Robust Registration of Multi-modal Medical Images Using Huber's Criterion

Nora Leïla Ouzir, Esa Ollila, Sergiy Vorobyov

Robust Registration of Multi-modal Medical Images Using Huber's Criterion

Nora Ouzir
Dept. of Mathematics
CentraleSupelec
France
nora.ouzir@centralesupelec.fr

Esa Ollila and Sergiy A. Vorobyov
Dept. of Signal Processing and Acoustics
Aalto University
Finland
esa.ollila@aalto.fi, sergiy.vorobyov@aalto.fi

Abstract—**Registration of multi-modal medical images is an essential pre-processing step, for example, for fusion or image guided-interventions. However, the alignment process is prone to high variability in tissue appearance between modalities, in addition to local intensity variations and artefacts. This work introduces a robust multi-modal registration approach that mitigates the undesirable effect of such variability. Robustness is achieved using Huber's loss function for the data fidelity and regularization terms. We propose a novel approach using Huber's criterion, which enables a jointly convex estimation of the motions and the associated scale parameters. We formulate the problem as a complex 2D transformation estimation and investigate a robust total-variation smoothing, as well as a dictionary learning-based data fidelity term. Experiments are conducted using two datasets of multi-contrast MR brain images.**

Index Terms—**Multi-modal registration, robust registration, Huber's criterion, coupled dictionary learning, MR-T1, MR-T2**

I. INTRODUCTION

The registration of multi-modal images, such as multi-contrast MR images or MR to computed tomography images, is an essential problem in medical image analysis. Because multiple scans (acquired from different imaging systems) can result in changes in the location or shape of anatomical structures, many processing tasks such as fusion, guided interventions or disease progression, require that the images are first registered, *i.e.*, aligned to the same coordinates.

In order to align a pair of images, registration methods usually seek to characterize either a feature-based or dense (*i.e.*, pixel-wise) spatial transformation that warps one image to the other. Dense or intensity-based registration methods face a common challenge because of the high variability in the tissue appearance between different modalities. Most of the current methods rely on mutual-information (MI) as a matching criterion [1]–[3]. MI is a global criterion that does not incorporate local information; this makes it also sensitive to local intensity variations. Moreover, it is not designed to handle imaging artefacts or noise. A different approach

first transforms the intensities into a domain where the images share the same appearance [4], [5]. Registration is then performed using classic similarity measures, such as the sum-of-squared differences, also known to be sensitive to intensity changes.

Robust loss functions provide an efficient way of incorporating resilience to intensity variations. Huber's loss function is a typical choice in this context [6]. Like most robust loss functions, Huber's loss requires the estimation of a scale parameter controlling outlier rejection properties. However, the simultaneous estimation of the parameters of interest and scale of the error terms is not jointly convex from the standard maximum likelihood (ML) perspective. An interesting jointly convex alternative is provided by Huber's criterion [6]. In [7], the Huber's criterion formulation has been generalized to the complex case in the context of multichannel sparse recovery.

In this work, Huber's loss function is used to incorporate robustness in the multi-modal image registration problem. The proposed robust method aims to reduce the impact of intensity variations and other registration outliers. We adopt a two stage methodology where one of the images is transformed to the appearance of the other modality before performing registration. In the first stage, a novel transformation method using coupled dictionary learning (CDL) [8] is proposed. Instead of using a global MI criterion, we follow a dense optical-flow approach [9]. Finally, a Huber's criterion-based formulation leads to a cost function that is jointly convex in the deformation to be estimated and the scale parameters associated with Huber's loss function.

II. ROBUST MULTI-MODAL REGISTRATION

A. Problem Formulation

Registration of two multi-modal input images $I \in \mathbb{R}^{N \times M}$ and $J \in \mathbb{R}^{N \times M}$ is achieved by estimating a dense non-rigid deformation $Z \in \mathbb{C}^{N \times M}$. At each pixel position (i, j) in the image, we consider that the deformation can be separated into a real horizontal and

imaginary vertical component, *i.e.,* $z_{ij} = u_{ij} + \mathrm{i}v_{ij}$, $i = 1, .., N$, $j = 1, ..., M$, and $\mathrm{i} = sqrt-1$.

Let $\boldsymbol{I}$ be the reference image and $\boldsymbol{J}^r$ be the registered version of $\boldsymbol{J}$. Then

$$\boldsymbol{J}^r(\boldsymbol{w}) = \boldsymbol{J}(\boldsymbol{w} + \boldsymbol{Z}) \qquad (1)$$

where $\boldsymbol{W} = \boldsymbol{X} + \mathrm{i}\boldsymbol{Y}$, with $\boldsymbol{X}$ and $\boldsymbol{Y}$ denoting the horizontal and vertical coordinates, respectively.

In the energy minimization framework, $\boldsymbol{Z}$ is typically obtained by minimizing a cost function composed of the sum a data fidelity term expressing the similarity between the input images, and a regularization term

$$\min_{\boldsymbol{Z}} \ \{E_{\text{data}}(\boldsymbol{Z}, \boldsymbol{I}, \boldsymbol{J}) + \lambda E_{\text{reg}}(\boldsymbol{Z})\} \qquad (2)$$

where $\lambda \in \mathbb{R}^+$ controls the amount of regularization. In this work, we use an optical-flow-based data fidelity term, combined with a spatial smoothing term based on total-variation [9]. The next subsection explains the first step of the proposed method, which focuses on the image transformation required for using optical-flow in the context of multi-modal imaging.

B. Optical-flow Using CDL

Optical-flow is a well-established approach for estimating dense deformations [9], [10]. As opposed to other classic methods that make use of the image differences directly, optical-flow has the key advantage of providing a linear data fidelity term. The linear relationship between the images and the deformation is based on the *brightness constancy assumption*, where it is assumed that the intensity of a pixel remains constant over short periods of time. A first order Taylor expansion leads to the classic optical-flow constraint equations

$$\partial_t \boldsymbol{I}_{ij} + \nabla_x \boldsymbol{I}_{ij} u_{ij} + \nabla_y \boldsymbol{I}_{ij} v_{ij} = 0 \qquad (3)$$

where $\partial_t \boldsymbol{I}$ denotes the temporal derivative of the intensities and ∇_x, ∇_y are the horizontal and vertical spatial gradient operators. In the context of multi-modal imaging, the brightness constancy assumption is does not hold because of the drastically different appearances between modalities. One solution to this problem is to transform the intensities of one input image to match the appearance of the second modality. Since it is not straightforward to find such a transformation, we propose to adopt a learning-based approach. Specifically, the implicit mapping between the modalities is learnt using CDL. This strategy allows us to obtain a data-driven mapping that does not require designing a specific transformation for each modality (or searching for relevant features to be registered beforehand). After mapping the images to the same modality, it is possible to use the classic optical-flow approach.

CDL [8] aims to capture the relationship between two input signals by finding their joint sparse representation in two *coupled* dictionaries. In this work, we learn two dictionaries $\boldsymbol{D}_I$ and $\boldsymbol{D}_J$ capturing the implicit appearance transformation between two medical imaging modalities. Specifically, these dictionaries are learnt by solving the following optimization problem [11]

$$\min_{\boldsymbol{D}_I, \boldsymbol{D}_J, \boldsymbol{A}} \|\boldsymbol{D}_I \boldsymbol{A} - \mathcal{P}(\boldsymbol{I}_{\text{train}})\|_{\text{F}}^2 + \|\boldsymbol{D}_J \boldsymbol{A} - \mathcal{P}(\boldsymbol{J}^r_{\text{train}})\|_{\text{F}}^2$$

$$\text{s.t.} \ \|\boldsymbol{A}_i\|_0 \le K, \|[\boldsymbol{D}_I]_j\|_2 = 1, \|[\boldsymbol{D}_J]_j\|_2 = 1, \forall j, i$$
$$(4)$$

where $\boldsymbol{I}_{\text{train}}$ and $\boldsymbol{J}^r_{\text{train}}$ are two registered training images, $\boldsymbol{A}$ is their joint sparse representation (with $\boldsymbol{A}_i$ denoting the ith column of $\boldsymbol{A}$), $\mathcal{P}$ is a binary patch extraction operator, and K stands for the sparsity level. Once the dictionaries are learnt, a new input image $\boldsymbol{J}$ is mapped to the modality of $\boldsymbol{I}$ by first finding its sparse representation $\boldsymbol{A}$ in the dictionary $\boldsymbol{D}_J$ (*e.g.,* using Orthogonal Matching Pursuit (OMP) [12]). The transformed image $\tilde{\boldsymbol{J}}$ is then reconstructed using $\boldsymbol{D}_I$ as

$$\tilde{\boldsymbol{J}} = \mathcal{P}^*(\boldsymbol{D}_I \boldsymbol{A}), \qquad (5)$$

where $\mathcal{P}^*$ is the operator that places each patch in its original position and averages overlapping patches. After transforming the image $\boldsymbol{J}$ and obtaining $\tilde{\boldsymbol{J}}$, the optical-flow equations (3) can be rewritten as

$$\boldsymbol{I}_{ij} - \tilde{\boldsymbol{J}}_{ij} + \nabla_x \boldsymbol{I}_{ij} u_{ij} + \nabla_y \boldsymbol{I}_{ij} v_{ij} = 0, \ \forall i, j \qquad (6)$$

where the temporal derivative $\partial_t \boldsymbol{I}$ has now been approximated by $\boldsymbol{I} - \tilde{\boldsymbol{J}}$. By denoting the complex gradient with respect to the coordinates $\boldsymbol{W}$ as ∇, *i.e.,* $\nabla \boldsymbol{I}_{ij} = \frac{1}{2}(\nabla_x \boldsymbol{I}_{ij} - \mathrm{i}\nabla_y \boldsymbol{I}_{ij})$, (6) leads to the data error term

$$e_d(z_{ij}) = \boldsymbol{I}_{ij} - \tilde{\boldsymbol{J}}_{ij} + 2\mathrm{Re}[\nabla \boldsymbol{I}_{ij} z_{ij}] \qquad (7)$$

C. Robust Cost Function Using Huber's Criterion

The data fidelity and regularization terms in (2) also require the choice of an appropriate loss function ρ. As explained in Section I, we employ Huber's loss function in order to deal with the various outliers that can be present in the context of multi-modal registration. Huber's loss function takes the following form [6]

$$\rho(e) = \begin{cases} |e|^2 & |e| \le \sigma c \\ 2c|e| - c^2 & |e| > \sigma c, \end{cases} \qquad (8)$$

where c is a user defined parameter controlling the outlier rejection and efficiency properties, and σ is the associated scale parameter. For example, a value $c = 1.345$ provides 95% asymptotic relative efficiency under Gaussian errors [13] (this is the value that will be used in our experiments in Section III). Note that Huber's loss function can be viewed as a hybrid ℓ_2-ℓ_1 loss function where the relatively small errors are treated with the ℓ_2-loss while large errors are penalized less heavily using the ℓ_1-loss.

1) Huber's Criterion: Solving (2) using Huber's robust loss function requires an estimate of the scale parameter for both E_{data} and E_{reg}. In the standard ML approach, one minimizes the negative log-likelihood of the data when the errors e are assumed to follow a probability density function of the form $f(x) \sim \exp(-\rho(x))$. For example, the ML approach using Huber's loss over the data errors (7) yields the following data fidelity term

$$E_{\text{data,ML}} = \sum_{i,j} \rho(e_d[z_{ij}]/\sigma_d) + NM\log(\sigma_d), \quad (9)$$

where σ_d is the scale parameter associated with e_d. The ML formulation in (9) has the disadvantage of not being jointly convex with respect to the transformation $\boldsymbol{Z}$ and the scale σ_d. This issue can be circumvented using Huber's criterion [13], which provides a jointly convex alternative to the ML approach. Joint convexity is achieved by reformulating $E_{\text{data,ML}}$ as

$$E_{\text{data,HUB}} = \sum_{i,j} \rho(e_d[z_{ij}]/\sigma_d)\sigma_d + 2NM\beta\sigma_d, \quad (10)$$

with $\beta > 0$ being a fixed scaling factor used to obtain Fisher-consistency for Gaussian errors and computed using the predefined parameter c (see [13] for more details on the computation of this parameter).

2) Final Cost Function: After transforming one of the images using the CDL approach presented in Subsection II-B, a cost function of the form (2) is constructed by combining the optical-flow-based term (7) with a regularization enforcing a smooth deformation. As explained above, we use Huber's criterion in order to bypass the non-convexity of the ML approach, and obtain the final jointly convex minimization problem

$$\min_{\boldsymbol{Z},\sigma_d,\sigma_r} \sum_{i,j} \rho\left[\frac{\boldsymbol{I}_{ij} - \tilde{\boldsymbol{J}}_{ij} + 2\text{Re}[\nabla\boldsymbol{I}_{ij}z_{ij}]}{\sigma_d}\right]\sigma_d \quad (11)$$

$$+\lambda \sum_{i,j} \rho\left[\frac{\nabla^* z_{ij}}{\sigma_r}\right]\sigma_r + 2NM(\beta\sigma_d) + 2\lambda NM(\beta\sigma_r)$$

where the first term corresponds to the optical-flow constraint using the transformed image $\tilde{\boldsymbol{J}}$ and the reference image $\boldsymbol{I}$, while the second term stands for the smoothness constraint weighted by the parameter λ. Specifically, the use of Huber's loss for the regularization term ensures that the gradient of the deformation field changes slowly, while allowing for some discontinuities. Note that the associated scale parameter is denoted σ_r and ∇^* stands for the complex conjugate gradient , *i.e.,* $\nabla^* z_{ij} = \frac{1}{2}(\nabla_x u_{ij} - \nabla_y v_{ij}) + \frac{i}{2}(\nabla_y u_{ij} + \nabla_x v_{ij})$.

D. Minimization

The minimization of (11) is carried out using a gradient descent approach for the deformations $\boldsymbol{Z}$, combined with a fixed point update of the scale parameters σ_d and σ_r. The fixed point iteration is obtained by equating the gradient of (11) to zero. The final registered image is obtained using (1).

III. EXPERIMENTS

The proposed method is compared to the multi-modality non-rigid demons algorithm [14], which is a multi-modal version of the classic *demons algorithm* [15]. This method incorporates the mutual-information criterion by transforming the images from one modality to the other using 2D mutual histograms. The estimated transformation is dense, and smoothed at each iteration of the algorithm using Gaussian filtering. A multi-resolution scheme is used to handle large deformations. We will refer to this method as MM-demon. The registration performance of each method is measured using the root-mean-squared error (RMSE) between the true and estimated deformations.

A. Datasets

The proposed robust registration method is evaluated using two publicly available datasets of mutli-contrast MR images, namely ALBERT [16] and *The Whole Brain Atlas*[1]. The former dataset contains neonatal MR brain images, while the latter depicts various brain diseases, such as strokes, Alzheimer's, etc. All multi-modal pairs of images are originally registered in both datasets, which allows us to have a ground-truth for performance evaluation. A subset of these registered images will also be used in the learning phase (the CDL step described in Subsection II-B). For each pair, we then apply different dense deformations to one of the images (this will be the image to be registered). An example of a pair of input images, as well as the image obtained after deformation, is shown in Fig. 1.
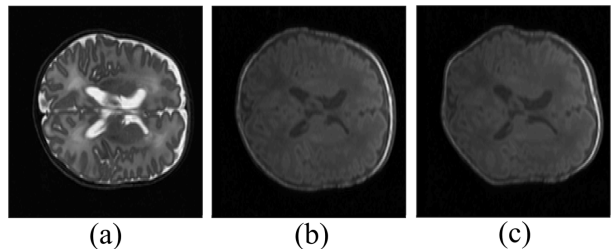

Fig. 1: Examples of an (a) input MR-T2 image ($\boldsymbol{I}$), (b) input MR-T1 image and (c) the T1 image ($\boldsymbol{J}$) after deformation.

Fig. 2: The transformed image ($\tilde{J}$) obtained for the T2 image shown in Fig. 1-a.

B. Transformation Learning Using CDL

The first step of the proposed method focuses on learning the implicit transformation between multi-contrast MR images. A total of 50000 patches of size 8×8 pixels were extracted from *The Whole Brain Atlas* dataset and used as a training set.[2] The sparsity parameter was set to $K = 15$ with a dictionary size of 64×900. Fig. 2 shows the image $\tilde{J}$ obtained by transforming the T1 image J in Fig. 1-a to the appearance of the T2 modality. Notice that the outer edge of the original T1 image does not appear in this transformed version of the T2 image.

C. Registration Performance

Two types of deformations were applied to the test images before registration. First, the performance was tested using simple translations of 1 pixel in both the horizontal and vertical directions. The second type of deformations were generated using a Free-Form-Deformation model based on B-splines, with different magnitudes. These non-rigid transformations were larger than those of the first category.

Tables I and II show the RMSEs obtained for all the images in both datasets. The errors obtained using the proposed method with $c \to \infty$ (*i.e.,* non-robust setting denoted by NR) are provided for comparison. The results show that the proposed method outperformed the MM-demon algorithm for all experiments. However, we can observe some differences in the gap between the errors for the different types of deformations. Specifically, we can see a more significant improvement in the case of small translations (Table I). This can be explained by the fact that the proposed method does not use a multi-resolution scheme in order to deal with large deformations.

Overall, the proposed robust method performed better than its non-robust version, showing the benefits of robustness. The pixels that are treated differently with these two approaches correspond to data or spatial outliers. These are the pixels that result in errors greater than

---

[2]These training images were excluded from the registration evaluation set.

Dataset	Brain Atlas	Neonatal
MM-Demon	2.75	1.91
Proposed NR	0.88	0.53
Proposed R	0.74	0.49

TABLE I: RMSE for rigid deformations

Dataset	Brain Atlas	Neonatal
MM-Demon	3.02	2.21
Proposed NR	2.28	1.80
Proposed R	2.08	1.65

TABLE II: RMSE for non-rigid deformations
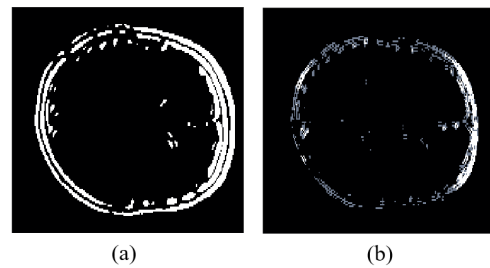

(a)   (b)

Fig. 3: Resulting (a) data and (b) spatial outliers for the images in Fig. 1. The pixels corresponding to outliers are shown in white color.

the outlier rejection threshold defined by c and σ_d or σ_r. Fig. 3 shows an example of data and spatial outliers obtained for the images in Fig. 1. One can see that in this example, outliers corresponded to the outer edge of the brain. This is the fat layer that can be observed in the T1 images but is relatively dark in the T2 modality (see Fig. 1). This shows how the proposed robust formulation can account for high variabilities in tissue appearance.

IV. CONCLUSION

This paper has presented a robust method for the registration of multi-modal images. A first preprocessing step using coupled dictionary learning has been used to transform the appearance of the images from one modality to the other. In a second step, registration has been performed by minimizing a robust cost function composed of an optical-flow and spatial regularization terms using Huber's loss function. One more contribution of this work focuses on taking advantage of Huber's criterion formulation, providing a jointly convex cost function with respect to the image deformation field and its scale parameters. Finally, experimental results obtained using two datasets of multi-contrast MR images have shown the improved robustness of the proposed approach.

REFERENCES

[1] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proceedings of the IEEE*, vol. 91, no. 10, pp. 1699–1722, 2003.

[2] J. Woo, M. Stone, and J. L. Prince, "Multimodal registration via mutual information incorporating geometric and spatial context," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 757–769, 2015.

[3] M. P. Wachowiak, R. Smolikova, Yufeng Zheng, J. M. Zurada, and A. S. Elmaghraby, "An approach to multimodal biomedical image registration utilizing particle swarm optimization," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 289–301, 2004.

[4] A. Wong and J. Orchard, "Robust multimodal registration using local phase-coherence representations," *J. Signal Process. Syst.*, vol. 54, no. 13, p. 89100, Jan. 2009.

[5] A. Guimond, A. Roche, N. Ayache, and J. Meunier, "Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 58–69, 2001.

[6] P. Huber, *Robust Statistics*. Wiley, New York, 1981.

[7] E. Ollila, "Multichannel sparse recovery of complex-valued signals using huber's criterion," in *2015 3rd Int. Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, 2015, pp. 26–30.

[8] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, 2012.

[9] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 13, p. 185203, Aug. 1981.

[10] N. Ouzir, A. Basarab, O. Lairez, and J. Y. Tourneret, "Robust optical flow estimation in cardiac ultrasound images using a sparse representation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 741–752, 2019.

[11] F. Veshki and S. Vorobyov, "An efficient coupled dictionary learning method," *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1441–1445, Oct. 2019.

[12] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44 vol.1.

[13] A. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[14] D. Kroon and C. H. Slump, "MRI modalitiy transformation in demon registration," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 963–966.

[15] J.-P. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243 – 260, 1998.

[16] I. S. Gousias, A. D. Edwards, M. A. Rutherford, S. J. Counsell, J. V. Hajnal, D. Rueckert, and A. Hammers, "Magnetic resonance imaging of the newborn brain: manual segmentation of labelled atlases in term-born and preterm infants," *NeuroImage*, vol. 62, no. 3, p. 14991509, September 2012. [Online]. Available: www.brain-development.org