A GENERIC APPROACH TO IMAGE CO-REGISTRATION.

R. Fransens1, C. Strecha1, G. Caenen1,2, and L. Van Gool1,3

1KUL, Dept. of Electrical Engineering, Kasteelpark Arenberg 10, B-3001 Heverlee
2IncGEO, Research Campus Hasselt, Kempische Steenweg 297/2, B-3500 Hasselt
3ETH, Dept. of Information Technology and Electrical Engineering, Sternwartstrasse 7, CH-9092 Zurich

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 few challenging real examples.

Key words: Image co-registration, mutual information, change detection.

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 then 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 nowadays accompanied by a (rough) georeference. For a more elaborate overview of past and current registration techniques we refer to surveys in the literature: [5, 6].

1.2. Outline of the paper.

This paper describes a framework that enables 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 image co-registration as an optimisation problem. Additionally a modular implementation is discussed using principles from object oriented programming. Next, section 3 and 4 elaborate respectively on the geometric transformations and similarity measures that are supported by the framework. Section 5 deals with optimisation strategies and speed-ups. In section 6 a series of experiments is carried out on ground-truth and real data. One particular application is highlighted in section 7. An extension of the framework towards change detection is explained using an illustrative example. Finally, section 8 concludes the paper with a summary of the registration results and a short discussion on how the current framework can be further improved and extended.
2. MATHEMATICAL FORMULATION.

For the sake of clarity, we start off with introducing some definitions and conventions that will be used throughout the paper. Next, these basic building blocks are combined to state the co-registration problem.

2.1. Image functions and raster images.

An image is assumed to be a differentiable function $I$, that maps pixels $x = [x, y]^t$ from a domain $D \subset \mathbb{R}^2$ to an $n$-dimensional set of spectral values $[I_1(x) \ldots I_n(x)]^t$.

A raster image $I = \{I_{ijk}\}$ is an $M \times N \times n$ array containing a discrete set of samples drawn from $I$. The samples are taken from a lattice $\{x_{ij}\}$ by means of a sampling function $S$. Typically, a canonical lattice $\{x_{ij} = [j \ q]^t, i = 1 \ldots M, j = 1 \ldots N\}$ is used. Hence, $I_{ijk} = S(x_{ij}; I)$.

In practice, the image $I$ is not known. We can therefore try to reconstruct it from a known sampling $I$ by means of a reconstruction function $R(x; I)$.

In this paper we will considerably simplify the mathematics by assuming simple reconstruction and sampling functions. Although these are non-optimal in some respects, they yield accurate results in practice. For a more elaborate discussion on sampling functions and their corresponding optimal reconstruction functions we refer to the literature.

$$S(x; I) = I(x)$$

$$[R(x; I)]_k = P_{00}I_{[y][x]k} + P_{11}I_{[y][x]k} + P_{10}I_{[y][x]k} + P_{01}I_{[y][x]k}$$

with $\lfloor \cdot \rfloor$ the floor-function and $\lceil \cdot \rceil$ the ceil-function. $P_{ij} = p_{ij}$ are the bilinear interpolation weights computed from

$$p_0 = x - \lfloor x \rfloor = 1 - p_1$$

$$q_0 = y - \lfloor y \rfloor = 1 - q_1$$

Figure 1 illustrates these functions on a 1D-signal.

2.2. Geometric transformations.

A geometric transformation $T$ maps points $x$ in the plane $(\mathbb{R}^2)$ to other points $T(x)$ in the plane. Parametric transformations $T(\cdot; p)$ can be described completely by means of a parameter vector $p$.

**Example 1** A rigid transformation has three parameters: the two components of the translation vector and a rotation angle. Hence $p = [\theta \ t_x \ t_y]^t$.

$$T(x; p) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

2.3. Similarity measures.

Similarity of spectral signatures is expressed by a real-valued similarity measure or criterion function $C$, that takes as an input two arrays of the same size.

**Example 2** The Sum of Squared Differences (SSD) criterion computes the Euclidian distance between the arrays $X$ and $Y$.

$$SSD(X, Y) = \frac{1}{MN} \sum_{i,j,k} (X_{ijk} - Y_{ijk})^2$$

where $M \times N \times n$ are the dimensions of both arrays.

2.4. Image co-registration.

The co-registration of two raster images $I$ and $J$ is obtained as follows. One of the raster images, say $I$ is considered as the reference or the master image. The slave image $J$ is deformed as to fit the master image as good as possible. The deformation is constrained by a transformation $T(\cdot; p)$ and the fit is quantified by a criterion $C(\cdot, \cdot)$. Consequently, we can formulate the co-registration as an optimisation problem.

$$\hat{p} = \arg \max_p f(p) = \arg \max_p C(I, J')$$

where $J'_{ijk} = R(T(x_{ij}; p); J)$ represents a resampled version of $J$ according to the transformation $T(\cdot; p)$. Figure 2 depicts the construction of the objective function. The evaluation of the objective function $f$ for a given parameter vector $p$ is computed in three steps.

1. compute the transformed grid $T(x_{ij}; p)$,
2. create $J'$ by resampling $J$,
3. compute $C(I, J')$. 

---

Figure 1. Sampling and reconstruction functions. Left: the original signal (blue line) is sampled. Right: bilinear reconstruction of the signal (red line) based on the samples.

Figure 2 depicts the construction of the objective function.
Deterministic optimisation strategies usually require the gradient of the objective function for attaining a (local) optimum. The components of the gradient vector of (5) is obtained as:

$$\frac{\partial f}{\partial p_i}(p) = \sum_{i,j,k} \frac{\partial C}{\partial Y_{ijk}}(I, J') \cdot \langle \frac{\partial R_k}{\partial x} (T(x_{ij}; p); J), \frac{\partial T}{\partial p_i}(x_{ij}; p) \rangle \quad (6)$$

where we applied the chain rule to resolve the dependencies. $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^2$, and $R_k$ reconstructs only the $k$-th band of the slave image. It is certainly worthwhile taking a closer look at (6) 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 gradient function of the reconstructed 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 $J$ 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 (6).

### 2.5. Object-oriented implementation

Object oriented programming allows for generic and modular implementation of the proposed co-registration framework. 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 3 summarises this scheme.

### 3. GEOMETRIC TRANSFORMATIONS.

Typical global parametric transformations display a nested (sub)group structure: translations, rigid motions, similarity transformations, affine transformations and planar homographies. Transformations such as linear splines, polynomial and rational transformations gradually allow local deformations with increasing number of parameters (resp. spline grid size or polynomial degree). The thin-plate spline transform and the weighted mean transform are governed by the displacement of a predefined set of control points. Finally, the most degrees of freedom are obtained using a free-form transformation that defines a displacement for every point $x$ in the plane by means of a flow field $u(x)$. To avoid ill-conditioned parameter estimations, smoothness constraints are imposed on the flow field.
From an application point of view, the relevant question is: What kind of transformation can account for the possible geometric distortions caused by sensor types, sensor positioning and the topography of the scene. This question forms the basis of the practical co-registration problem. Imagery of planar scenes, acquired with perspective (pinhole) cameras can be co-registered using a homography after removing the artefacts of lens distortion (usually a polynomial transformation). For non-planar scenes we immediately arrive at free-form transformations. In cities the depth discontinuities of the buildings violate the smoothness assumption of the flow-field and one needs to resolve to anisotropic flow-fields which are typically dealt with in depth-from-stereo settings.

However, the proposed framework allows an elegant integration of topographic information in the form of a DSM into the co-registration. For the simple case where we dispose of an orthorectified image \( I \) that has an associated (co-registered) DEM \( Z(x) \) we can define the following transformation:

\[
T(x; p, Z) = P_p \left( \begin{bmatrix} x \\ Z(x) \end{bmatrix} \right)
\]

with \( P_p \) the projection matrix and \( p \) the calibration parameters (possibly both internal an external) of the camera. Co-registration here boils down to calibrating the second image.

### 4. SIMILARITY MEASURES.

Similarity measures quantify the match between the spectral content of the master and slave image. Depending on the nature of both images (sensor, camera parameters, \ldots) the measure should account for the photometric transformation at hand. In this paper, we discuss three criteria types, with increasing complexity: Sum of Squared Differences (SSD), 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, \ldots) 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:

\[
\begin{align*}
\text{SSD} & : s = m \\
\text{NCC} & : s = am + b
\end{align*}
\]

where \( m \) denotes the pixel value of the master image, \( s \) denotes the pixel value of the slave image, and \( a \) and \( b \) 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 [8, 7]. 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.

\[
\begin{align*}
\text{MI}(X, Y) &= H(X) + H(Y) - H(X, Y) \\
&= \text{div}_{\text{KL}}(p(x)p(y), p(x, y))
\end{align*}
\]

where \( H(\cdot) \) denotes an entropy measure (in our case Shannon entropy) and \( \text{div}_{\text{KL}}(\cdot, \cdot) \) 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 \( p(x, y) = p(x)p(y) \) 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. OPTIMISATION.

#### 5.1. Multi-resolution approach.

Non-convex optimisation often suffers from local minima, especially when the starting point 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 its way down the pyramid as follows. After optimisation, the solution \( p_i \) at pyramid level \( i \) is “upscaled” to level \( i + 1 \) \((p_{i+1})\) 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.
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 $I$ can be reduced to a percentage of all pixels accompanied by their locations $x$. 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).

5.3. 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.

6. 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 photometric and geometric distortions. 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.


A set of 24 arbitrarily chosen RGB remote sensing images of size $800 \times 600$ was subjected to a series of tests. Each image was both geometrically and photometrically distorted; for a given transform type $T$ a set of raster images $I_{ij}$ is generated from $I$ where $i = 1, 2, \ldots$ indicates the level of geometric distortion and $j = 1, 2, \ldots$ indicates the photometric distortion level. The geometric distortion was obtained by means of a radially symmetric Gaussian distortion in 2D:

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

Applying $d_i$ to the four corners of the image grid yields a set of pairs $(x_k, d_i(x_k))$ 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 $\sigma = 10$. Figure 4 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.

6.1.1. 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 5, Figure 6 and Figure 7. 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 7(b) illustrates, increasing the number of samples (dense equals the use of 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.

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. Even though 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.

6.1.2. Influence of the magnitude of the geometric distortion.

For this experiment, the convergence results were grouped per geometric distortion level. \( i = 1 \ldots 6 \). Figure 8 shows that the amount of geometric distortion does not affect the quality of the convergence significantly for NCC (SSD 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).

6.1.3. Influence of photometric distortion.

For this experiment the convergence results were grouped per spectral noise level \( j = 1 \ldots 6 \). The results in Figure 9 reveal that image noise (here referred to as photometric distortion) has a significant effect on the quality of the convergence of NCC (SSD 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 distortion level, in the
large majority of all experiments a sub-pixel convergence (RMSE < 0.5) was achieved.

6.1.4. 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 10 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.

Another interesting plot is depicted in Figure 11. 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.

6.2. 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.

6.2.1. 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. 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. 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).

6.3. Miscellaneous Experiments.

To conclude our experiments, a few details from a co-registration example of a topographic map and an aerial photo are depicted in Figure 12. 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.

7. CHANGE DETECTION.

In the introduction we already pinpointed a number of applications that rely on 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 \( x \) has undergone change when its spectral value \( m = I(x) \) in the master image is not compliant with its counterpart \( s = J(T(x)) \) in the slave image. The dissimilarity or change, which we will denote \( D(m, s) \), depends on the criterion at hand. For SSD, this can be expressed in terms of Euclidean distance

\[
D(m, s) = \| m - s \| 
\]

For NCC, this is the Euclidean distance between the spectral value of the master and the affine transformed value of the slave

\[
D(m, s) = \| m - as - b \| 
\]

where \( a \) and \( b \) are regression estimates minimizing \( D(m, s) \) 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 \( m \) in the master and \( s \) in the slave image: \( p(m, s) \). The amount of change is then inversely proportional to the likelihood that \( s \) is observed for pixel \( T(x) \) given that \( m \) was observed for \( x \). This quantity is called the conditional likelihood \( p(s|m) \) and is computed from the joint pdf as \( p(s|m) = p(s, m)/p(m) \). With a final refinement we can define the probabilistic change measure associated with the MI criterion as

\[
D(m, s) = \max \{ 1 - p(s|m), 1 - p(m|s) \} 
\]

Notice that this dissimilarity measure takes both conditional likelihoods as arguments. This is necessary because a certain master-slave pixel value combination \( (m, s) \) is only unlikely if both conditional probabilities are low. If only one conditional probability is low, say \( p(s|m) \) is low but \( p(m|s) \) is high, this merely signals that there is a relatively low number of slave pixels taking on the pixel value \( s \). In this particular example, all these slave pixels are in correspondence with similarly valued master pixels \( p(m|s) \) is high), hence this is not an improbable or suspicious match. Similar arguments hold for the case in which \( p(m|s) \) is low but \( p(s|m) \) is high. The proposed dissimilarity measure has the added advantage of symmetry. Figure 13 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 14 is constructed using equation as follows:

\[
C(x) = D(I(x), J(T(x))) 
\]

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.

8. SUMMARY AND CONCLUSIONS.

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.
ACKNOWLEDGMENTS

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.

REFERENCES


