag{1}$$

where f_1 and f_2 are the two images and $\mathcal{I}_{MI}$ denotes the mutual information of the two distributions. This can be written as:

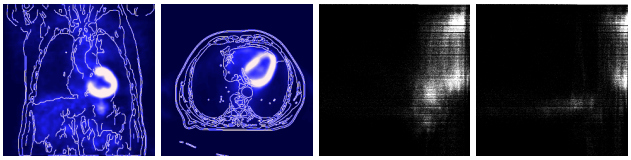
$$\mathcal{I}_{MI}\big(f_1(x), f_2(x+u(x))\big) = \int_{\mathbb{R}^2} p_u(i_1,i_2) \log \frac{p_u(i_1,i_2)}{p_{f_1}(i_1) p_{f_2}(i_2)} di_1 di_2, \tag{2}$$

where $i_1 = f_1(x)$, $i_2 = f_2(x+u(x))$, and $p_{f_1}(i_1)$, $p_{f_2}(i_2)$, $p_u(i_1,i_2)$ are the marginal and joint intensity distributions estimated from $f_1(x)$ and $f_2(x+u(x))$.

Maximizing constraint (1) will favor correspondence between pixels that is no longer dictated by the assumption that their intensities should be similar, but by the assumption that the resulting distributions of the intensities of matched pixels should be maximally dependent. The correspondence of intensity pairs therefore arises purely from the two matched images.

In most practical applications, however, this low-level information is corrupted or incomplete. The two matched views may be deteriorated by noise, or certain parts may be missing or occluded — in the context of medical images for example due to the presence of a tumor in one of the two images. In this case, the corrupted low-level information will be insufficient to accurately determine the correct intensity correspondence. The matching of the occluded region to respective image areas in the other image may even deteriorate the estimated intensity transformation between the matched images and hence it will bias the simultaneously estimated non-rigid transformation. Moreover, the local optimization by gradient descent requires an initial estimate of the intensity transformation between the two images. If the

0-7695-2646-2/06 $20.00 (c) 2006 IEEE



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

Figure 1. Different registration problems are characterized by different joint intensity distributions. (a) Registration of coronal PET and CT slices, (b) registration of axial PET and CT slices. (c) Joint histogram for the coronal registration problem, (d) joint intensity histogram for the axial registration problem.

initialization is too far from the final matching, then the initial estimate of the intensity transformation may be too far from the true one to allow the correct registration of the two structures. At the same time, one may have prior knowledge about which intensities are more or less likely to be in correspondence. It turns out that non-rigid registration can be strongly improved by reverting to such prior knowledge.

In the context of image registration such prior knowledge has hitherto been imposed on the displacement field T. Beyond a number of fairly sophisticated regularity constraints such as non-quadratic smoothness priors [16], Roth and Black [13] recently suggested to actually learn the statistics of optical flow from training sequences and to subsequently impose these as priors for variational motion estimation. Our work differs from theirs in that we propose to learn and impose statistical priors not on the displacement field u, but on the simultaneously estimated joint intensity distribution p_u characterizing the intensity transformation from one image to the other.

In the context of medical image registration, Leventon and Grimson introduced a joint intensity distribution as a prior for rigid registration [10]. Zöllei et al. showed that this method makes some implicit assumptions about the desired solution which do not always hold [18]. Chung et al. [2] found empirically that the Kullback-Leibler (KL) divergence is superior to the log likelihood used by Leventon and Grimson. Guetter et al. [6] further developed this idea, introducing the KL divergence with respect to a known distribution into the energy functional for non-rigid image matching process, which leads to a matching algorithm that favors a specific joint intensity distribution. While the additional term provides for improved matching results, its introduction is rather add-hoc. Moreover, it merely allows to impose a *single* learned joint intensity distribution. In the context of medical images, we found that a single joint intensity distribution is not sufficient to describe the variability of observed intensity correspondences: Given a set of pairs of matched images from different modalities, one finds great variations among the estimated joint distributions. Figure 1 shows coronary slices as obtained with a PET and a CT scanner, registrations of this pair of coronary slices and of

KLD	PET/CT Whole Body	PET/CT Lungs	SPECT/CT Kidneys
PET/CT Whole Body	0.0	0.7740	3.9609
PET/CT Lungs	0.4871	0.0	3.8275
SPECT/CT Kidneys	2.6614	2.5604	0.0

Table 1. KL divergences (KLD) for sample medical data showing the dissimilarity between joint intensity distributions (each of which was computed from the registration of respective image/volume pairs), as shown in Figure 1. The matching of slices requires different priors on the intensity correspondence, depending on which imaging modalities, which slice locations and which acquisition protocols are used.

a respective pair of axial slices. The inferred joint intensity histograms characterizing the intensity correspondence look quite different. Table 1 shows the KL distances between pairs of joint intensity distributions, each of which is estimated from a registered pair of medical images. These observed variations are due to different pairings of imaging modalities (PET, CT, SPECT), different acquisition protocols, or simply due to the selection of slices from different areas of the same scan, i.e. to a variation in the field of view.

In this paper, we will develop a statistical framework that allows to impose statistically learned prior knowledge about the joint intensity distribution into image registration methods. We assume that we are given an entire set of correctly registered image pairs. From these we can compute respective joint intensity distributions and construct a nonlinear statistical prior given by a kernel density estimate on the space of joint intensity distributions. It can be introduced into the registration process in the framework of Bayesian inference. As a consequence, the subsequent image matching process is not only driven by a maximization of statistical dependence of the individual intensity distributions, but it will also favor matching results for which the resulting joint intensity distribution is statistically consistent with the set of learned joint intensity distributions.

2. Image Registration as Bayesian Inference

Assume we are given a representative set of pre-registered image pairs $\{f_1^j, f_2^j\}_{j=1,\ldots,m}$, where $f_k^j : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$. These image pairs may be obtained from various image modalities and slice locations. Each registered image pair gives rise to a specific joint intensity distributions $p_j(i_1, i_2)$, stating which intensities i_1 and i_2 are likely to be in correspondence for the given image pair. The goal of the present paper is to derive means to impose this knowledge into variational image registration algorithms.

In the framework of Bayesian inference, the problem of

nonrigid image registration can be solved by finding the most likely displacement field u and joint intensity distribution p_u, given the two images f_1 and f_2 and given the set of previously learned joint intensity distributions $\{p_j\}_{j=1,\dots,m}$, under the constraint that the joint intensity distribution p_u is given by the one arising from $f_1(x)$ and $f_2(x+u)$. That is we propose to maximize the conditional distribution

$$\begin{aligned}\mathcal{P}\big(u,\, p_u \,\big|\, f_1, f_2, \{p_j\}\big) \\ &\propto \mathcal{P}\big(f_1, f_2 \,\big|\, u, p_u, \{p_j\}\big)\; \mathcal{P}\big(u, p_u \,\big|\, \{p_j\}\big) \quad (3)\\ &\propto \mathcal{P}\big(f_1, f_2 \,\big|\, u, p_u\big)\; \mathcal{P}(u)\; \mathcal{P}\big(p_u \,\big|\, \{p_j\}\big),\end{aligned}$$

with respect to the displacement field u. Proportionality in the above expressions means that we have only neglected factors which do not depend on the displacement field u and thus do not affect the maximization. In the second step in (3), we have made the assumption that the prior decouples into a geometric prior $\mathcal{P}(u)$ on the displacement field and a prior on the joint intensity distribution p_u.

Thus the optimization problem in (3) separates into three factors, which can be interpreted as follows: The first factor provides the measurement likelihood, stating how likely the two images are given the correspondence induced by the displacement field u. The second factor in (3) indicates the a priori probability of a displacement field u. And the last factor specifies how consistent the estimated joint intensity distribution is with respect to the previously learned ones. We propose to model these expressions as follows.

2.1. Consistency with Learned Distributions

Given a set of joint intensity distributions $\{p_j\}_{j=1,\dots,m}$ obtained from a set of optimally registered image pairs, we can revert to concepts from kernel density estimation [14, 4, 9, 5] in order to derive the following prior on the space of joint intensity distributions[1]:

$$\mathcal{P}(p_u \mid \{p_j\}) \propto \frac{1}{m} \sum_{j=1}^{m} \exp\left(-\frac{\mathcal{I}_{KL}(p_u, p_j)}{\sigma}\right), \quad (4)$$

where

$$\mathcal{I}_{KL}(p_u, p_j) = \int_{\mathbb{R}^2} p_u(i_1, i_2) \log \frac{p_u(i_1, i_2)}{p_j(i_1, i_2)}\, di_1 di_2 \quad (5)$$

denotes the KL divergence measuring the dissimilarity between the intensity distribution p_u (induced by matching f_1 and f_2 under the displacement u) and the previously learned joint distribution p_j. In the optimization of (3), the distribution (4) therefore imposes statistical similarity between the infered intensity correspondence p_u and the previously observed joint intensity distributions $\{p_j\}_{j=1,\dots,m}$. The kernel width σ in the density estimator is fixed to the average nearest neighbor distance computed for the set of joint intensity distributions $\{p_j\}$:

$$\sigma = \frac{1}{m} \sum_{i=1}^{m} \min_{j \neq i} \mathcal{I}_{KL}(p_i, p_j) \quad (6)$$

More sophisticated estimates, for example using cross validation, are conceivable, we refer the reader to [14].

2.2. Mutual Information Maximization

To model the second factor in (3) we revert to the well-known concept of maximal mutual information, by stating that two images f_1 and f_2 are more likely to be aligned if the two distributions of corresponding intensities $f_1(x)$ and $f_2(x+u)$ are more dependent:

$$\mathcal{P}(f_1, f_2 \,|\, u, p_u) \propto \exp\Big(\alpha_1 \mathcal{I}_{MI}\big(f_1(x), f_2(x+u)\big)\Big), \quad (7)$$

with the mutual information $\mathcal{I}_{MI}$ defined in (2). In the statistical inference (3) this constraint will favor displacement fields u that maximize the statistical dependency between the two intensity distributions.

2.3. Smoothness Prior on the Displacement Field

The last factor in (3) allows to impose a prior on the displacement field u stating which displacement fields are *a priori* more or less likely. As proposed in [13], one could also learn such priors from training sequences of optic flow fields. Since our contribution is the statistical modeling of priors on the intensity transformation between the two images, we shall merely impose a common smoothness prior on the displacement field:

$$\mathcal{P}(u) \propto \exp\left(-\alpha_2 \int |\nabla u|^2\, dx\right). \quad (8)$$

This smoothness prior was pioneered in the seminal work of Horn and Schunck [8]. More sophisticated priors are conceivable, for example non-quadratic (robust) smoothness priors that allow for discontinuities in the estimated displacement fields (cf. [1]).

3. Variational Formulation

Now that the three factors in the inference problem (3) are specified, we can maximize this probability by minimizing its negative logarithm, which is given by an energy of the form:

$$E(u) = E_{prior}(u) \,+\, \alpha_1\, E_{MI}(u) \,+\, \alpha_2\, E_{smooth}(u). \quad (9)$$

These three energies impose several constraints: The energy E_{prior} guarantees that the joint intensity distribution

[1] A theoretical basis of probability distributions on infinite-dimensional function spaces is provided by the theory of Gaussian processes [12].

induced by a displacement field u is consistent with previously observed joint intensity distributions. According to (4), it is given by:

$$E_{prior}(u) = -\log\left(\sum_{j=1}^{m} \exp\left(-\frac{\mathcal{I}_{KL}(p_u, p_j)}{\sigma}\right)\right). \quad (10)$$

The second term in (9) yields the well-known mutual information criterion:

$$E_{MI}(u) = -\mathcal{I}_{MI}\left((f_1(x), f_2(x+u)\right), \quad (11)$$

and the prior on the displacement field u gives the smoothness constraint:

$$E_{smooth}(u) = \int |\nabla u|^2 \, dx. \quad (12)$$

Minimization of the energy (9) by gradient descent leads to a partial differential equation for u of the form:

$$\begin{aligned} \frac{\partial u}{\partial t} &= -\frac{\partial E(u)}{\partial u} \quad (13) \\ &= -\frac{\partial E_{prior}(u)}{\partial u} - \alpha_1 \frac{\partial E_{MI}(u)}{\partial u} - \alpha_2 \frac{\partial E_{smooth}(u)}{\partial u}. \end{aligned}$$

For the gradient of I_{MI}, we refer to Hermosillo et al. [7]. The gradient of the smoothness constraint leads to a well-known diffusion term $\triangle u$.

The gradient of E_{prior} is given by[2]:

$$\frac{\partial E_{prior}(u)}{\partial u} = \frac{1}{\sigma} \sum_{j=1}^{m} \gamma_j \frac{\partial \mathcal{I}_{KL}(p_u, p_j)}{\partial u}, \quad (14)$$

with normalized weights

$$\gamma_j = \frac{\hat{\gamma}_j}{\sum_i \hat{\gamma}_i}, \quad (15)$$

where:

$$\hat{\gamma}_j = \exp\left(-\frac{\mathcal{I}_{KL}(p_u, p_j)}{\sigma}\right). \quad (16)$$

The remaining challenge is to compute the gradient of $\mathcal{I}_{KL}(p_u, p_j)$ with respect to the displacement field u. To this end, we will make use of the Gateaux derivative giving the gradient in direction $\tilde{u}$:

$$\frac{\partial \mathcal{I}_{KL}}{\partial u}\Big|_{\tilde{u}} = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\Big(\mathcal{I}_{KL}(p_{u+\epsilon\tilde{u}}, p_j) - \mathcal{I}_{KL}(p_u, p_j)\Big) \quad (17)$$

Using the definition of the joint intensity distribution

$$p_u(i_1, i_2) \equiv \frac{1}{|\Omega|} \int_{\Omega} \mathrm{G}_\rho\Big(i_1 - f_1(x), i_2 - f_2(x+u)\Big) dx$$

where G_ρ is a two-dimensional Gaussian distribution of width ρ, a straight-forward computation shows that:

$$p_{u+\epsilon\tilde{u}} = p_u + \epsilon \int \tilde{u} \, \frac{\partial G_\rho}{\partial i_2}(i_1 - f_1, i_2 - f_2) \nabla f_2 \, dx, \quad (18)$$

where f_2 and ∇f_2 are evaluated at $x + u(x)$. Inserting the expansion (18) into (17) and further linearization gives the directional derivative

$$\frac{\partial \mathcal{I}_{KL}(p_u, p_j)}{\partial u}\Big|_{\tilde{u}} = \int \left(\frac{\partial \mathcal{I}_{KL}(p_u, p_j)}{\partial u}\right) \tilde{u}(x) \, dx, \quad (19)$$

with the gradient given by:

$$\begin{aligned} &\frac{\partial \mathcal{I}_{KL}(p_u, p_j)}{\partial u} = \frac{1}{|\Omega|} \cdot \\ &\left[\mathrm{G}_\rho * \left(\frac{\partial_{i_2} p_u(i_1, i_2)}{p_u(i_1, i_2)} - \frac{\partial_{i_2} p_j(i_1, i_2)}{p_j(i_1, i_2)}\right)\right](f_1, f_2) \cdot \nabla f_2, \end{aligned}$$

where as above f_2 and ∇f_2 are evaluated at $x + u(x)$.

The interpretation of the additional term (14) in the evolution of the displacement field u is quite intuitive: It induces a change in the estimated displacement field u that aims at minimizing the KL-distance $\mathcal{I}_{KL}(p_u, p_j)$, thereby drawing the current intensity distribution p_u toward the previously learned distributions $\{p_j\}$. More precisely, the energy gradient exerts a force on the estimated intensity distribution toward each learned intensity distribution p_j, which is modulated by a weight γ_j that decays exponentially with the distance between the intensity distributions — see equation (16). Thus this additional term comes into play only for those learned distributions that are most consistent with the currently estimated intensity distribution. And this is precisely the mechanism by which the algorithm "decides" which intensity distributions among the learned ones are to be used for a given registration task.

To further clarify this effect of the multimodal energy (10) we refer to the visualization in Figure 2: In this schematic drawing, each joint intensity distribution p_j is represented as a black 2D point. The energy (4) generated by all learned points is shown as a shaded surface. It essentially extends the KL-divergence to a dissimilarity with respect to an entire set of joint intensity distributions. During the optimization process it constrains the displacement field such that the corresponding intensity distribution remains within the valleys of low energy. This ensures that the joint intensity distribution will favor similarity to previously learned intensity distributions during the optimization.

What does it mean that the current joint intensity distribution is forced to be similar to one or the other previously learned intensity distribution? To this end, let us consider the following very simple example. Assume we have learned two joint distributions, where the first one states that white pixels in image 1 are always associated with black

[2]Note that using a Gaussian kernel $\exp(-\mathcal{I}_{KL}^2/(2\sigma^2))$ rather than an exponential one in (4) will lead to an additional factor of I_{KL}/σ in (14), which might provide better convergence properties as the gradient goes to zero for $p_u \to p_j$. We plan to investigate this in future work.

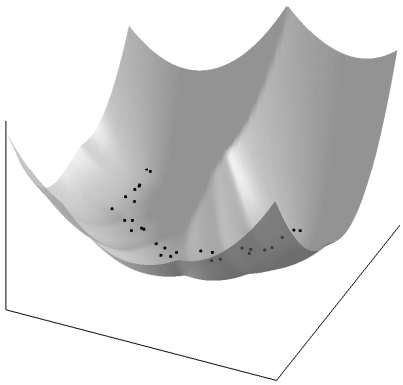


Figure 2. Schematic plot of energy (10). Each black point represents a joint intensity distribution. The energy (10) measures the dissimilarity between a given joint intensity distribution and the previously learned distributions.

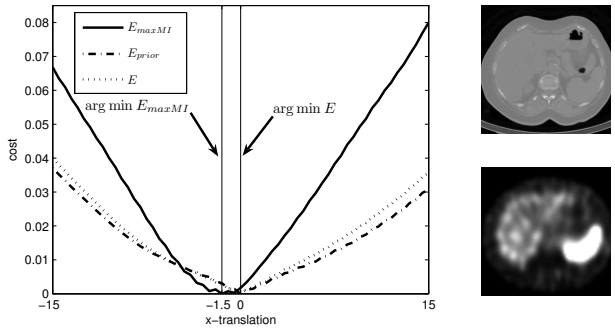


Figure 3. SPECT-CT registration performance analysis. SPECT slice was shifted horizontally within a range of $-15mm$ to $15mm$. Mutual information is noisy around the optimum and its minimum actually corresponds to an incorrect alignment. The integration of a prior on the joint intensity distribution provides for a larger basin of attraction and enables the estimation of the correct alignment.

pixels in image 2 and vice versa, while the second one states that the matching of white-to-white and black-to-black is most likely. Then enforcing similarity to one or the other by energy (10) has the following effect: If during optimization pairs of white pixels are associated through the displacement field, then this induces proximity to the second learned intensity distribution, and the prior will automatically enforce that black should also be associated with black – because a matching of white-to-white on one hand but black-to-white on the other is not consistent with any of the two learned intensity distributions. In other words: The matching of certain intensities will provide clues for the matching of others, as indicated by the learned joint distributions.

The above example illuminates the idea of imposing a prior on the space of joint intensity distributions. Note that this is fundamentally different from learning a *single* joint intensity distribution, as proposed for example by Leventon and Grimson [10]. Firstly, our method allows for a large variety of different intensity distributions. Secondly, the inherent selection mechanism allows the algorithm to infer statistical relations between matching of different intensity pairs, as in the simple example of two joint distributions discussed above.

4. Experimental Results

In the following, we will evaluate the proposed statistical framework for image registration. In Section 4.1, a quantitative study on a SPECT - CT image pair shows that priors on the joint intensity distribution can improve the mutual-information-based registration process by increasing the basin of attraction and by shifting the location of the energy minimum to the correct one. In Section 4.2, a study on the registration of a PET - CT image pair shows that the proposed multimodal prior on the joint intensity distribution outperforms a simpler unimodal prior, because the multimodal one allows the registration process to "select" among appropriate joint distributions. Section 4.3 shows that the proposed prior allows to cope with partial occlusions in a face registration task.

All implementations are done within a multi-resolution framework, giving computation times around 10 seconds for image pairs of size 450×450.

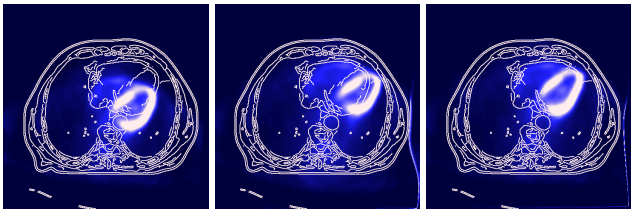
4.1. Quantitative Evaluation

Assume we are given a perfectly aligned image data set, such as the SPECT - CT image pair acquired by a Siemens Symbia T2 hybrid scanner in Fig. 3(a). Now we use these data to study the performance of competing objective functions, e.g. $E_{maxMI} = \max(\mathcal{I}_{MI}) - \mathcal{I}_{MI}$ for MI, E_{prior} in (10), and the total energy E in (9). In this experiment, the SPECT slice was shifted horizontally, while the CT image remained fixed, and the respective values of all three objective functions are computed, see Fig. 3(a). The energy plots show quantitatively that incorporating a prior (computed from the correctly aligned image pair) will lead to a superior registration algorithm. While this is only shown for the case of translation, one can expect similar improvements for non-rigid deformations.

4.2. PET - CT Medical Image Registration

Given several training image pairs, the proposed prior can incorporate a variety of joint intensity distributions. The following experiment will show that among this complementary information, the proposed algorithm selectively choses the intensity information appropriate for a specific registration task.

The training data is composed of two sets of aligned



(a) (b) (c)

Figure 4. PET/CT registration. (a) deformed PET/CT, (b) registration result using average prior, (c) registration result using both priors. While using the averaged joint intensity distribution as a prior leads to misregistration, the proposed multi-modal prior on the joint intensity distribution allows for the correct registration.

PET - CT slices[3] acquired by a Siemens hybrid scanner, Figs 1(a), 1(b). The superimposed edge maps (white) illustrate structural information of the CT image and visualize the quality of alignment. Note the significant difference in Figs 1(c) and 1(d) between the two shown joint intensity distributions, which reflect a typical scenario as it occurs in clinical applications.

An artificial deformation is applied to the PET slices and compared to the recovered displacement fields of using only a single prior distribution vs. using two complimentary. The two priors represent the joint distribution of the axial and coronal PET - CT slices shown in Fig. 1. In the case of a single prior, the average of the two is used for fair comparison.

Figure 4 illustrates the advantage of using several prior distributions as opposed to only using one. The weighting factors of energy (9) are set with a preference towards the prior energy E_{prior}, i.e. α_1 is chosen to be small. The width σ is determined using equation (6), and α_2 is chosen to allow for a smooth displacement field.

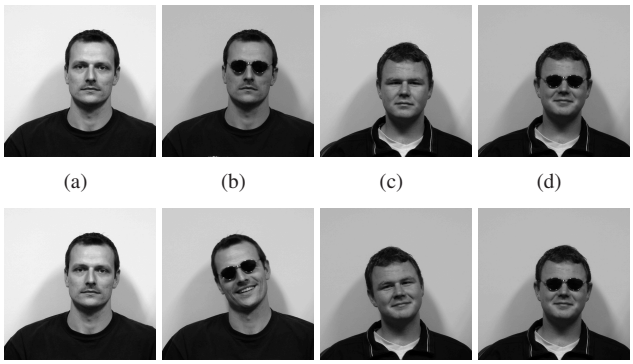
The results of recovering the significant deformation between the PET and CT images (see Figure 4(a)) are shown in Figs. 4(b) and 4(c). Using an average distribution misleads the algorithm and registration fails, see Fig. 4(b). However, the proposed method can fully utilize the given priors and correctly "selects" the closest joint intensity distribution. As a result, the underlying deformation is fully recovered (Fig. 4(c)).

This experiment shows the strength of introducing a space of joint intensity distributions, while the algorithm is able to chose the best available prior information. In case no best information is available, the prior energy decays to zero, and performance will be at least as good as using a context-free similarity measure.

[3]PET and SPECT are nuclear imaging techniques which visualize centers of high glucose acticity in the human body.

4.3. Face Registration in the Presence of Occlusion

Non-rigid multi-modal registration can serve as a preprocessing step for face recognition, where facial and/or head motion must be recovered in order to establish correspondence. In the following experiment, we illustrate how prior knowledge on the joint intensity distribution improves the registration in the presence of lighting variation and occlusion. The experiment is to recover facial expressions and head movement between two images. The second image is taken under different lighting conditions with the person wearing sun glasses. The objective functions of comparison are (i) purely MI based registration and (ii) the proposed combined approach using prior knowledge in eq. (9). The first row of Fig. 5 shows the two pairs of manually registered training data used to construct the prior, i.e. $m = 2$. To compare the performance, the same images have been registered (i) by the context-free MI criterion and (ii) by additionally imposing a prior on the space of joint intensity distributions. The parameters used here are similar to previous experiment. The second row of Fig. 5 shows the reference and alignment images that are subject to registration. Those images show multi-modality due to a slight illumination change, but moreover due to the appearance of the sun glasses. Furthermore, Figure 5 illustrates the differences of learned data towards the current data.



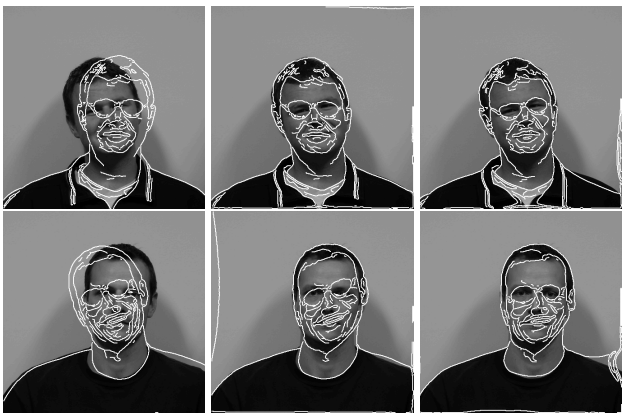
(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 5. Face images used for training and registration. (a)-(d) training images, (e)-(h) reference and alignment images that are subject to registration. The latter pose a challenging registration task and slightly differ from the training data.

There are two runs for each objective function that are being compared. Figure 6 shows the achieved results for pure MI and for imposing a space of prior information. Since the underlying transformation is unknown, the edge map of the alignment image is superimposed on the reference image for performance comparison. Column 6(a) shows the initial positions of the faces, column 6(b) shows the results using pure MI, and column 6(c) plots the results of the energy in equation (9). Comparing the edge maps it can be noticed that the proposed energy (9) is superior to

using pure MI. The MI method matches the outline of the persons correctly but fails to match the glasses in the aligment image on the eye region of the reference image. The combined method, however, succeeds for both faces in establishing correspondence, see column 6(c). Note that our method selects the prior intensity distribution, which corresponds best to the current input images.



(a) (b) (c)

Figure 6. Face image registration results. Column (a) shows initial alignment of the two images, column (b) the final registration for pure MI-based energy, and column (c) illustrates the final registration using energy (9). The energy (9) shows to be superior to context-free MI energy by minimizing the distance towards previously learned intensity distributions.

5. Conclusion

In this paper, we proposed a multimodal prior on the joint intensity distribution in order to enhance image registration problems. While MI was shown to provide a powerful registration criterion, it remains a purely low-level criterion. Our formulation allows to enhance this existing registration method in order to integrate prior knowledge about likely intensity correspondences, which is statistically learned from multiple pairs of pre-registered training images. Experimental results on both medical and face images demonstrate that our approach outperforms purely MI based image registration. Future directions of research include the extension of our approach to 3D, and a more complete qualitative and quantitative experimental study, in order to confirm robustness and accuracy of this approach for multi-modal datasets, in particular for larger training sets.

Acknowledgments

We thank Andrew Blake for helpful discussions and comments on the manuscript and support from Frank Sauer, Gareth Funka-Lea, and Joachim Hornegger. This work is funded by Siemens Corporate Research.

References

[1] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV (4)*, pages 25–36. Springer, 2004.

[2] A. C. S. Chung, W. M. Wells, A. Norbash, and W. E. L. Grimson. Multi-modal image registration by minimising Kullback-Leibler distance. In *MICCAI (2)*, pages 525–532, 2002.

[3] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3D multi-modality medical image registration using feature space clustering. In *CVRMed*, pages 195–204, 1995.

[4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE PAMI*, 24(5):603–619, 2002.

[5] D. Cremers, S. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *IJCV*, 2006. To appear.

[6] C. Guetter, C. Xu, F. Sauer, and J. Hornegger. Learning based non-rigid multi-modal image registration using Kullback-Leibler divergence. In *MICCAI (2)*, pages 255–262, 2005.

[7] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras. Variational methods for multimodal image matching. *IJCV*, 50(3):329–343, 2002.

[8] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[9] X. Huang, D. Metaxas, and T. Chen. Metamorphs: Deformable shape and texture models. In *CVPR*, pages 496–503, 2004.

[10] M. E. Leventon and W. E. L. Grimson. Multi-modal volume registration using joint intensity distributions. In *MICCAI*, pages 1057–1066, 1998.

[11] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE TMI*, 16(2):187–198, 1997.

[12] C.-E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2005.

[13] S. Roth and M. Black. On the spatial statistics of optical flow. In *ICCV*, pages 42–49, 2005.

[14] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1992.

[15] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Patt. Rec.*, 32(1):71–86, 1999.

[16] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE–based computation of image motion. *IJCV*, 45(3):245–264, 2001.

[17] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[18] L. Zöllei, J. W. Fisher, and W. M. Wells. A unified statistical and information theoretic framework for multi-modal image registration. In *IPMI*, pages 366–377, 2003.