A GENERIC APPROACH TOWARDS AUTOMATIC IMAGE CO-REGISTRATION

R. Fransens [a], C. Strecha[a], G. Caenen [a,b], L. Van Gool[a,c]

[a] KUL, Dept. of Electrical Engineering, Kasteelpark Arenberg 10, B-3001 Heverlee,
{fransen,cstrecha,caeneng,vangool}@esat.kuleuven.be
[b] IncGEO, Research Campus Hasselt, Kempische Steenweg 297/2, B-3500 Hasselt
geert.caenen@incgeo.be
[c] ETH, Dept. of Information Technology and Electrical Engineering, Sternwartstrasse 7, CH-9092 Zurich
vangool@vision.ee.ethz.ch

KEY WORDS: Image co-registration, Mutual Information

ABSTRACT:

This paper presents a mathematical framework for dealing with area-based image co-registration problems in a generic and modular way. Based on this framework an efficient implementation is devised allowing "plug & play" support for a whole gamut of geometric transformations, image similarity measures (criteria) and optimisation methods. A typical drawback of area-based methods is the exhaustive use of memory and the slow optimisation speed. Therefore, this paper outlines an effective subsampling strategy that considerably speeds up the registration process. The final part of the paper is devoted to an extensive evaluation of the algorithm using both ground-truth data and a series of challenging real examples.

1. INTRODUCTION

1.1 Image Co-registration

Accurate co-registration of remotely sensed imagery is an important step in the analysis of Earth Observation data. For example, time series analysis, data fusion and change detection all require accurately co-registered images in order to produce useful and meaningful results.

The problem of image registration has been studied extensively from different viewpoints among which computer vision, medical imaging and remote sensing. Currently, a whole gamut of registration techniques is available, many of which rely on the interaction of a human operator. The classical approach requires the manual selection of homologue points. The global transformation is than interpolated from the known displacements of these points using different methods. It is important to note that the image content is only used by the human operator when selecting and matching the control points.

Automatic co-registration, on the other hand, exploits the image content for finding the correct alignment of images. A first class of automatic registration approaches attempts to automate the process of feature selection and feature matching. These *feature-based* methods are fast and work particularly well when the images at hand have similar spectral content. Special care must be taken to use robust descriptors and matching techniques that account for, or are invariant to, deformations of the image. Features can be corners (e.g. [1]), wavelet features (e.g. [3]), edges, contours (e.g. [2]) or even entire image segments that resulted from an image segmentation. In the literature, a second broad class of automated approaches is categorised as *area-based* registration. In this case, information from all pixels is accounted for. These approaches are much slower and memory consuming but have proven successful in registering multi-modal or heterogeneous data types. Unlike feature-based methods, the initial alignment of the two images should be reasonably well. Fortunately, in remote sensing, most image data is accompanied with a (rough) georeference. For a more elaborate overview of past and current registration techniques we refer to surveys in the literature: [4],[5].

1.2 Outline of the paper

This paper describes a framework that gives rise to a generic and modular implementation of area-based registration methods. To overcome the slow optimisation, inherent to area-based methods, a subsampling procedure is proposed, giving rise to a *hybrid* registration method.

The rest of this paper is organised as follows. Section 2 introduces the necessary mathematics for formulating the registration problem as an optimisation problem. Implementation issues as well as improvements (multi-resolution strategy and subsampling) are discussed. Next, section 3 and 4 elaborate respectively on the geometric transformations and similarity measures that are supported by the framework. In section 5 a series of experiments is carried out on ground-truth and real data. Finally, section 6 concludes the paper with a summary of the registration results and a discussion on how the current framework can be improved and extended towards applications such as change detection.

2. MATHEMATICAL FORMULATION

2.1 Generic Co-registration Framework

For the sake of clarity, we start with a few notational conventions that will considerably simplify the expressions in this section. An image function is a differentiable function I, that maps pixels $\mathbf{x} = [x\ y]^T$ from a domain $D \subset \mathbb{R}^2$ to an n-dimensional set of spectral values. A raster image $\mathbf{I}$ is considered as a sampled version of I. Given a grid of sample coordinates (usually a lattice) $\{\mathbf{x}_i\ ,\ i = 1 \ldots N\}$, the raster image $\mathbf{I}$ is constructed as a vector of grey values:

$$\mathbf{I} = \left[I(\mathbf{x}_1) \ldots I(\mathbf{x}_N) \right]^T \qquad (1)$$

Multi-band raster images can be represented as matrices, storing the different spectral bands in different columns. In the following, we will consider one-dimensional (grey-value) raster images to avoid complex notations. The reader should bear in mind that the results of this section apply to multi-band images.
In practice, the image function I is not known. In fact, all that is known is a sampling of I on a certain grid (a raster image $\mathbf{I}$). Because the continuity (and smoothness) of I is crucial to our approach, the image function I is defined as an interpolation of the available raster image $\mathbf{I}$. For our experiments we used bilinear interpolation which is fast and meets the desired smoothness property.
A geometric transformation T maps points $\mathbf{x}$ in the plane to other points in the plane. Parametric transformations $T(\bullet;\mathbf{p})$ can be described completely by means of a parameter vector $\mathbf{p}$. For example, a rigid transformation has three parameters: the two components of the translation vector and a rotation angle.

$$T(\mathbf{x};\underbrace{t_x,t_y,\theta}_{\mathbf{p}}) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \mathbf{x} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \qquad (2)$$

If we want to transform a raster image, this is done by adjusting its sampling grid according to the transformation.

$$\mathbf{I}(\mathbf{p}) = \left[I(T(\mathbf{x}_1;\mathbf{p}))...I(T(\mathbf{x}_N;\mathbf{p})) \right]^T \qquad (3)$$

Similarity of raster images is expressed by a real-valued similarity measure or criterion function C, that takes two vector (matrix) arguments of the same size.

$$C : (\mathbf{X},\mathbf{Y}) \mapsto C(\mathbf{X},\mathbf{Y}) \qquad (4)$$

For example, the Sum of Squared Differences (SSD) criterion computes the Euclidian distance between the vectors $\mathbf{X}$ and $\mathbf{Y}$.

$$SSD(\mathbf{X},\mathbf{Y}) = \sum_{i=1}^{N} (X_i - Y_i)^2 \qquad (5)$$

Co-registration of an image I_2 (called the "slave" image) with a "master" image I_1 boils down to finding the correct parameter vector $\mathbf{p}$ that yields an "optimal" criterion value, i.e.

$$f(\mathbf{p}) = C\left(\mathbf{I}_1, \mathbf{I}_2(\mathbf{p})\right) \qquad (6)$$

attains an extremum at $\mathbf{p}$. The nature (maximum or minimum) of this extremum depends on the criterion. The problem of co-registration has now been formulated as an optimisation problem. The evaluation of the objective function f for a given parameter vector $\mathbf{p}$ is computed in 3 steps.

1. compute the transformed grid $T(\mathbf{x}_i;\mathbf{p})$
2. create $\mathbf{I}_2(\mathbf{p})$ by sampling I_2
3. compute $C(\mathbf{I}_1,\mathbf{I}_2(\mathbf{p}))$

Most analytic optimisation algorithms require the gradient of the objective function to implement an efficient strategy for attaining the optimum (e.g. gradient ascent). In our case the gradient of the objective function takes on the form

$$\frac{\partial f}{\partial p_j} = \sum_{i=1}^{N} \frac{\partial C}{\partial Y_i}\left(\mathbf{I}_1, \mathbf{I}_2(\mathbf{p})\right) \cdot \left\langle \frac{\partial I_2}{\partial \mathbf{x}}\left(T(\mathbf{x}_i;\mathbf{p})\right), \frac{\partial T}{\partial p_j}(\mathbf{x}_i) \right\rangle \qquad (7)$$

where we applied the chain rule to resolve the dependencies. Here, $\langle\bullet,\bullet\rangle$ denotes the inner product in $\mathbb{R}^2$. It is certainly worthwhile taking a closer look at the expression for the derivative in equation (7), as a few components represent well-known quantities. The first factor is the derivative of the criterion with respect to its second argument. The second factor is the image gradient function of the slave image and the third factor is the Jacobian of the transformation. In summary, computing the gradient of the objective function consists of the following consecutive steps:

1. Compute the Jacobian of T on the grid points
2. Evaluate the image gradient of I_2 on the transformed grid points.
3. Compute the derivative w.r.t. its second argument and evaluate it on the master image and the transformed slave image.
4. Combine the results of 1,2 and 3 using (7).

2.2 Object Oriented Implementation

Object oriented programming allows generic and modular implementation of the proposed co-registration framework.

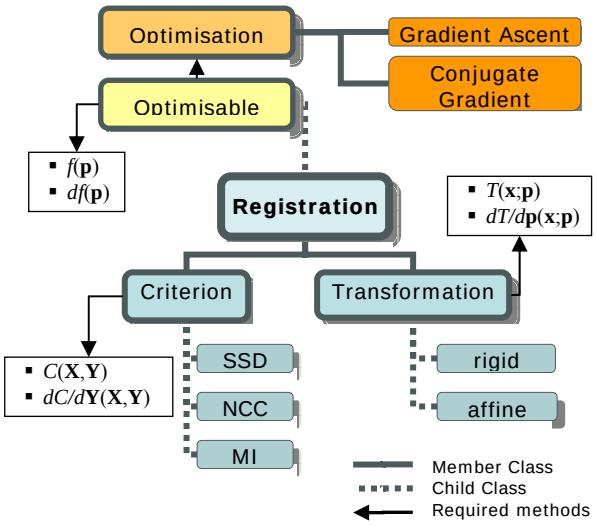


Figure 1 Simplified Object Model for co-registration. The registration algorithm is formulated as an optimisation problem. Hence it is a derived class from Optimisable. The latter object is required to implement the objective function f and its gradient df. Registration will provide this implementation using concatenations of implementations from Criterion and Transformation. For the sake of clarity, image data has been left out of the diagram.

Based on the principle of polymorphisms, it suffices to create abstract classes (interfaces) for a criterion, a transformation and an optimisation method. Then, specific implementations for criteria and transformations can be derived from these interfaces, provided that they implement evaluations and derivatives. As we formulated the co-registration as an optimisation problem, the co-registration class itself is a derived class from an optimisation problem (i.e. optimisable class). An optimisable class is required to implement an objective function

and a gradient. The latter can be obtained from its members: a criterion, a transformation and the images. Different optimisation strategies can avail of this information to attain the desired optimum. Figure 1 summarises this scheme.

2.3 Multi-resolution approach

Non-convex optimisation often suffers from local minima, especially when the initial "guess" is too far off. Therefore, for each image, a pyramid of downsampled versions is created. The effect of this operation is twofold: at higher pyramid levels the amount of geometric distortion decreases (in terms of pixel displacement) and the objective function becomes smoother as the image data is smoothed.
The co-registration is started at the highest pyramid level, level 0 and works is way down the pyramid as follows. After optimisation, the "solution" $\mathbf{p}^i$ at pyramid level i is "upscaled" to level $i+1$ ($\mathbf{p}^{i+1}$) and used as an initialisation for the next pyramid level. The upscaling for each parameter p_i has to be done carefully, as the upscaling transform causes not only a scaling but also a small offset of the canonical coordinate system of both image rasters.

2.4 Subsampling

One of the obvious shortcomings of area-based co-registration is that the exhaustive use of all pixels causes the algorithm to run slowly, especially for criteria that yield a high computational complexity. Both the objective function and its gradient can be approximated using only a subset of the pixels: the raster image $\mathbf{I}_1$ can be reduced to a percentage of all pixels accompanied by their locations $\mathbf{x}$. This set of locations can be considered as a special grid. In general, a random sample of the master image domain will provide arbitrarily good approximations, depending on the number of samples. In some cases, a smaller, biased sample could be more efficient (e.g. interest points).

2.5 Optimisation methods

We experimented with different optimisation methods, including conjugate gradient, quasi-newton (DFP and BFGS) and variable step gradient descent. We obtained the best results with the simple gradient descent optimisation strategy. Furthermore, gradient descent also compares favourable to the other methods in terms of execution time. The reason for this probably lies in the non-convexity of the image co-registration criterion functions.

3. GEOMETRIC TRANSFORMATIONS

In our framework we distinguish between parametric transformations and non-parametric transformations. The main reason for this is that the optimisation techniques for both types are quite different.

3.1 Parametric Transformations

The tool supports all well-known parametric transformations: translation, rigid transformation (2), similarity transformation, affine transformation, planar homography and linear spline. These "classical" parametric transformations account for global differences in alignment.
In case of depth or relief related disparities a localised transformation model is required. Examples of such families are thin-plate splines (TPS), higher order polynomials, weighted mean (WM), piecewise linear (PL),… A recent study [6] pinpoints the pro's and cons of these non-rigid transformations. Most of these transformation families require a set of homologue points identifying matches between the master and slave image. In our setting, envisaging fully automatic registration, such information is not available. However, if a set of control points is available only for the master image, such transformations can be parameterised in terms of the displacements of the control points, i.e.

$$\mathbf{p} = \begin{bmatrix} dx_1 & dy_1 & \ldots & dx_n & dy_n \end{bmatrix}^T \tag{8}$$

Control points (x_i,y_i) can be either automatically selected on a regular grid, or using an interest point detector.

3.2 Non-Parametric Transformations

A broad-purpose type of geometric distortion is optical flow (OF). Here, the displacement of pixels is only constrained by a global (or local) smoothness. As such, OF is able to recover from local discontinuities without needing to identify a control point in this neighbourhood. The usage of optical flow for co-registration is explained thoroughly in [9].

4. SIMILARITY MEASURES

Similarity measures quantify the match between the spectral content of the master and slave image. Depending on the type of both images (sensor, camera parameters,…) the measure should account for the spectral transformation at hand. In this paper, we discuss three criteria, with increasing complexity: Sum of Squared Differences (SSD) (5), Normalised Cross-correlation (NCC) and Mutual Information (MI). SSD should be used when the master and slave image are captured with the same sensor under similar conditions, i.e. the photometric transformation between corresponding spectral values should be close to the identity transformation. When, due to different illumination conditions, an (approximately) linear photometric transformation occurs between master and slave, NCC is able to account for this. Finally, if the data is heterogeneous (different sensor, different number of bands, …) MI is the only remaining choice.

4.1 Parametric Models : SSD and NCC

Both SSD and NCC implicitly assume a parametric model for the transformation between spectral values:

$$\begin{aligned} &\text{SSD-model}: \mathbf{s} = \mathbf{m} \\ &\text{NCC-model}: \mathbf{s} = a\mathbf{m} + \mathbf{b} \end{aligned} \tag{9}$$

where $\mathbf{m}$ denotes the pixel value of the master image, $\mathbf{s}$ denotes the pixel value of the slave image, and a and $\mathbf{b}$ are the respective scaling and offset of the linear transformation.

4.2 Non-Parametric Model : MI

MI is non-parametric in the sense that it can capture arbitrary complex relations between datasets. In medical imaging, maximisation of MI has been demonstrated to be a general and reliable approach to register multimodal images [6],[8]. Unlike e.g. correlation based measures, MI assumes no functional relationship between the values of both images. Rather, it

measures the statistical dependency between corresponding image values.

$$\begin{aligned} \mathrm{MI}(X;Y) &= \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y) \\ &= \mathrm{div}_{\mathrm{KL}}\left(p(\mathbf{x})\,p(\mathbf{y}), p(\mathbf{x},\mathbf{y}) \right) \end{aligned} \tag{10}$$

where H(•) denotes an entropy measure (in our case Shannon entropy) and $\mathrm{div}_{\mathrm{KL}}(\bullet,\bullet)$ is the Kullback-Leibler divergence (distance) measure. The latter provides an intuitive explanation for MI. If two random variables X and Y are statistically independent, we can write $p(\mathbf{x},\mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$ and consequently their MI is zero. The divergence therefore measures how far we are from this situation. The larger the MI-value, the more dependent both random variables are.

As MI is a statistical measure, relying on an (unknown) joint density of the data, it has to be approximated by an estimator. Different solutions can be used. We refer to the literature for more details. In [9] an elaborate description and a few comparative experiments for different estimators are carried out.

5. EXPERIMENTS

The quality of any co-registration is difficult to evaluate. In this paper we provide three types of experiments. Firstly, a large-scale ground-truth experiment is conducted with the sole purpose of evaluating the algorithm's robustness against noise and the magnitude of the geometric distortion. Secondly, a number of experiments was carried out on a real database. Co-registration of misaligned raster and rendered vector layers was evaluated using a separate set of ground control points. Finally, a series of challenging real examples was processed and evaluated qualitatively using checkerboard images as a visual aid.

5.1 Ground-truth experiments

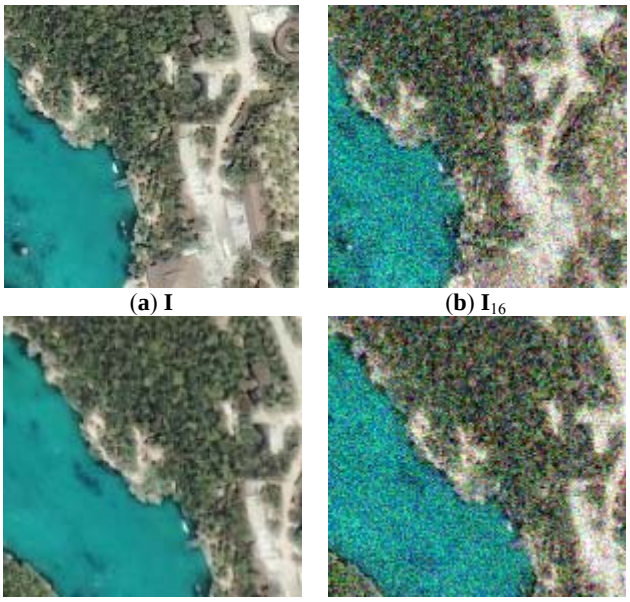
A set of 24 arbitrarily chosen RGB remote sensing images of size 800x600 was subjected to a series of tests. Each image was both geometrically and photometrically distorted; for a given transform type T a set of images $\mathbf{I}_{ij}$ is generated from $\mathbf{I}$ where i=1,2,…,6 indicates the level of geometric distortion and j=1,2,…,6 indicates the photometric noise level. The geometric distortion was obtained by means of a radially symmetric Gaussian distortion in 2D:

$$d_i : \mathbf{x} \mapsto d_i(\mathbf{x}) = \mathbf{x} + \mathbf{n}\,, \quad \mathbf{n} \sim \mathcal{N}\left(\mathbf{0}, (5i)^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \tag{11}$$

Applying d_i to the four corners of the image grid yields a set of pairs $(\mathbf{x}_k, d_i(\mathbf{x}_k))$ that is fed into a least square fitting algorithm for the requested transform type T. The outcome (best fit) yields the desired ground truth transform parameters that are used to resample the image.

The photometric distortion was applied on each colour band separately (after geometric distortion and resampling) by using a Gaussian noise distribution with $\sigma = 10j$. Figure 2 depicts details (patches) of four extreme samples generated from an example image.

The accuracy of the convergence was measured by computing the RMSE between the ground-truth transformation and the result of the optimisation using all pixels from the master image grid.



(**a**) **I** (**b**) $\mathbf{I}_{16}$ (**c**) $\mathbf{I}_{61}$ (**d**) $\mathbf{I}_{66}$

Figure 2: Details of generated raster images for ground-truth experiments. (a) shows a detail of the original raster image, (b), (c) and (d) show the corresponding details (patches) of the distorted raster images.

5.1.1 Performance of criteria and transform types

A first series of experiments examines the convergence quality for different combinations of criteria and transforms. The results are depicted in Figure 3 (SSD), Figure 4 (NCC) and Figure 5 (MI).

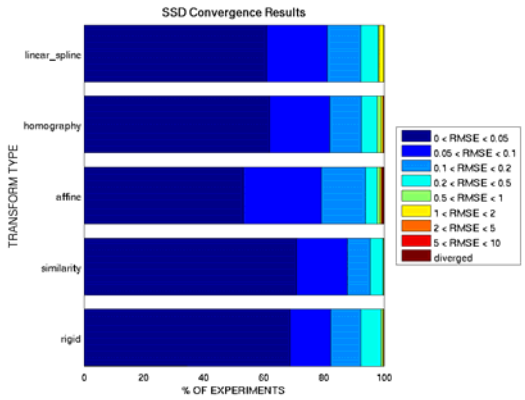


Figure 3: Convergence results for different transform types using the SSD criterion for all 864 experiments.

The overall accuracy of the registration is very good and in the majority of experiments, deep sub-pixel accuracy (RMSE < 0.1) is achieved. SSD and NCC perform equally well, while MI results are significantly worse. For all the experiments a subsampling approach using 10000 pixels was used which, with the exception of MI, did not affect the RMSE that was obtained with dense registration. As Figure 5(**b**) illustrates, increasing the number of samples (dense equals the use all pixels) does improve the RMSE for MI. This points out that the MI measure requires a large number of samples for providing a good approximation. Adjusting the kernel width of the histogram estimator could improve the results and will be tested in future experiments.

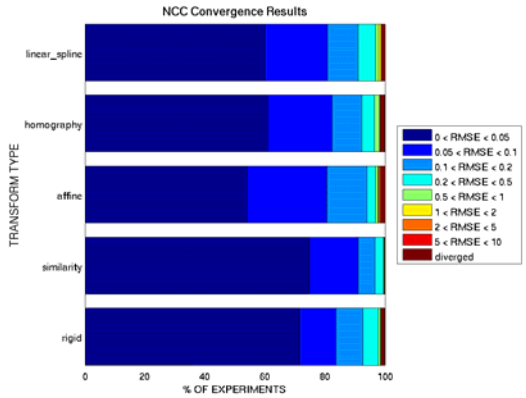
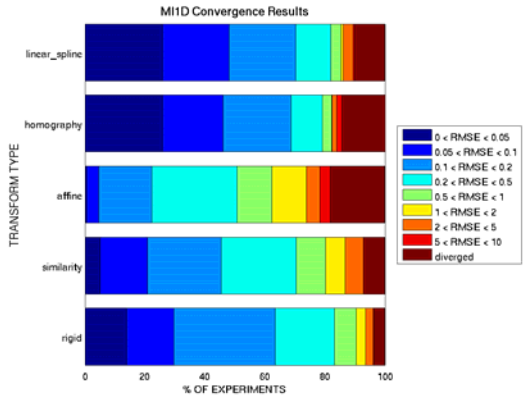
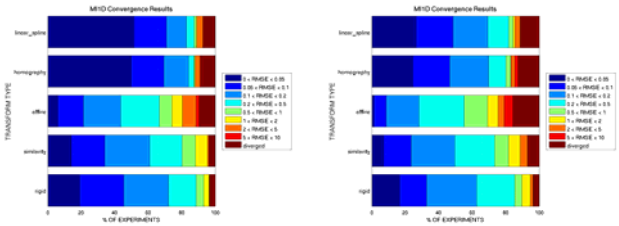


Figure 4: Convergence results for different transform types using the NCC criterion for 864 experiments.



(**a**) subsampled, 32 bins



(**b**) dense, 32 bins (**c**) subsampled, 64 bins

Figure 5: Convergence results for different transform types using the MI criterion for 864 experiments. (**b**) increasing the number of samples improves the result, while (**c**) increasing the number of bins barely makes a difference.

When comparing the results for different transformations, it seems that the linear spline and the homography do particularly well, even with MI. This seems contradictory as the number of parameters for both transformations (8) is higher than for e.g. an affine transformation (6). An explanation can be found in the "nature" of the parameters. Both the linear spline and the homography are parameterised in terms of displacements of the four corners of the image. Changing either of the parameters has an effect of equal magnitude on the overall geometric distortion. For an affine transformation, scaling and translation behave quite differently in this respect. Eventhough the optimisation routine tries to balance the influence of each parameter proportional to the magnitude of their effect, the former type of parameterisation clearly remains advantageous.

5.1.2 Influence of the magnitude of the geometric distortion.

For this experiment, the convergence results were grouped per geometric distortion level. (i=1…6). Figure 6 shows that the amount of geometric distortion does not affect the quality of the convergence significantly for SSD (NCC and MI yield similar graphs). If the distortion grows too large, the algorithm simply won't converge (notice the increase in divergences for noise level 6).

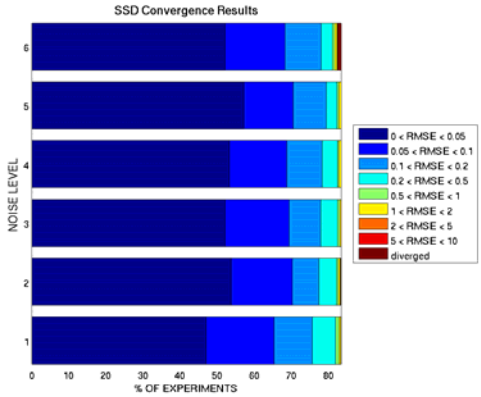


Figure 6: Convergence results for different geometric distortion levels using the SSD criterion for 720 experiments.

5.1.3 Influence of spectral noise

For this experiment the convergence results were grouped per spectral noise level (j=1…6). The results in Figure 7 reveal that image noise has a significant effect on the quality of the convergence of SSD (NCC and MI yield similar graphs). This is to be expected as the convergence process of an image-based registration is governed by the similarity of spectral content of the pixels. If this information becomes more and more corrupted, the matching performance will evidently deteriorate. Nevertheless, even at the highest evaluated spectral noise level, in the large majority of all experiments a sub-pixel convergence (RMSE < 0.5) was achieved.'

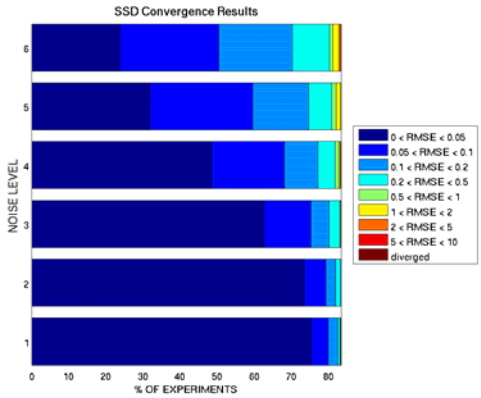


Figure 7: Convergence results for different spectral noise levels using the SSD criterion for 720 experiments (all transformations).

5.1.4 Algorithmic complexity

In general, area-based method are slow due to the fact that they take into account information from all pixels. By using the

subsampling strategy, the computational complexity is significantly reduced as can be seen in Figure 8 for the SSD-criterion. The same graphs can be reproduced for NCC and MI, albeit that NCC is typically twice as slow as SSD and MI is four times as slow as SSD. All experiments were carried out on an AMD Athlon 64 X2 Dual Core 3800+, 1GHz PC.

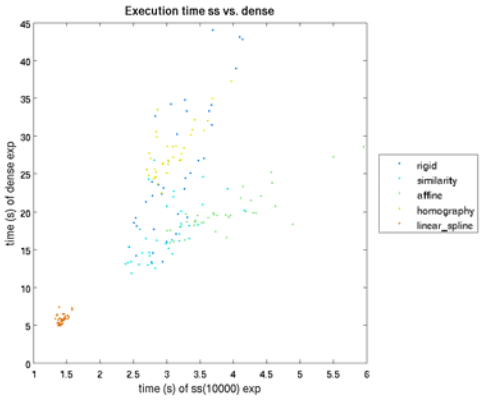


Figure 8: Comparison of execution times for SSD experiments using the dense and subsampling approach.

Another interesting plot is depicted in Figure 9. Here, the complexity of the MI criterion is analysed in terms of the number of bins. The increasing number of bins considerably slows down the algorithm due to the necessary histogram filling and smoothing operations.

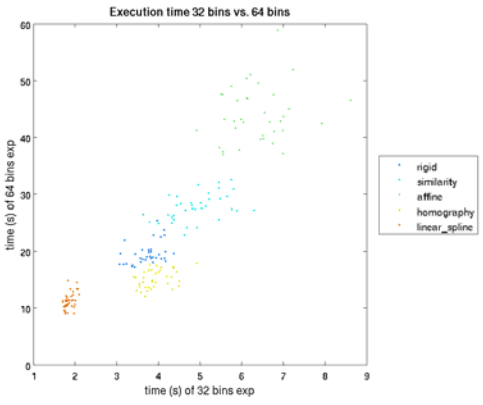


Figure 9: Comparison of optimisation time for MI experiments carried out with 32 bins and 64 bins.

5.2 Real examples

The ground-truth examples have revealed interesting results on the behaviour of the algorithm in terms of image noise, criteria and transformation families. Whether or not the tool is able to tackle real co-registration problems can only be verified by using relevant datasets. In this way, the usage of different criteria and transformations can be evaluated.

5.2.1 Google Earth experiment

For this experiment, 11 pairs of screenshots from Google Earth were used. The orthophoto (raster layer) was used as the master image, the rendered street network was used as the slave image. As Figure 10 shows, at some locations, the vector layer does not match the raster layer very accurately. For each of the pairs, a set of control points was manually selected and matched between the master and slave image and used as ground truth. The RMSE of the initial situation is compared with the RMSE after co-registration using MI in Table 1.

Figure 10: Example of misalignment of raster and vector layer.

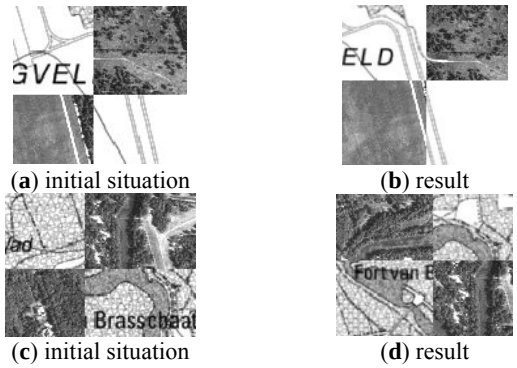
region (country)	before registration	after **affine** registration	after **homography** registration
countryside (UK)	9.7871	7.0572	6.1145
countryside (FR)	8.0000	4.2557	1.9963
sarrebourgh (FR)	6.5879	4.7661	5.8124
Berlin (D)	6.9887	7.2402	7.1242
Munchen (D)	6.4292	5.3857	6.8045
Bologna (I)	8.3075	7.9933	7.7724
countryside (P)	6.9642	37.5665	20.6179
unknown (ES)	12.5300	3.1402	4.4292
Zaragoza (ES)	10.1489	23.5197	10.8079
countryside (USA)	6.4031	4.8768	5.3966
Kansas City (USA)	7.0285	3.4956	2.6116

Table 1: RMSE of Google Earth data before and after co-registration using MI for different regions.

Most of the results show an improved co-registration, although for a few examples the RMSE became worse. After visual inspection of the results, the incorrect match was found to be a local optimum, due to either inconsistencies in the road network or the absence of spectral contrast in some regions (e.g. Portugal countryside).

5.2.2 Various experiments

The experiment section is concluded with some co-registration examples obtained combining various sensors and data types.



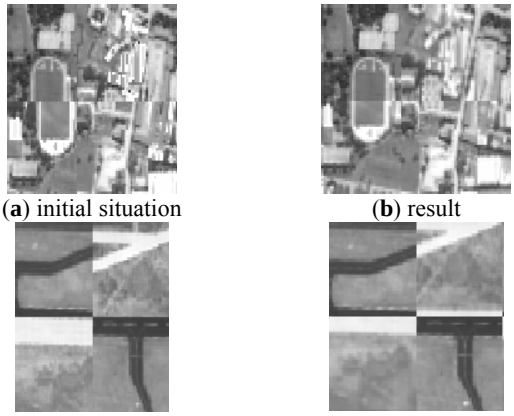
(**a**) initial situation (**b**) result

(**c**) initial situation (**d**) result

Figure 11: Details of an affine co-registration of an aerial image (grey) with a topographic map using MI. (a)

and (c) show the initial situation, (b) and (d) the result.

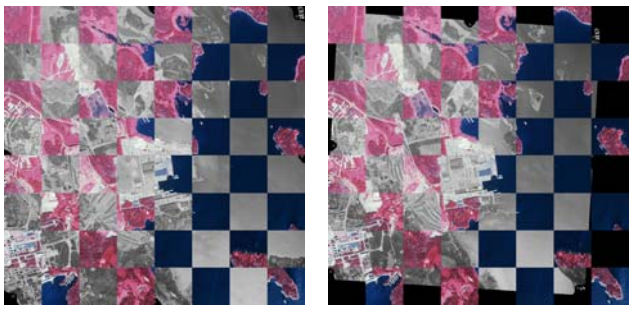
The results are presented as checkerboards, where the white squares show the contents of the master and the black squares the contents of the slave. Figure 11 depicts details of a checkerboard of a co-registration of an aerial image and a topographic map under affine transformation using MI. The example shown in Figure 12 is obtained by co-registering a visual and an IR image using MI and optical flow. Finally, Figure 13 illustrates the rigid co-registration of a visual and a NIR image using MI.



(**a**) initial situation (**b**) result

(**c**) initial situation (**d**) result

Figure 12: Details of optical flow co-registration of an aerial image (grey) with an IR image using MI. (a) and (c) show the initial situation, (b) and (d) the results



(**a**) initial situation (**b**) result

Figure 13: Rigid co-registration of an aerial image (grey) and a 3-band NIR image using MI. (a) shows the initial situation, (b) the result.

6. SUMMARY AND DISCUSSION

This paper outlined a framework for area-based co-registration. Considerable effort was spent on implementing the different building blocks in a generic, coherent and modular fashion. A subsampling heuristic provided an efficient strategy for speeding up the co-registration process. We summarise concisely the results and findings of the paper:

- Area-based registration provides highly accurate results, but remains slow compared to point-based strategies. The framework could benefit from parallelising the computations and providing efficient support for large images (out of core processing). Also, the effect of biased subsampling, in order to reduce the number of samples, will be investigated.
- In terms of optimisation of geometric transformation parameters, "grid-based" transform types are favoured, because of their superior convergence behaviour. Future extensions of the transform library should therefore also incorporate local transform types on regular/irregular grids, because of their ability to adapt to local deformations.
- The MI criterion shows a worse convergence behaviour compared to SSD and NCC. Further investigation is required to determine whether better histogram estimators (e.g. different kernel width) can be deployed to obtain better results.

In the introduction we already pinpointed a number of applications that rely an accurate co-registration of the underlying data. Change detection is an important example. The necessary tools for change detection are already provided by the proposed registration framework. A pixel $\mathbf{x}$ has undergone change when its spectral value $\mathbf{m} = I_1(\mathbf{x})$ in the master image is not compliant with its counterpart $\mathbf{s} = I_2(T(\mathbf{x}))$ in the slave image. The dissimilarity or change, which we will denote $D(\mathbf{m},\mathbf{s})$, depends on the criterion. For SSD, this can be expressed in terms of Euclidean distance $D(\mathbf{m},\mathbf{s}) = \|\mathbf{m}-\mathbf{s}\|$ between the spectral values. For NCC, this is the Euclidean distance between the spectral value of the master and the affine transformed value of the slave $D(\mathbf{m},\mathbf{s}) = \|\mathbf{m}-a\mathbf{s}-\mathbf{b}\|$, where a and $\mathbf{b}$ are regression estimates minimizing $D(\mathbf{m},\mathbf{s})$ for all given matches. For MI, the desired change is derived from the joint probability density function that is approximated in the estimation of MI.

The joint pdf expresses the co-occurrence of spectral values $\mathbf{m}$ in the master and $\mathbf{s}$ in the slave image: $p(\mathbf{m},\mathbf{s})$. The amount of change is then inversely proportional to the likelihood that $\mathbf{s}$ is observed for pixel $T(\mathbf{x})$ given that $\mathbf{m}$ was observed for $\mathbf{x}$. This quantity is called the conditional likelihood $p(\mathbf{s}|\mathbf{m})$ and is computed from the joint pdf as $p(\mathbf{s}|\mathbf{m}) = p(\mathbf{s},\mathbf{m})/p(\mathbf{m})$. With a final refinement we can define the "probabilistic change measure" associated with the MI criterion as

$$D(\mathbf{w},\mathbf{v}) = \max\left\{1 - p(\mathbf{w}\mid\mathbf{v}), 1 - p(\mathbf{v}\mid\mathbf{w})\right\} \qquad (12)$$

Notice that this dissimilarity measure takes both conditional likelihoods as arguments. This is neccessary because a certain master-slave pixel value combination $(\mathbf{m},\mathbf{s})$ is only unlikely iff both conditional probabilities are low. If only one conditional probability is low, say $p(\mathbf{s}|\mathbf{m})$ is low but $p(\mathbf{m}|\mathbf{s})$ is high, this merely signals that there is a relatively low number of slave pixels taking on the pixel value $\mathbf{s}$. In this particular example, all these slave pixels are in correspondence with similarly valued master pixels ($p(\mathbf{m}|\mathbf{s})$ is high), hence this is not an improbable or suspicious match. Similar arguments hold for the case in which $p(\mathbf{m}|\mathbf{s})$ is low but $p(\mathbf{s}|\mathbf{m})$ is high. The proposed dissimilarity measure has the added advantage of symmetry.

Figure 14 shows an illustrative example of change detection on satellite imagery that was taken at two different instances in time. Clearly, the spectral content is different, yet similar structures are perceived. A co-registration was carried out with a homography using MI. The change image shown in Figure 15 is constructed using equation (12) as follows:

$$I_{\text{change}}(\mathbf{x}) = D\left(I_1(\mathbf{x}), I_2(T(\mathbf{x}))\right) \qquad (13)$$

The change image has one band and is displayed using a false colour map, where dark blue means low values (no change) and red indicates high values (change). Notice that, despite the multi-modal image contents, the probabilistic change map highlights the drying up of the lake (bright spots at the top), the

appearance of a few new buildings (bright spots in the middle) and the appearance of a new channel (light blue in the left bottom). These preliminary results are promising, and the full exploration of this type of probabilistic change detection is part of future work.

(**a**) old image (**b**) new image

Figure 14: Example of two SPOT images captured at different instances in time.

7. REFERENCES

[1] Bentoutou, Y., Taleb, N., Kpalma, K., Ronsin, J., 2005. *An automatic Image Registration for Applications in Remote Sensing*, IEEE Trans. On Geoscience and Remote Sensing, 43(9), pp. 2127-2137.

[2] Li, H., Manjunath, B. S., Mitra, S. K., 1995. *A contour-based approach to multisensor image registration*. IEEE Transactions On Image Processing, 4(3), pp. 320-334.

[3] Le Moigne, J., Campbell, W.J., Cromp, F.J., 2002. An automated parallel image registration technique based on the correlation of wavelet features. IEEE Transacations on Geoscience and remote sensing, 40(8), pp. 1849-1864.

[4] Brown, L.G. 1992. *A survey of Image Registration Techniques*. ACM Comput. Surv. 24(4), pp. 325-376.

[5] Zitova, B., Flusser, J. 2003. *Image Registration Methods: A survey*. Image Vis. Comput., 21(11), pp. 977-1000.

[6] Zagorchev, L., Goshtasby, A., 2006. *A comparative study of transformation functions for non-rigid image registration*, IEEE Trans. On Image Processing, 15(3), pp. 529-538.

[7] Maes, F. Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality Image Registration by maximization of Mutual Information. *IEEE Trans. Med. Imag.,* 16(2), pp. 187-198.

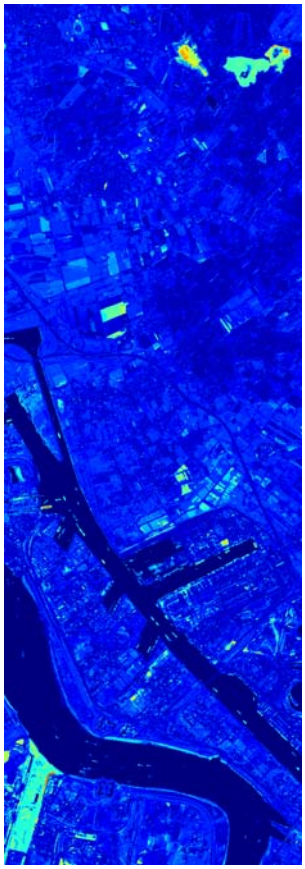


Figure 15: Change image after homography co-registration using MI.

[8] Wells, W. M. III, Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of Mutual Information. *IEEE Trans. Med. Imag. Anal.*, 1, pp. 35-51.

[9] Fransens, R., Strecha, C., Van Gool, L., 2004. *Multimodal and multiband image registration using Mutual Information*. ESA-EUSC 2004: Theory and Applications of Knowledge driven Image Information Mining, with focus on Earth Observation; EUSC, Madrid (Spain).

[10] Johnson, K., Cole-Rhodes, A., Zavorin, I., Le Moigne, J., 2001. *Mutual Information as a similarity measure for remote sensing image registration*. Proc. SPIE Aerosense, Geo-Spatial Image and Data Exploitation II, 4383, pp. 51-61.

Acknowledgements

We would like to acknowledge AERODATA, GIM and GOOGLE for providing the necessary test material for evaluating the registration tool. The SPOT images are under license by GIM and used by GIM in the project for technical

evaluation of the IG-ART product in collaboration with IncGEO.