An extension of Digital Image Correlation for intermodality image registration

Erika Tudisco[a], Clément Jailin[b], Arturo Mendoza Quispe[b], Alessandro Tengattini[c], Edward Andò[c], Stephen Hall[d], Gioacchino Viggiani[c], François Hild[b], Stéphane Roux[b]

[a] *Division of Geotechnics, Lund University, Sweden*
[b] *LMT-Cachan (ENS de Cachan/CNRS/Univ. Paris-Saclay), Cachan, France*
[c] *Univ. Grenoble Alpes, CNRS, Grenoble INP[1], 3SR, F-38000 Grenoble, France*
[d] *Division of Solid Mechanics, Lund University, Sweden*

Abstract

The question of registering two images (or volumes) acquired with different modalities, and thus exhibiting different contrast, is addressed based on an extension of global Digital Image (or Volume) Correlation. A specific comparison metric is to be introduced allowing the signature of the different phases to be related. A first solution consists of Gaussian mixture to describe the joint distribution of gray levels, which not only provides a matching of both images, but also offers a natural segmentation indicator. A second 'self-adapting' solution does not postulate any *a priori* model for the joint histogram, and registers both images based on their initial histogram. The algorithm is implemented with a pyramidal multiscale framework for robustness. The proposed multiscale technique is tested on two 3D images obtained from X-ray and neutron tomography respectively.

Keywords: Digital image correlation; Image fusion; Image registration; Neutron tomography; X-ray tomography

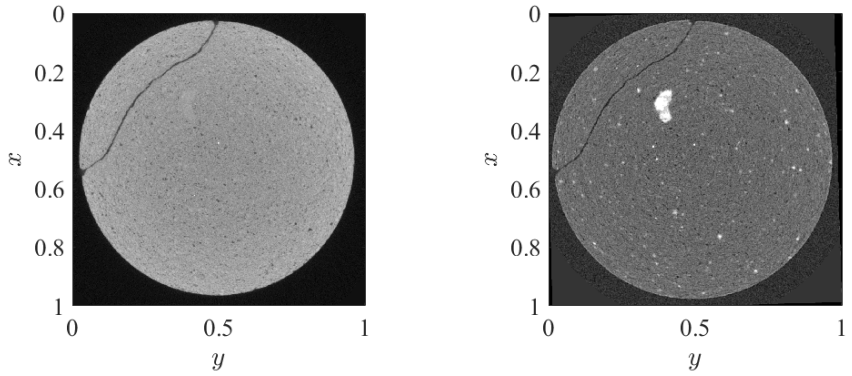
1. Introduction

Various modalities provide different information on materials because of their different sensitivities, and thus their different contrast. Yet to benefit

[1] Institute of Engineering Univ. Grenoble Alpes

from multiple modalities it is essential to be able to obtain a local characterization of them for the *same* material point. However, images are generally not acquired (or reconstructed) in the same frame. Hence rigid body motions have to be considered to make the two frames coincident. Additionally, image resolutions may be different, so that the transformation has to be enriched with a possible dilation/contraction. The purpose of the present study is to identify the transformation $\boldsymbol{F}$ (*i.e.,* rigid body motions and isotropic dilation, that is 4 or 7 parameters respectively in 2D or 3D) calling for the registration of the two images.

As an illustration, Figure 1 shows two horizontal slices of the same cylindrical specimen imaged by X-ray and neutron tomography. The contrast is different in both images, although some specific (and matching) patterns can be seen. This 3D example will be used to validate the proposed registration procedure.



(a) X-ray (b) Neutron

Figure 1: An example pair of images obtained from slices manually selected from reconstructed a) x-ray and b) neutron tomography volumes

There exist simple ways to find such a transformation: a set of remarkable points can be chosen in both images and from their respective coordinates, the transformation $\boldsymbol{F}$ can be determined [1]. In three dimensions, a minimum of three points is needed, whereas two points are sufficient in two dimensions. Redundancy provided by using more points allows robustness to be improved and more importantly the validation to be thoroughly probed. This technique is expected to allow for a matching with an accuracy of typically a few pixels.

Moreover, the selection of remarkable points may be tedious especially in three dimensions.

Multimodal (multi-sensor) registration poses a particular difficulty due to the fact that the contrast of both images is different, and hence it is difficult to assess their coincidence from a proper metric. Thus sometimes artificial examples are considered where the exact solution is known and hence quadratic differences (or equivalent signal to noise ratio) between the test case and its known reference is used to evaluate a given strategy before it is applied to real data [2]. Yet the question of the representativeness of the artificial reference case and its noise mostly remain unknown. As previously discussed, multimodal registration can be achieved by selecting some features that are deemed robust, such as sharp gradients, and specifically their orientation as the gradient magnitude cannot generally be compared. Strategies consisting in matching such features and attributing "points" for each match so as to maximize the registering score have been proposed [3], although they call for a necessarily subjective appreciation, as for the above mentioned remarkable points.

Traditionally, methods based on mutual information (MI) [4] represent the leading technique in multimodal registration [5]. Put simply, the MI of two images is the amount of information that one image contains about the other and *vice versa* [6]. Such a metric provides a very general (probabilistic) framework that does not call for prior expectation or assumptions (such as linear relationships) but only for a predictable relationship. However, one of its biggest pitfalls is the arbitrariness of the measure attached to "gray levels", and the lack of "distances" in the space of joint gray levels. Similarly, the knowledge of gray levels in the spatial vicinity of a point is not used. Another drawback is that most implementations require the images to be registered to be defined as random variables (*i.e.,* determine their probability density functions); for which either discrete or continuous approaches need to be used (such as estimation through Parzen windows [7]).

The first approach for this type of MI-based methods was proposed by Viola and Wells [7]. Since then, different approaches have been explored in order to obtain non-rigid multimodal registration: by using multiple local windows connected through a Gaussian window function ensuring continuity and smoothness of the deformation field [8]; by using a "correlation ratio" [9] that is based on the MI but drastically improves on it by considering the spatial information; by combining MI with a term based on the image gradient to be registered [6] (*i.e.,* multimodal images can have drastically different

intensities but since images fundamentally depict the same microstructures, gradients in two multimodal images are assumed to be similar); or by modeling the deformed image as a viscous fluid that deforms under the influence of forces derived from the gradient of the MI [10] (built upon [11]).

However, image registration, and particularly Digital Image Correlation (DIC), has shown a much higher resolution potential [12]. Resolutions of one hundredth of pixel can be achieved due to the exploitation of all pixels (or voxels) in a given region of interest. Such resolution is precious for solid mechanics applications [13], and hence it would be appealing to benefit from a similar matching accuracy when dealing with images issued from different sensors. However, the difficulty is that a suitable metric has to be considered to estimate the "distance" between the two images when in coincidence.

Let us note that, although not very frequent, within the methodology of DIC, some prior work has explored the question of image registration with different modalities, but with a rather straightforward correspondence between the different contrasts. For instance, the calibration of distortions in an SEM has been studied aimed at registering a design of a speckle marking (a binary image) with an actual acquisition of the surface where the marking has been applied by a Pt e-lithography technique [14]. In this case an affine transformation of the gray levels revealed sufficient. In Refs. [15, 16], the authors used the joint histogram of two simultaneous X-ray and neutron acquisitions assuming that the images were identically located without displacement and distortion. A more difficult example deals with an extension of classical stereo-vision or stereocorrelation [17, 12] to hybrid stereocorrelation. In that case, two images, one acquired with a standard optical camera, and a second one with an IR camera, taken from two different points of view were matched so as to extract a 3D shape and displacement of the surface of a specimen [18].

The outline of the article is the following. After having introduced the problem and notations in Section 2, it is proposed to address the registration as a minimization based on a "potential" Φ. The algorithm, including a multiscale version, follows closely what is done in global DVC, and is presented in Section 3. Different formulations can be proposed for the potential as discussed in Section 4. However, within a probabilistic framework, the minimization of the potential can be seen as the maximization of a log-likelihood providing clues to formulate different descriptions. Two particular cases are proposed. The first is based on a Gaussian Mixture model to account for the joint histogram. It is shown to be very well suited to the problem at hand

and has the additional benefit of leading to a segmentation into phases. The second is model-free and hence does not rely on any specific prior modeling of the microstructure into phases. In contrast, it aims to learn the potential from the current joint histogram of gray levels. To illustrate the merits and limits of the proposed algorithm, Section 5 provides a 3D test case in which the two approaches are applied. The results show that both approaches are successful at registering the images, and provide comparable estimates of transformations. A brief summary of the proposed procedure and achieved results is given in Section 6 together with some perspectives.

2. Formulation

The two images are $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$, where f and g denote the gray level values provided by each modality. The registration consists in finding the transformation $\boldsymbol{F}$ such that the material point at position $\boldsymbol{x}$ in the first image coincides with the same material point at position $\boldsymbol{x}' = \boldsymbol{F}.\boldsymbol{x}$ in the second one. In order to account for translation, rotation and dilation, it is convenient to use homogeneous coordinates $\boldsymbol{x} = (x_1, x_2, x_3, 1)$, where the last unit coordinate allows the change in origin to be considered within the same linear framework. The sought transformation is linear and hence it can be represented by the 4×4 matrix $\boldsymbol{F}$ with a trivial conventional row $F(4,.) = (0,0,0,1)$. The non trivial part is a 3×4 matrix that may also describe an arbitrary homogeneous deformation. The above assumptions imply that the deviatoric part of the strain (5 components) is null, so that the remaining degrees of freedom are $12-5 = 7$ consistent with the 3 translations, 3 rotations and 1 dilation. $\boldsymbol{F}$ can be restricted to the above 7 unknowns or left free with 12 unknowns, in which case the estimated deviatoric strain may be a way to validate the approach or to detect possible distortions in one modality or the other.

In the following, the second possibility of 12 unknowns is chosen. More specifically, $\boldsymbol{F}$ is written as

$$\boldsymbol{F} = \begin{bmatrix} \begin{bmatrix} \boldsymbol{I} + \boldsymbol{\varepsilon} + \boldsymbol{\omega} \end{bmatrix} & \begin{Bmatrix} \boldsymbol{t} \end{Bmatrix} \\ \{0 \quad 0 \quad 0\} & 1 \end{bmatrix} \tag{1}$$

where $\boldsymbol{I}$ is the 3×3 identity matrix, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\omega}$ respectively the symmetric and anti-symmetric parts of the $\boldsymbol{F}(1-3, 1-3)$ upper block, $\boldsymbol{t}$ is a translation

vector. The origin of the coordinate system is located in one corner of the images (minimum x_i coordinates are 0). The anti-symmetric tensor $\boldsymbol{\omega}$ can be interpreted as a rotation, while $\boldsymbol{\varepsilon}$ is the infinitesimal strain tensor.

If both images were acquired with the same modality, it would have been expected that for all $\boldsymbol{x}$ in the Region of Interest

$$f(\boldsymbol{x}) = g(\boldsymbol{F} \cdot \boldsymbol{x}) \tag{2}$$

i.e., the brightness conservation assumption, which is at the basis of digital image correlation [13]. The above relationship can be relaxed, for instance to account for different brightness or contrast, and is rewritten as

$$\Phi(f(\boldsymbol{x}), g(\boldsymbol{F} \cdot \boldsymbol{x})) = 0 \tag{3}$$

where $\Phi(f, g) = \alpha_1 f - \alpha_2 g - \alpha_3$ is an affine transformation. In particular, image correlation based on normalized cross-correlation for instance is naturally immunized against such an affine transform Φ [12]. Image correlation is generally formalized to tolerate a gaussian white noise that affects both images. This leads to relaxing the above formulation so that the transformation $\boldsymbol{F}$ is sought as the minimizer of the functional $\mathcal{T}$ over all affine transforms

$$\begin{aligned} \boldsymbol{F} &= \text{Argmin}_{\boldsymbol{G}} \mathcal{T}[\boldsymbol{G}] \\ \mathcal{T}[\boldsymbol{G}] &= \sum_{\boldsymbol{x} \in ROI} \Phi^2(f(\boldsymbol{x}), g(\boldsymbol{G} \cdot \boldsymbol{x})) \end{aligned} \tag{4}$$

When considering different modalities, the change in gray levels is generally much more drastic than just an affine transformation. Yet, the assumption that is physically sound is that the material consists of few phases $i = 1, ..., N$ that have a well defined signature (f_i, g_i) in the (f, g) plane. In other words, wherever the phase is located in space, it will give the same gray level with the same modality. Note however that even if the above statement seems quite natural, for instance the artifact called "beam hardening" in X-ray tomography [19] violates this assumption. Figure 2 shows a log-scale joint histogram of the two images shown in Figure 1 prior to any registration. It is natural to interpret the most salient peaks as corresponding to the different phases (f_i, g_i), at the reservation that prior to registration, a wrong voxel positioning may give rise to spurious peaks in the joint histogram, as the f and g coordinates may refer to a different phase.

The spirit of the proposed approach is thus to design a "potential" Φ that is locally minimum for the different phases (f_i, g_i), and to determine $\boldsymbol{F}$

from the minimization of $\mathcal{T}$. Different approaches can be followed in order to choose this potential. The discussion on its determination is deferred for a further section, and the minimization algorithm is first discussed.

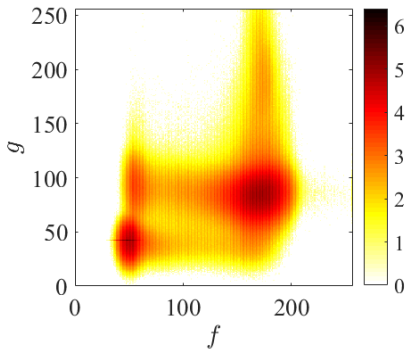


Figure 2: $\log_{10}$-scale joint histogram of gray levels from the sections shown in Fig. 1 with no registration

3. Registration algorithm

The registration problem is nonlinear and it is proposed to solve it with a Gauss-Newton algorithm, assuming that the initial determination of $\boldsymbol{F}$ is close to the solution. Thus the algorithm consists in the determination of successive corrections to the current estimate of the transformation $\boldsymbol{F}^{(n)}$ at iteration n.

3.1. Gauss-Newton algorithm

Let us introduce the following notations for the corrected image

$$\widetilde{g}^{(n)}(\boldsymbol{x}) \equiv g(\boldsymbol{F}^{(n)} \cdot \boldsymbol{x}) \tag{5}$$

At iteration n, the correction is denoted as $\boldsymbol{G}^{(n+1)} = \boldsymbol{I} + \boldsymbol{\delta F}^{(n+1)}$, such that $\boldsymbol{F}^{(n+1)} = \boldsymbol{G}^{(n+1)}\boldsymbol{F}^{(n)}$. The functional to be minimized takes the following expression

$$\mathcal{T}[\boldsymbol{F}^{(n+1)}] = \sum_{\boldsymbol{x}\in ROI} \Phi^2(f(\boldsymbol{x}), \widetilde{g}^{(n)}(\boldsymbol{G}^{(n+1)} \cdot \boldsymbol{x})) \tag{6}$$

Expanding the above integral up to second order in $\boldsymbol{\delta F}^{(n+1)}$ yields

$$\begin{aligned}\mathcal{T}[\boldsymbol{F}^{(n+1)}] &= \sum_{\boldsymbol{x}\in ROI} \Phi^2\left(f, \widetilde{g}^{(n)} + (\boldsymbol{\nabla}\widetilde{g}^{(n)} \cdot \boldsymbol{\delta F}^{(n+1)} \cdot \boldsymbol{x})\right) \\ &= \sum_{\boldsymbol{x}\in ROI} \begin{array}{l}\big[[\Phi^2] + ([\Phi^2]_{,2})(\boldsymbol{\nabla}\widetilde{g}^{(n)} \cdot \boldsymbol{\delta F}^{(n+1)} \cdot \boldsymbol{x}) \\ + (1/2)([\Phi^2]_{,22})(\boldsymbol{\nabla}\widetilde{g}^{(n)} \cdot \boldsymbol{\delta F}^{(n+1)} \cdot \boldsymbol{x})^2\big]\end{array}\end{aligned} \tag{7}$$

where the notations $[\Phi^2]_{,2}$ and $[\Phi^2]_{,22}$ are used to refer to the partial derivatives with respect to the second argument once or twice respectively and Φ^2 as well as its derivatives are computed at point $(f, \widetilde{g}^{(n)})$ (omitted for conciseness).

Because the unknowns are themselves the component of a matrix, $\delta F_{ij}^{(n+1)}$ it is convenient to relabel them with a single index m such that $\delta F_{m}^{(n+1)} \equiv \delta F_{i(m)j(m)}^{(n+1)}$, thus defining implicitly $i(m)$ and $j(m)$, in the spirit of Voigt notation for the strain tensor [20]. Since the above expression of the functional is quadratic in $\boldsymbol{\delta F}^{(n+1)}$, its determination is the solution to a linear system

$$M_{mp}^{(n)} \delta F_{p}^{(n+1)} = A_m \tag{8}$$

where

$$M_{mp}^{(n)} = \sum_{\boldsymbol{x}\in ROI} ([\Phi^2]_{,22})(x_{j(m)}\widetilde{g}_{,i(m)}^{(n)})(x_{j(p)}\widetilde{g}_{,i(p)}^{(n)}) \tag{9}$$

and

$$A_m^{(n)} = -\sum_{\boldsymbol{x}\in ROI} ([\Phi^2]_{,2})(x_{j(m)}\widetilde{g}_{,i(m)}^{(n)}) \tag{10}$$

In those equations x_i refers to the i-th component of the vector $\boldsymbol{x}$.

The correction to the transformation is then taken into account as

$$\boldsymbol{F}^{(n+1)} = (\boldsymbol{I} + \boldsymbol{\delta F}^{(n+1)})\boldsymbol{F}^{(n)} \tag{11}$$

It may be worth noting that standard DIC and DVC (*i.e.,* with the same modality) is recovered for $\Phi^2(f, g) = (1/2)(f - g)^2$, where Hessian and Jacobian respectively take the well-known [21] following expressions

$$M_{mp} = \sum_{\boldsymbol{x}\in ROI} (x_{j(m)}\widetilde{g}_{,i(m)}^{(n)})(x_{j(p)}\widetilde{g}_{,i(p)}^{(n)}) \tag{12}$$

and

$$A_m = \sum_{\boldsymbol{x}\in ROI} \left(f - \widetilde{g}^{(n)}\right)(x_{j(m)}\widetilde{g}_{,i(m)}^{(n)}) \tag{13}$$

Within the classical DIC and DVC frameworks, a very precious tool is the residual field, $\rho(\boldsymbol{x})$, namely the image difference between reference and corrected deformed images. This field, which is resolved at the voxel scale, shows success of the registration, and possibly where it did not, so that the registration may be revisited with other assumptions or corrections. When extended to the above general intermodality framework, it is observed that the equivalent residual field is

$$\rho(\boldsymbol{x}) = \frac{\partial \Phi^2(f(\boldsymbol{x}), g(\boldsymbol{F} \cdot \boldsymbol{x}))}{\partial g} \tag{14}$$

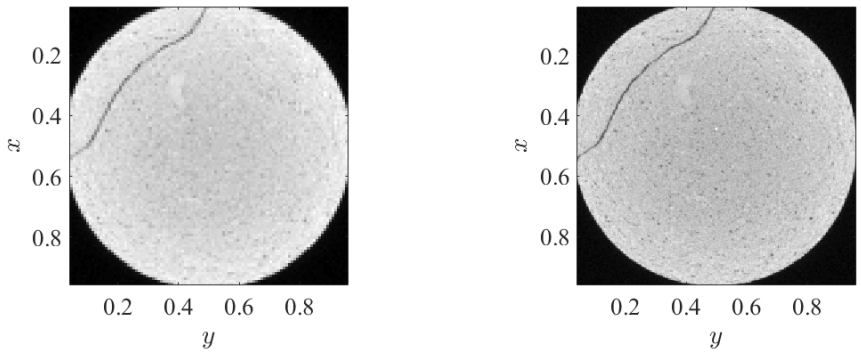
computed at convergence.

3.2. Multiscale approach

One of the major difficulties encountered in DIC and DVC is the initialization of the transformation $\boldsymbol{F}$. If it is too far from the actual solution, the above algorithm may simply diverge, or converge to a secondary minimum of $\mathcal{T}$. The criterion to track convergence is based on the ratio of the maximum displacement over all voxels of the region of interest and the correlation length of the image as estimated from the pair correlation function. When the ratio exceeds unity, there is a risk of convergence toward spurious secondary minima.

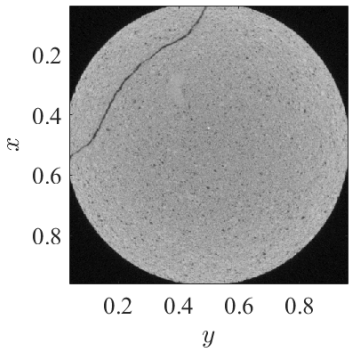
One easy (and cheap) way of enhancing the robustness of DIC and DVC is to coarse-grain the images [22, 23]. This coarse-graining consists of a low-pass filtering with a filter at a characteristic scale of 2 pixels, and a decimation that retains only one pixel out of two in each spatial direction. The result is smaller image whose size has been cut down by a factor of 2^d where d is the space dimension. The resulting pixels can be seen as "super-pixels" that account for the underlying group of 2^d fine scale pixels. The simplest coarse graining step consists in replacing each elementary group of 2^d pixels, by a single super-pixel with the arithmetic average of the gray levels of the group. This coarse-graining operation can be applied recursively, thereby defining a pyramidal construction, where each pyramid level represents the coarsened image of the level below. When expressed in "super-pixels," the maximum displacement decreases by a factor of 2 at each level of the pyramid. Conversely, the correlation length cannot decrease below one pixel, and thus it is easy in a few pyramidal levels to restore a safe convergence criterion. Once convergence has been achieved at a given level n of the pyramid, it

may be used to initialize a new DVC computation at the level $(n-1)$, down to level 0 which conventionally is the original image. In the implementation discussed below, a Gaussian pyramid implementation [24] has been chosen. Figure 3 shows a set of four coarse-grained images from level 3 to 0 (original image) for the example of the section of the X-ray tomography (Figure 1(a)).

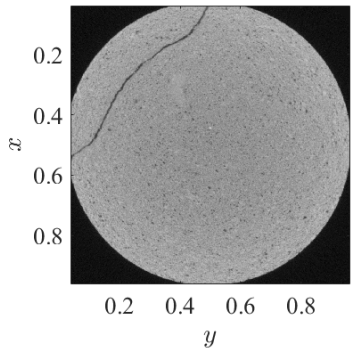


(a) Level 4

(b) Level 3



(c) Level 2



(d) Level 1

Figure 3: Pyramid of X-ray tomographic images (section along median z plane)

In the particular context of multimodal DVC, the only new question to address is how to design the potential Φ at different levels of the pyramid. Pure phases that are present as large grains tend to display a sharper peak in the joint histogram as the pyramid level increases, as noise is very much reduced in the coarsening operations. Phases that are present as very small

grains are progressively assimilated with their surrounding and contribute to broadening their distribution. Hence, there is no general rule that would be independent of the microstructure, and it is suggested to observe the joint histogram and adjust the Φ potential at each level of the pyramid. The multiscale procedure is summarized in Algorithm 1.

Algorithm 1 Multiscale intermodal DVC algorithm

```
Initialize F
for PyramidLevel=n:-1:0 do
    Compute pyramidal f(x)
    Compute pyramidal g(x)
    Compute g̃^(0)(x)
    Compute joint histogram (f, g̃^(0))
    Adjust Φ
    while ‖δF‖ > ε do
        Compute M_mp
        Compute A_m
        Solve M_mp δF_p = A_m
        Update F
        Compute g̃^(n)(x)
    end while
end for
Display residual "distance" (Φ(f, g̃^(n))^2)_,2
Compute joint histogram
Extract segmentation
```

4. Design of potential Φ

4.1. What is expected from the potential?

In such a simple case of having only a few pure phases, it is natural to tailor a specific $\Phi(f, g)$ functional relying on very few parameters. One practical way of designing this potential is to consider the joint histogram of gray levels, and select the points that correspond to the centroids of clustered data that suggest the occurrence of a given phase. Figure 2 shows such a joint histogram using a log scale for the density of probability. It is to be noted that when images are not registered yet, this joint histogram will display spurious populations of voxels whose gray values (f, g) correspond to

different phases. Those points share the abscissa and ordinate of existing (but different) phases. It will be shown in the following that the region around $f \approx 55$ and $g \approx 95$ as seen in Fig. 2 is such a spurious point.

In reality, pixel (or voxel) data are not punctual, but they rather correspond to an integration over a finite area or volume. Thus, voxels are not expected to consist only of pure phases. For those lying across the boundary between two phases i and j, it is expected that the voxel value be $f = \alpha f_i + (1-\alpha) f_j$ and $g = \alpha g_i + (1-\alpha) g_j$, where α ranges continuously over the interval $[0, 1]$. Hence a distribution of gray levels all along the segment connecting two pure phases is expected. The probability density along this line is related to the density of i-j grain boundaries crossed by voxels. The probability is low for a coarse microstructure, and increases for a finer one. Likewise, a term that lowers the penalty given to a mixture between phases may be introduced in the potential Φ.

Let us emphasize that the image resolution itself may have a very significant impact on the joint histogram and hence on the potential to be used. Although trivial, this observation justifies that the potential should in principle be adjusted at each level of a multiscale procedure.

4.2. Likelihood interpretation

The previous discussion points to the design of specific Φ functionals, based on the observation of joint density distribution for $\boldsymbol{h} = (f, g)$ gray levels. The latter potential should ideally be adjusted once registration has been achieved, and this is the very motivation for computing Φ^2. At best, an iterative treatment of the problem may be envisioned where better and better approximations of the potential are looked for.

In fact the previous approach can be rewritten in terms of a probabilistic inference approach. From a registered pair of images, one may compute the probability distribution function of $\boldsymbol{h}$, $p(\boldsymbol{h})$. Then, the likelihood that the two images $f(\boldsymbol{x})$ and $g(\boldsymbol{F} \cdot \boldsymbol{x})$ coincide can be written

$$\mathcal{L}[\boldsymbol{F}] \propto \prod_{\boldsymbol{x} \in ROI} p(f(\boldsymbol{x}), g(\boldsymbol{F} \cdot \boldsymbol{x})) \tag{15}$$

The log-likelihood assumes a convenient form

$$\log(\mathcal{L}[\boldsymbol{F}]) = \sum_{\boldsymbol{x} \in ROI} \log(p(f(\boldsymbol{x}), g(\boldsymbol{F}.\boldsymbol{x}))) + \text{constant} \tag{16}$$

Maximizing the probability or equivalently log-likelihood thus should lead us to the registration solution. Hence it is deduced from the comparison with the postulated form of the DVC functional $\mathcal{T}$ that a natural expression of the potential Φ^2 is (disregarding a useless additional constant)

$$\Phi^2(\boldsymbol{h}) = -\log(p(\boldsymbol{h})) \tag{17}$$

Rather than Φ^2 being a subjectively designed potential, the above log-likelihood interpretation provides a well-defined and objective foundation. The difficulty however, is that the joint distribution and hence the cost function can only be known once registration has been achieved, and the latter calls for the cost function.

Hence a first strategy can be designed whereby it is assumed that the starting point of the registration is already good enough to provide a decent estimate of Φ^2. From it, a registration transformation $\boldsymbol{F}$ is estimated, and hence a corrected joint histogram is computed that allows Φ^2 to be updated, and iteratively, a fixed point solution is sought, which would correspond to registration. In order not to be trapped in local minima, and to compute gradients and curvatures safely, the joint histogram p is first filtered by convolution with a Gaussian kernel whose standard deviation is 3 gray levels (on an 8-bit scale), in order to compute Φ^2 as in Eq. (17). This approach is denoted as LL (for log-likelihood) in the following.

Let us underline that, in contrast with several approaches cited in the introduction, the above procedure for determining Φ is no longer just a reasonable choice, or a particular heuristics, such as the quadratic differences, peak SNR or mutual information. It exploits a functional that provides a maximum likelihood estimator without any *a priori* assumption. The only caveat is that no guarantee for convergence can be shown. It is however clear that close to the solution a converged solution that coincides with the desired one can easily be obtained. The principal limitation is therefore the maximum tolerable distance from the solution that still leads to the solution. In this respect, the multiscale procedure is a way to make the algorithm more robust to large initial mismatches.

4.3. Gaussian mixture approach

A classical way of describing a two-parameter distribution is the Gaussian mixture model [15, 16]. It consists in representing the distribution of gray

levels pairs $\boldsymbol{h} = (f, g)$ as

$$p(\boldsymbol{h}) = \sum_i \phi_i \mathcal{N}(\boldsymbol{h}; \boldsymbol{\mu}_i, \boldsymbol{\sigma}_i) \tag{18}$$

where $\mathcal{N}(\boldsymbol{h}; \boldsymbol{\mu}, \boldsymbol{\sigma})$ is the normal distribution of $\boldsymbol{h}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\sigma}$. Introducing

$$\lambda_i(\boldsymbol{h}) = (1/2)(\boldsymbol{h} - \boldsymbol{\mu}_i)\boldsymbol{\sigma}_i^{-1}(\boldsymbol{h} - \boldsymbol{\mu}_i) \tag{19}$$

p reads

$$p(\boldsymbol{h}) = \sum_i \exp(-\lambda_i(\boldsymbol{h}) + \log(\phi_i)) \tag{20}$$

if the Gaussians are well separated, then

$$\begin{array}{rcl} p(\boldsymbol{h}) & \approx & \max\limits_i \left[\exp(-\lambda_i(\boldsymbol{h}) + \log(\phi_i))\right] \\ & \approx & \exp\left[-\min\limits_i(\lambda_i(\boldsymbol{h}) - \log(\phi_i))\right] \end{array} \tag{21}$$

Using the previous relationship between p and Φ^2, the following simple expression of the potential results

$$\Phi^2(\boldsymbol{h}) = \min_i(\lambda_i(\boldsymbol{h}) - \log(\phi_i)) \tag{22}$$

This second strategy (labeled GM in the following) appears comparable to the previous one in the sense that Φ^2 still represents the negative log-likelihood. However, as compared to the previous LL approach, which is a plain depiction of the joint histogram, a specific model for the joint distribution is proposed that naturally focusses on the most populated peaks. Therefore spurious secondary peaks may be ignored if the number of Gaussian components is small enough, and if an approximate registration is available initially. Thus, this GM approach is expected to help convergence from more distant initial registration than LL. However, because the approximation involved is forcing a simplifying model in GM, it is anticipated that the uncertainty of the finally obtained registration will be larger than that of the LL approach.

Let us also mention a side benefit of the GM approach: as the potential appears as consisting of a few paraboloids, the $\boldsymbol{h}$ plane is naturally partitioned into sub-domains each belonging to a single paraboloid, the boundaries between subdomains being simple conic curves. The label i where the minimum is reached in Eq. (22) can be assigned to each gray level pair $\boldsymbol{h}$,

$\psi(\boldsymbol{h}) = i$. Each subdomain i is characterized by a "phase" whose signature is the gray level pair $\boldsymbol{\mu}_i$ of the apex of the paraboloid. Moreover, a linear function and a constant function can be introduced respectively for $[\Phi^2]_{,2}$ and $[\Phi^2]_{,22}$, in each subdomain of this partition. After registration, assigning the label $\psi(\boldsymbol{h})$ to each voxel where the gray level pair is $\boldsymbol{h}$ allows the image to be directly segmented based on the signatures of each phase.

Such a segmentation is not directly available following the LL approach, although one could still consider the basin of attraction of each minimum of the Φ^2 potential using a steepest descent.

5. Test case

In order to illustrate the performance of both proposed approaches, LL and GM, a 3D case is chosen where two modalities X-ray and neutron tomographies have been performed with a similar resolution.

5.1. Experimental details

The studied material is a Bentheim sandstone with an average porosity of about 23 %, a mean grain diameter of 300 μm and a composition of 95 % quartz, 3% kaolinite and 2 % orthoclase **[SR Erika: wt.%, vol% or at%??? – I can't find this information in the paper, I asked a colleague, waiting for the answer]** [28]. The sample shape is cylindrical with 50 mm diameter and 100 mm height. It was scanned independently using X-ray and neutron tomography.

Both measurements were carried out at the Helmholtz-Zentrum Berlin (HZB). Neutron images were acquired at the beamline CONRAD [25] with L/D of 500, a voxel size of about 30 μm and an acquisition time for each radiography of 30 seconds, which gives a total time of about 6 hours for a complete scan of 600 images. X-ray tomographies were acquired through a lab source with a voltage of 120 kV, a current of 83 μA, a voxel size of about 30 μm and an acquisition time for each radiography of 1.3 s, which gives a total time of about 2.5 hours for a complete scan of 1300 projections (each projection resulted from the mean of 5 repeated images). To compare ...

In the following, an un-optimized Matlab implementation has been used to illustrate the feasibility of the approach. In order to limit memory usage, two volumes of height $\Delta z = 128$ voxels where first extracted in both 3D-images. They were also 2×2-binned in the transverse (x, y) directions so that the cross-sections were 900×900 voxels for both modalities. Last, they were

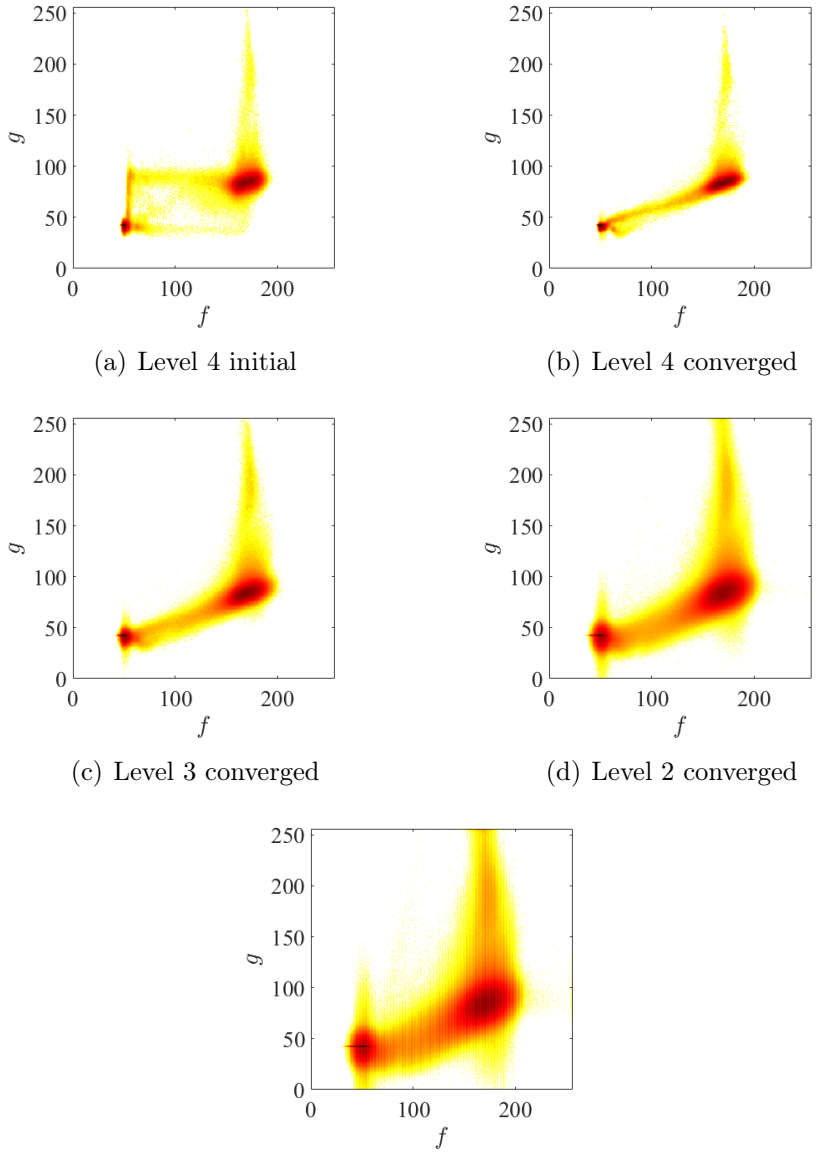
re-encoded into 8 bits. A four-level pyramidal scheme was used where level 1 corresponds to the $900 \times 900 \times 128$ volume. However, in order to preserve a decent thickness at all floors of the pyramid, it was chosen not to coarse-grain the z direction. Hence at level 4, the volume shape was approximately cubic $112 \times 112 \times 128$, but its microstructure was highly anisotropic as each level-4 voxel resulted from an original $8 \times 8 \times 1$ group of voxels in the original image. The price to pay for this anisometric re-scaling is that one should be careful when transferring the result of one level as the initialization of the next one, and resorting to physical coordinates rather than voxel numberings is a simple way to avoid ambiguities.

5.2. Results of the GM approach

The first test was performed using the Gaussian mixture (GM) route developed in Section 4.3. Although automatic selection algorithms are available to fit Gaussian mixture distributions, our attempts to use them lead to rather poor fits (more often than not the means were pushed to the limits of accessible gray levels (0 or 255). Thus a simple procedure was designed for fitting the highest peaks, fitting the highest of the residual between the measured joint histogram and the already identified Gaussian components recursively. Albeit simple, this procedure reveals itself to be robust and adequate for our usage. The number of components was *a priori* prescribed from the shape of the joint histogram. As the latter is highly dependent on the pyramid level, it was selected independently per level. Here, it went from two components for level 4 to four components at level 1. At the coarsest level, choosing two Gaussian components forbids the spurious third peak (see Fig. 4(a)) to be mistaken for a real one, thereby favoring its annihilation with the transformation $\boldsymbol{F}$.

Figures 4(b-e) show the change of the joint histogram of both images after registration at different levels of the pyramid. It is observed that the spurious patterns that appear as horizontal or vertical lines in Fig. 4(a) already disappeared after convergence at level 4. The joint histogram appears already at this stage to display very few signs of mismatch. At lower levels, only a broadening of the histograms is visible, which simply results from the higher resolution and the higher noise magnitude.

Table 1 gives the number of iterations at each level of the pyramid, together with the estimated parameters of the transformation $\boldsymbol{F}$. The convergence criterion is chosen to be the stationarity of the deformation gradient, and relaxation is stopped when the change between two iterations is less than



(a) Level 4 initial

(b) Level 4 converged

(c) Level 3 converged

(d) Level 2 converged

(e) Level 1 converged

Figure 4: (a) $\log_{10}$-scale joint histogram at various stages of the GM registration from (a) the initial step at level 4, and after registration at levels 4 to 1 respectively in (b) to (e) of registered image

Table 1: Evaluation of the transformation parameters at convergence for each pyramid level, using the GM approach. The convergence criterion was chosen to be such that the norm of the deformation gradient part of $\boldsymbol{\delta F}$ should be less than 5×10^{-4}

Pyramid level (iterations)		$L = 4$ (47)	$L = 3$ (12)	$L = 2$ (12)	$L = 1$ (10)
ϵ_{xx}	(%)	1.78	1.72	1.69	1.66
ϵ_{yy}	(%)	1.75	1.68	1.63	1.60
ϵ_{zz}	(%)	-2.04	-2.90	-2.72	-2.88
ϵ_{yz}	(%)	-0.36	-0.24	-0.21	-0.21
ϵ_{xz}	(%)	-0.21	-0.01	0.03	0.03
ϵ_{xy}	(%)	-0.05	0.01	0.04	0.04
ω_x	(deg.)	-0.11	-0.07	-0.06	-0.07
ω_y	(deg.)	0.43	0.52	0.58	0.59
ω_z	(deg.)	0.59	0.68	0.70	0.71
T_x	(vox.)	-14.1	-15.7	-16.1	-16.0
T_y	(vox.)	-11.8	-10.1	-9.4	-9.1
T_z	(vox.)	9.8	9.1	9.3	9.6

5×10^{-4} or when the number of iterations reaches a maximum number here chosen to be 60 (with a minimum of 10). It is observed that the number of iterations is large at level $L = 4$ (where the volume is much smaller than at later stages and hence computation is much faster). After this first step, the number of iterations is close to the imposed minimum. Consistently, it is observed that the transformation is already very well determined after convergence at level $L = 4$. Further corrections from one level to the next are of order 10^{-3} in strain. Only the ϵ_{zz} strain shows a larger correction from $L = 4$ to $L = 3$, where it almost reaches its final stabilized value. It is also interesting to observe that directions x and y behave in a similar fashion, with very close values. However, they both differ quite significantly from direction z. Since the latter is the rotation axis, it is deduced that it is not coupled with the perpendicular directions that, in contrast, are processed together in the reconstruction, and hence cannot be different. Thus, one possible explanation of this difference is that one detector array is not square but rather rectangular.

As above discussed, a residual field, $\Phi^2_{,2}$, is constructed at convergence.

It is shown in Figure 5(a). The fact that neither f or g, can be recognised in the residual field is an indication that the result can be trusted.

One output of the proposed registration is a natural segmentation. Starting from any gray level pair (f, g), following the steepest descent in the potential, directly provides a partition into (here) four basins of attraction, and hence any (f, g) can be labelled by the indices of the different basin roots. This labelling of each voxel by a specific "phase" is a natural segmentation issued from the proposed approach. Figure 5(b) shows the result of such a labelling issued from the GM approach. It is noteworthy that no other treatment has been performed (such as spatial filtering or any other morphological operator) to dampen the noise.

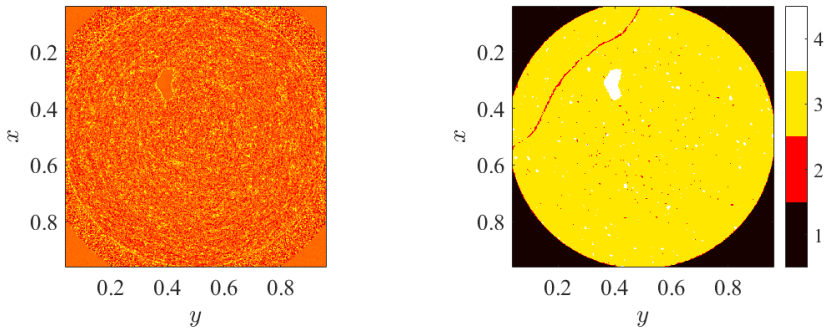
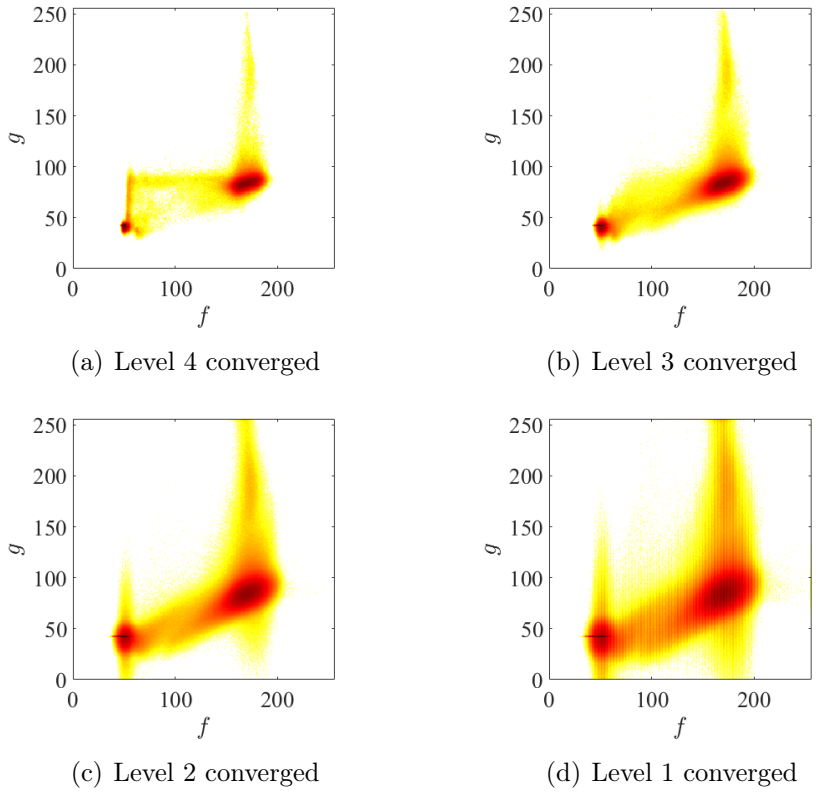


Figure 5: (a) Residual at convergence (level 1) and (b) segmented map (with four phases) after registration using the GM approach (section along median z plane)

5.3. Results of the arbitrary likelihood LL approach

The advantage of the LL approach lies in the fact that there is less room for personal judgement, as the potential is based on the observed histogram. However, as could be seen in the previous section in Fig. 4, depending on the pyramid level, the joint histogram itself appears to change significantly across pyramidal levels. In particular at the initial state, registration is rather poor and the resulting spurious features in the joint histogram will not be corrected in the present approach as they were in the GM approach (simply because very few (*i.e.,* 2) components were chosen initially). Here the only processing of the joint histogram is a convolution by a Gaussian whose width

is three gray levels in order to remove high frequency fluctuations and hence local minima.



(a) Level 4 converged

(b) Level 3 converged

(c) Level 2 converged

(d) Level 1 converged

Figure 6: Observed joint histograms after convergence at four pyramidals level from $L = 4$ to $L = 1$

Figure 6 shows the initial potential at different pyramid levels together with the converged joint histogram (using the same log-scale). As anticipated, because some spurious features are present in the initial histogram, the early convergence is quite poor and at the end of the $L = 4$ relaxation, very few changes occur. In spite of this, at level $L = 3$, the vertical branch at the lower f values disappears almost entirely. The following levels are similar to the GM case, as well as the final joint histogram.

Table 2 gives the results of the DVC registration based on the LL approach. It is observed first that the number of iterations, in the higher levels, is much larger than for the GM approach as the maximum iteration number

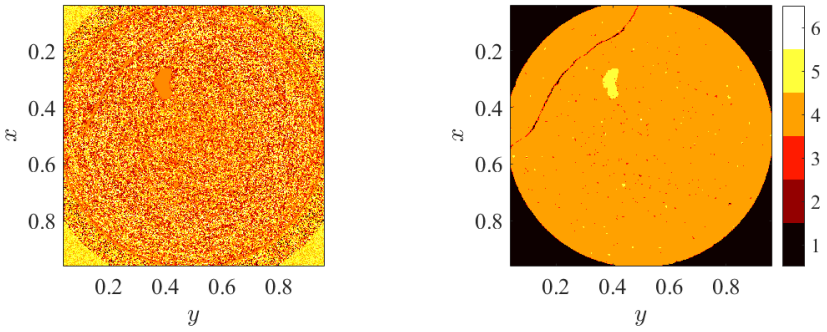
is reached. It is also observed consistently with the previous observations that the estimation of the transformation $\boldsymbol{F}$ at level 4 is totally wrong. After convergence at level $L = 3$, much more decent values are obtained, and show for the most part a good stability until the end of the registration at $L = 1$. In particular $L = 2$ and $L = 1$ results show only minute changes. Moreover, some of these estimates of the transformation $\boldsymbol{F}$ are in very good agreement with the previous approach. All strain components differ from the GM estimates by less than 10^{-4}, but the ϵ_{zz} whose difference with the GM estimate is close to 1%. Rotations are also almost identical. Translation differences are also in the voxel range (and 2 voxels along z). One possible reason for larger differences in the z-direction may come from the fact that the sample size is also much shorter along the z axis (128) than in the transverse directions (900).

Table 2: Evaluation of the transformation parameters at convergence for each pyramid level, using the LL approach. The convergence criterion was chosen to be such that the norm of the deformation gradient part of $\boldsymbol{\delta F}$ should be less than 5×10^{-4}

Pyramid level (iterations)		$L = 4$ (60)	$L = 3$ (60)	$L = 2$ (41)	$L = 1$ (10)
ϵ_{xx}	(%)	0.24	1.58	1.63	1.60
ϵ_{yy}	(%)	0.10	1.39	1.53	1.52
ϵ_{zz}	(%)	-2.91	-3.87	-1.82	-1.96
ϵ_{yz}	(%)	-0.64	-1.38	-0.24	-0.21
ϵ_{xz}	(%)	-0.26	0.66	0.01	0.03
ϵ_{xy}	(%)	0.10	0.16	0.03	0.04
ω_x	(deg.)	-0.22	-0.61	-0.03	-0.02
ω_y	(deg.)	0.15	0.85	0.70	0.72
ω_z	(deg.)	0.13	0.66	0.73	0.73
T_x	(vox.)	-4.2	-16.7	-16.1	-15.9
T_y	(vox.)	0.7	-6.4	-7.9	-7.9
T_z	(vox.)	9.5	12.1	11.6	11.6

Figure 7(a) shows the residual map at the final step. This again shows a good convergence where no bias is visible. As previously discussed, for the GM approach, the different Gaussian components could be associated with a phase. At convergence, for each voxel, the pair of gray levels of both

modalities could be ascribed to a specific phase. For the LL case, a similar albeit less obvious procedure, is designed by identifying pure phases (*i.e.,* labels) at maxima of the joint distribution. Then using a steepest ascent algorithm, any gray level pair will flow toward a maximum that defines the corresponding phase. This procedure groups gray levels into domains from the basin of attraction of the maxima. The smoothing of the distribution using a Gaussian filter as earlier discussed for the LL approach limits the number of maxima. Figure 7(b) shows the distribution of phases, where the number of phases has been set to 6 without prior subjective statement.



(a) Level 1 (b) Level 1

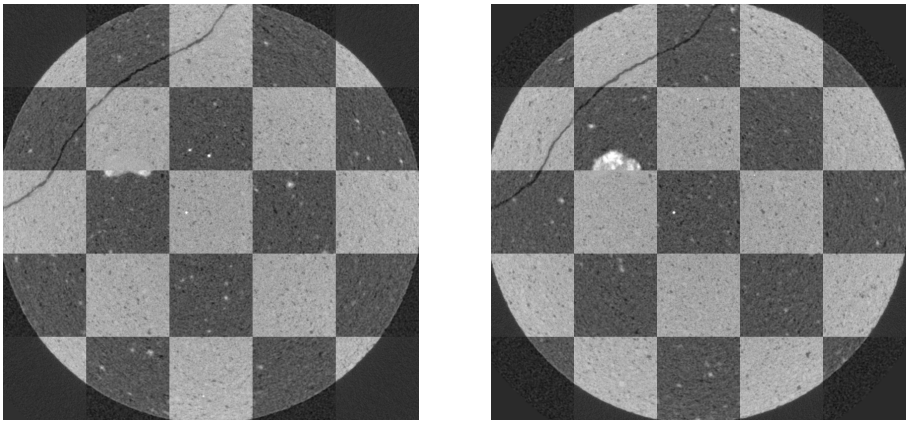
Figure 7: (a) Residual field after registration using the LL approach. (b) Segmented map (section along median z plane)

In order to check whether the observed difference is due to a slow relaxation preventing to reach the actual solution, it was tried to initialize one approach with the result of the other one. Each approach led to the same results as those already reported (*i.e.,* with no influence of the initialization), so that slow convergence cannot be incriminated. Thus, the observed difference in the transformation between both approaches presumably comes from the joint distribution function, which is not adequately depicted in the GM approach. In contrast, the LL approach that does not postulate priors to describe the distribution gives a more reliable picture, and hence it appears as more satisfactory, although it is difficult to provide a more objective appreciation of this from the registration results.

Last, it is worth emphasizing that tailored potentials may be designed on demand in order to achieve a registration based on a specific prior judgement

on the trustful region of the joint histogram. However, a systematic procedure that would allow for a simplified and yet faithful picture of the joint distribution is the most desirable. Two cases were shown as examples, the first one, GM, was presumably oversimplifying while the second, LL, was not constraining enough.

In order to check the quality of registration, Fig. 8 shows two complementary mosaic pictures, where after registration composite images by assembling checkerboards composed of squares taken from one modality or the other. This allows continuities to seen with ease whenever relevant, and also to observe that some patterns are hardly visible in one modality and very noticeable in the other one.



(a) (b)

Figure 8: Complementary composite images formed by a checkerboard assembly of squares coming form the two modalities after registration. The absence of discontinuity can be checked along all boundaries

6. Conclusions

It has been shown that the registration of two images acquired with different modalities can be performed through an extension of DVC through the minimization of a potential Φ^2. Furthermore, it is shown that if this potential coincides with (the cologarithm of) the joint probability distribution, then the registration solution is the one that maximizes the likelihood. From this observation two variants for the potential were formulated:

- a Gaussian mixture model (GM) where the joint distribution is fitted to a set of a few Gaussian distributions
- a less intrusive choice (LL) where the measured joint distribution is directly used to design the potential used in the registration.

A real case study on two-modality tomographic images allows the two above mentioned algorithms to be probed. The GM algorithm is very stable with a fast convergence in the multiscale framework. Although less stable, and displaying a slower convergence, the LL algorithm is also able to provide a good registration for which the mean transformation parameters were in good agreement with those obtained from the GM algorithm.

A voxel accuracy can be reached, revealing, in the case under study, artifacts related presumably to anisometric detectors: the strain along the tomographic rotation axis direction and perpendicular to it shows a few percent difference. Additionally, registration may lead to a segmentation of phases according to their contrast in both modalities.

The registration of these two-modality images opens the way to enrich/correct a low resolution, low contrast, or noisy image from one modality with the other one. For the exploitation of the simultaneous dual X-ray and neutron tomography (instead of separate acquisitions [26]), project "NEXT-Grenoble", such an approach should be precious in enhancing the dialog between both acquisitions.

Acknowledgement

[SR We acknowledge the support of XXXXX, and useful discussions with YYYY.]

[SR Helmholtz-Zentrum Berlin ? ILL ? ANR ?]

References

[1] R. Xia, J. Zhao, Y. Liu, A robust feature-based registration method of multimodal image using phase congruency and coherent point drift, Eighth International Symposium on Multispectral Image Processing and Pattern Recognition, 8919, 891903, (2013)

[2] D.K. Sahu, M.P. Parsai, Different image fusion techniques — a critical review, International Journal of Modern Engineering Research, 2(5), 4298-4301, (2012)

[3] C.S. Xydeas, V. Petrovic, Objective image fusion performance measure, Electronics Letters, 36(4), 308-309, (2000)

[4] A. Papoulis, *"Probability, Random variables and Stochastic processes"*, McGraw-Hill, New York, (1984)

[5] B. Zitova and J. Flusser, Image registration methods: a survey, Image and Vision Computing, **21**, (11), 977–1000, 2003.

[6] J. P. Pluim, J. A. Maintz, and M. A. Viergever, Image registration by maximization of combined mutual information and gradient information, *in* International Conference on Medical Image Computing and Computer-Assisted Intervention, 452–461, Springer, 2000.

[7] P. Viola and W. M. Wells III, Alignment by maximization of mutual information, International Journal of Computer Vision, **24**, (2), 137–154, 1997.

[8] T. Gaens, F. Maes, D. Vandermeulen, and P. Suetens, Non-rigid multimodal image registration using mutual information, *in* International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 1099–1106, Springer, 1998.

[9] A. Roche, G. Malandain, X. Pennec, and N. Ayache, The correlation ratio as a new similarity measure for multimodal image registration, *in* International Conference on Medical Image Computing and Computer-Assisted Intervention, 1115–1124, Springer, 1998.

[10] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, A viscous fluid model for multimodal non-rigid image registration using mutual information, Medical Image Analysis, **7**, (4), 565–575, 2003.

[11] C. Chefd'Hotel, G. Hermosillo, and O. Faugeras, Flows of diffeomorphisms for multimodal image registration, *in* Proceedings of IEEE International Symposium on Biomedical Imaging, 2002, 753–756, IEEE, 2002.

[12] M.A. Sutton, J.-J. Orteu, H.W. Schreier, *"Image Correlation for Shape, Motion and Deformation Measurements — Basic Concepts, Theory and Applications"*, Springer (New-York, NY, USA, 2009)

[13] P.K. Rastogi, E. Hack eds., *"Optical Methods for Solid Mechanics: A Full-Field Approach"*, Wiley-VCH, (Weinheim, Germany), (2012)

[14] A. Guery, F. Latourte, F. Hild, S. Roux, Characterization of SEM speckle pattern marking and imaging distortion by Digital Image Correlation, Measurement Sci. & Tech. **25**, (1), 015401, (2013)

[15] A. Kaestner, M. Morgano, J. Hovind, E. Lehmann, Bimodal Imaging Using Neutrons and X-rays, Proc. Digital Industrial Radiology and Computed Tomography, (2015)

[16] A. Kaestner, D. Mannes, J. Hovind, P. Boillat, E. Lehmann, Combined neutron and X-ray imaging on different length scales, Proc. 6th Conference on Industrial Computed Tomography, (2016)

[17] J.J. Orteu, 3-D computer vision in experimental mechanics, Optics and Lasers in Engineering, **47**(3), 282–291, (2009)

[18] A. Charbal, J.-E. Dufour, F. Hild, M. Poncelet, L. Vincent, S. Roux, Hybrid stereocorrelation using infrared and visible light cameras, Exp. Mech. **56**, 845–860, (2016)

[19] R. A. Brooks, G. Di Chiro, Beam hardening in x-ray reconstructive tomography, Physics in Medicine and Biology, **21**(3), 390, (1976).

[20] P. Helnwein, Some Remarks on the Compressed Matrix Representation of Symmetric Second-Order and Fourth-Order Tensors, Computer Methods in Applied Mechanics and Engineering. **190**, (22-23), 2753–2770, 2001.

[21] B. Wagne, S. Roux, F. Hild, Spectral Approach to Displacement Evaluation from Image Analysis, European Physical Journal: Applied Physics **17**, 247–252, (2002)

[22] F. Hild, B. Raka, M. Baudequin, S. Roux, F. Cantelaube, Multi-scale displacement field measurements of compressed mineral wool samples by digital image correlation, Appl. Optics **41**, 6815–6828, (2002)

[23] S. Roux, F. Hild, P. Viot, D. Bernard, Three dimensional image correlation from X-Ray computed tomography of solid foam, Composites Part A **39**, 1253–1265, (2008)

[24] P. Burt, E. Adelson, The Laplacian Pyramid as a Compact Image Code, IEEE Transactions on Communications, **COM-31**(4), 532–540, (1983)

[25] Hilger, A., Kardjilov, N., Strobl, M., Treimer, W., Banhart, J., The new cold neutron radiography and tomography instrument conrad at hmi berlin, Physica B: Condensed Matter **385**, 1213-1215, (2006)

[26] E. Tudisco, S.A. Hall, E.M. Charalampidou, N. Kardjilov, A. Hilger, H. Sone Full-field measurement of strain localisation in sandstone by neutron tomography and 3D-volumetric digital image correlation. Physics Procedia 69, 509–515, (2015)

[27] T. Ranchin, L. Wald, M. Mangolini, The ARSIS method: a general solution for improving spatial resolution of images by the means of sensor fusion, Fusion of Earth Data, Proceedings EARSeL Conference, Cannes, France, 6-8 February (1996), (Paris: European Space Agency)

[28] Klein, E., Baud, P., Reuschlé, T., Wong, T., Mechanical behaviour and failure mode of bentheim sandstone under triaxial compression. Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy **26**, 21-25,(2001)