

ADAPTIVE REGISTRATION OF REMOTE SENSING IMAGES USING SUPERVISED LEARNING

This paper describes a novel approach for registration of time series of remote sensing images, using supervised learning and a region-based strategy to adapt the registration to image characteristics.

Line Eikvil* (Line.Eikvil@nr.no)
Marit Holden* (Marit.Holden@nr.no)
Ragnar Bang Huseby* (Ragnar.Bang.Huseby@nr.no)

*) Norwegian Computing Center, P.O. Box 114, Blindern, 0314 Oslo, Norway

ABSTRACT/RESUME

This paper describes a system for co-registration of time series which uses a learning-based strategy. During a training phase the system learns to recognize regions in an image suited for registration. It also learns the relationship between image characteristics and registration performance for a set of different registration algorithms. This enables intelligent selection of an appropriate registration algorithm for each region in the image, while regions unsuited for registration can be discarded. The approach is intended for co-registration of sequences of images acquired from identical or similar earth observation sensors. It has been tested for such sequences from different types of sensors, both optical and radar, with varying resolution. For images with moderate differences in content the registration accuracy is in general good with an RMS error of one pixel or less.

ADAPTIVE REGISTRATION OF REMOTE SENSING IMAGES USING SUPERVISED LEARNING

1. INTRODUCTION

The study of time series of satellite images is an important task in many remote sensing applications, for instance for observing different environmental phenomena. In such applications a model-based geo-referencing is first performed based on satellite parameters. Then a co-registration of the images in the time series is used to improve the alignment of the images and refine the accuracy. A combination of manual and automatic registration techniques is typically used for this co-registration, but fully automatic techniques do exist.

The image registration process tries to determine the most accurate match between the images, and automatic techniques typically combine similarity metrics and matching strategies to achieve this. There exists a range of different algorithms, and selection of the appropriate combination depends on the application and the image specifics. Hence, a single registration technique will generally not be sufficient when handling a range of images. For a user that needs to work on different types of time series, it would therefore be useful to have a more general tool for image registration. In this paper we present a novel registration approach, which aims at offering the user such a tool. The idea is to have a system with a library of different registration algorithms, and to provide a tool on top of this which makes the system able to intelligently choose, at run-time, an appropriate algorithm based on image characteristics.

Others have presented systems for image registration that offer the user a selection of different methods to choose from. Fedorov *et al.* (2002) have designed a system consisting of three different algorithms for control point extraction which can easily be extended to include more methods. Le Moigne *et al.* (2004) present a modular image registration framework, offering a selection of algorithms. A registration algorithm is here defined as a combination of features, similarity measure and matching strategy, where a user may choose components suitable for the problem at hand. Hence, the user is responsible for selecting the right combination, and this combination will be used for the entire image. This requires that there exists an algorithm that will work for the entire image. The user also needs to know the algorithms well, and be able to select the algorithm that is best suited for each problem. Alternatively, the user can test and compare several algorithms.

Approaches, where the system itself is able to automatically select the best registration algorithm for a problem, have been mentioned as potentially useful, for instance by Rignot *et al.* (1991) and Fonseca and Costa (1997). However, no one has presented methods for automatic selection of the best algorithm, and systems incorporating such solutions have yet to appear.

In the approach that we propose, automatic selection of the best algorithms is achieved by using supervised learning, where the system learns the correspondence between image characteristics and algorithm performance. In addition, we use a region-based strategy to make the approach locally adaptive. This allows for use of different registration algorithms for different regions and permits regions not suited for registration to be discarded. The approach is an extension of a method introduced in a previous workshop paper (Eikvil *et al.*, 2005).

2. METHODS

The approach that we suggest works as follows. In a training phase, the system learns the correspondence between image features and performance for a set of registration algorithms. At run-time, this enables the system to predict the performance for each algorithm from the image features. By applying this strategy to image regions rather than the whole image, the approach is made locally adaptive. Features are extracted from regions, and the performance of the available

algorithms is predicted for each region. For this prediction we have chosen to use a neural net. Based on the performance prediction, regions and algorithms are selected. For each selected region, a local registration can then be performed with the appropriate algorithm, and from the resulting set of local transforms, a smooth global transform can be estimated. Figure 1 illustrates the steps involved in our approach. The details of each step will be treated in Sections 2.1 to 2.4.

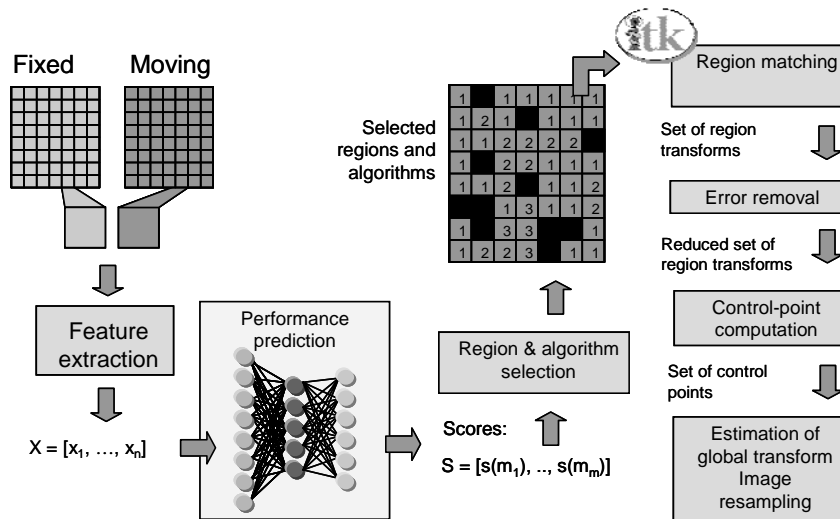


Figure 1: The first few steps of the registration approach consist of feature extraction, performance prediction and region and algorithm selection, which results in an image where some regions are discarded (marked as black cells), while different registration algorithms may be selected for each of the remaining regions (illustrated as cells marked 1, 2, 3). This is followed by a local co-registration of image regions, removal of potential erroneous region transforms and a final estimation of a global transform.

2.1. Region-based strategy

We propose to use an approach where the registration algorithm is selected automatically based on image characteristics. In addition, as the characteristics of a remote sensing image may vary across the scene, it can be useful to permit the use of different algorithms for different regions. At the same time, it could be desirable to be able to identify and discard regions that are not suited for registration. Hence, we have chosen to use a region-based strategy in our approach. As corresponding regions from the images to be registered need to be comparable in shape and size, we have chosen a simple and robust strategy where the images are simply divided into smaller rectangular sub-images of equal size. Some discussion on the selection of region size can be found in Section 3.6.

In the rest of this paper we will, when considering pairs of images to be co-registered, use the notation *fixed image* and *moving image*, where the fixed image is the image defining the coordinate system, while the moving image is the one to be transformed to this coordinate system.

2.2. Feature extraction

We want to extract features from the image regions that convey image characteristics that are important for image registration. Below we discuss what would be important image characteristics in this context.

- Regions containing *characteristic patterns or details* will often provide more accurate matches. Hence, the features should say something about the information content of the image, e.g. whether the image is dominated by homogeneous or textured areas. For this purpose, texture and gradient features may be suitable.
- Information on whether there is a *correspondence* between the fixed and the moving image(s) will be important. Regions where there is no correspondence, or where the images have very different contents, should be discarded. For earth observation images, such situations can for instance be caused by clouds. Features based on differences between the fixed and moving image, may be useful to reveal these cases.
- Finally, the actual *relationship* between the image regions can also provide useful information. Many correlation-based methods for matching will, for instance, perform best when there is a linear relationship between the intensity of the images. Again, features conveying information about the relationship between images may be derived from differences between the images.

Based on the considerations above, we selected a set of features based on texture, image statistics and image differences. More details on these features will be given in the next section.

2.2.1. Features

To extract information about the *content and characteristic details* of the images we chose to use the following three types of features:

- *Texture features* that are computed from the Grey Level Co-occurrence Matrix (GLCM). These features were originally introduced by Haralick *et al.* (1973). They are based on second-order statistics, and are computed by finding repeated occurrences of grey-level configurations. The following texture features were used: *mean, variance, homogeneity, contrast, dissimilarity, entropy, ASM and correlation*. The features were extracted from windows corresponding to the image regions.
- *Registrability (RIG)*. This is a measure introduced by Chalermwat (1999) which is intended to represent a region's ability to provide unambiguous registration by measuring sensitivity to transformations. Regions that are sensitive to transformations are expected to be better suited for registration than regions that are not.
- *Gradient magnitude* computed by the Sobel operator is included.

To extract information about *correspondence and relationship* between the images, we use a set of features based on differences between image characteristics. These are computed as differences between the features computed from the fixed image and the moving image:

- *Differences in texture, registrability and gradient magnitude*. These are computed from the feature values derived with the methods described above.
- *Differences in statistical features*. The differences in values for a set of four normalized statistical features are computed for each pair of images (regions). These features are range, mean, variance and entropy.
- *Differences in zone means*. Each region is divided into a grid of nine zones, where the mean within each zone is calculated. The Euclidean distance and the variance of the differences between the zone means of corresponding regions are used as features. The features are normalized to reduce their dependence on the grey scale level of the image.

A total of 26 features are extracted from each region. The features are normalized such that the means and standard deviations are zero and one, respectively.

2.3. Selection of regions and registration algorithms

We need to establish a correspondence between extracted features and performance of the registration algorithms. This can be achieved in different ways:

- By establishing an a priori model for the correspondence between image characteristics and the expected performance of each algorithm.
- By using a training approach to determine the correspondence between image characteristics and algorithm performance.

In practice, these two are often combined. We have used a priori knowledge to choose the features, and a supervised training approach to establish the correspondence. The latter problem is viewed as a regression problem, where the objective is to use the extracted features to predict the performance for each registration algorithm. For this regression problem we have used a feed-forward neural network, which provides a flexible way to model complicated relations.

2.3.1. Training

Supervised learning is used in a separate training phase, where the system is trained on a large number of examples for which the distortion is known. These examples are obtained by first performing a careful manual co-registration of image pairs and then applying various known distortions to one of the images. The subsequent training is then performed by running all registration algorithms on all examples, and computing the algorithm performance in each case. Features based on image characteristics are also extracted for all the examples. Afterwards, the system is trained to learn the correspondence between image characteristics and algorithm performance.

For learning, and later prediction, we have chosen to use a neural network with one hidden layer. The number of input nodes corresponds to the number of image features, and the number of output nodes corresponds to the number of registration algorithms. For the training of this network, a measure of each algorithm's performance needs to be defined for the target values. We have decided to use a performance measure based on the distance from the true distortion. By using training examples for which the true distortion was known, the performance was measured as the Euclidean distance between the distortion estimated by the registration algorithm and the true distortion.

In cases where a registration algorithm fails completely, the distance from the true distortion may be somewhat unpredictable. To reduce the variability in the target values we have therefore used a soft truncation of the distances. Finally, as a distance equal zero corresponds to the best performance, we have changed the sign of the target value to get a performance measure that increases with increasing registration accuracy.

2.3.2. Performance prediction

The trained neural network will, during run-time, use extracted image features to automatically predict the performance for the set of available registration algorithms. This is done for each region in the image, resulting in a list of scores corresponding to the predicted performance for each algorithm for that region. These scores can then be used to select both the regions that are best suited for registration and the registration algorithm to be applied to each region.

2.3.3. Selecting regions and algorithms

The aim of the region selection is to rule out regions that may decrease the quality of the registration and also to reduce the computational load of the process. A low maximum score for

a region will indicate that it is not suited for registration. Hence, regions are selected according to their maximum score. In addition, care is taken to ensure a sufficient spatial distribution of regions over the image. When the selection of regions is finished, the algorithm to be used for each remaining region is simply selected by picking the one with the highest predicted performance.

2.4. Transform estimation

In this process, the registration algorithm selected for each pair of regions is used to estimate the transform needed to co-register that pair. The result is a set of local transforms. This set is analysed to remove potential errors, and then the remaining set of transforms is used to estimate the global transform. The details of this process are described in the next sections.

2.4.1. Registration of regions

The registration of regions is performed using the algorithms selected according to the performance prediction. The algorithms are fetched from a library offering a selection of registration algorithms with different characteristics.

Algorithm library

Any library offering a variation of registration algorithms could in practice be used. To find a suitable library, we have reviewed and evaluated a range of existing tools offering algorithms needed in the registration process. Based on this evaluation, the ITK/Insight library (Ibanez *et al.*, 2005) was selected to provide the basic registration algorithms. This is a C++ library, originally developed for use in medical imaging, which contains, among other things, a selection of similarity metrics and optimizers for image registration. Image registration is seen as an optimization problem. The similarity metrics are used to compare the fixed and the moving image and provide the criteria for the optimizer, while the optimizer searches through the space of transform parameters to find the best match. In our case, the transform can be selected as one of translation, rotation or affine transform.

The combination of metric and optimizer defines the registration algorithm. From the library we have selected a set of metrics and optimizers, designed to handle different situations. A subset of combinations of these algorithms constitutes our set of registration algorithms. This will be described in the next sections.

Similarity metrics

The similarity metrics are used to measure the correspondence between images (or regions). Three metrics, with different characteristics, were selected from the Insight toolkit:

- *Mean Squares*. This metric computes the mean squared pixel-wise difference in intensity between image A and B over a region. It is simple to compute and has a relatively large capture radius, but even linear changes in intensity can result in a poor match.
- *Normalized Correlation*. This metric computes pixel-wise cross-correlation and normalizes it by the square root of the auto-correlation of the images. The metric is insensitive to multiplicative factors between the images, but has a relatively small capture radius. It is robust to white noise, but sensitive to clutter, occlusion and nonlinear contrast changes.
- *Mutual Information*. Mutual information (MI) measures how much information one random variable (image intensity in one image) tells about another random variable (image intensity in the other image). Hence, the actual form of dependency does not have to be specified, and a complex correspondence between image values can be modelled. Three different

variants of MI have been included: Mattes MI (Mattes *et al.*, 2001), Viola-Wells MI, and Normalized Viola-Wells MI (Viola and Wells 1997).

Optimizers

The optimizer will optimize the similarity metric criterion with respect to the transform parameters. Three optimizers with different characteristics were selected from the Insight toolkit:

- *Gradient Descent*: This optimizer advances parameters in the direction of the gradient, the step size being governed by a learning rate. The drawback of this optimizer is that the steps depend on the values of the gradient. This can however be an advantage for problems where the derivatives are smooth and monotonic.
- *One Plus One Evolutionary* (Styner *et al.*, 2000): This optimizer follows a strategy that simulates the biological evolution of a set of samples in the search space. It generates random samples around the current position in the parametric space. It can perform better than gradient descent type optimizers when similarity metrics are noisy.
- *Regular Step Gradient Descent*: This optimizer advances parameters in the direction of the gradient, computing the step size with a bipartition scheme. The Regular Step Gradient Descent will advance at a more stable rate than the other two optimizers.

Registration algorithms

A registration algorithm is defined as a combination of a similarity metric and an optimizer. We selected a subset of 10 combinations from the set of metrics and optimizers described above. This set was found to offer a sufficient selection of algorithms with different characteristics. The set of combinations is summarized in Table 1.

Algorithm	Similarity metric	Optimizer
M1	Mattes Mutual Information	Regular Step Gradient Descent
M2	Normalized Correlation	One Plus One Evolutionary
M3	Mean Squares	Regular Step Gradient Descent
M4	Mattes Mutual Information	One Plus One Evolutionary
M5	Normalized Correlation	Regular Step Gradient Descent
M6	Viola-Wells Mutual Information	Gradient Descent
M7	Viola-Wells Mutual Information	One Plus One Evolutionary
M8	Normalized Viola-Wells Mutual Information	One Plus One Evolutionary
M9	Mean Squares	One Plus One Evolutionary
M10	Normalized Viola-Wells Mutual Information	Regular Step Gradient Descent

Table 1: Our library of registration algorithms.

2.4.2. Rejection of erroneous local transforms

At times, some of the local transform estimations may fail. For a best possible result, these should be excluded from the estimation of the global transform. Hence, we have included a procedure for detecting and removing the erroneous local transforms prior to the global transform estimation. In the system described, all local transforms are of the same type. Hence, they will all have the same transform parameters. Assuming that the relative distortion between the fixed image and the moving image is a translation, rotation, scaling, or a combination of these, the true value of each transform parameter can be approximated by a linear function of the coordinates of the centre of the region. These linear functions are unknown, but based on the

set of estimated local transforms such a function can be fitted for each transform parameter. The fitting is done by using iteratively re-weighted least squares (Heiberger and Becker 1992; Huber, 1972; Huber, 1981), which is a robust method. Large fitting error for a region indicates that the local transform estimation has failed for that region. Regions where this error exceeds a given threshold are therefore excluded from the final global transform estimation.

2.4.3. Global transform estimation

The estimation of the final global transform is performed based on the set of local transforms that remain after the removal of the erroneous ones. First, a set of point coordinates are selected from regions in the moving image, and then the target points are determined by applying the estimated transform for each region to the selected points. The number of points that are selected for each region depends on the type of transform (translation, rotation or affine) that was selected for the local registration.

The total set of selected points and corresponding target points makes up our set of tie-points (or control points). This set of points is used as input to a warping method provided by ENVI/IDL¹. This software offers a selection of common approaches for warping and resampling, which we have integrated in our registration tool. For the experiments reported here, we have chosen to use their RST (Rotation, Scaling and Translation) based warping method and bilinear interpolation.

3. EXPERIMENTAL RESULTS

The approach presented in this paper is intended for co-registration of time series of remote sensing images originating from identical or similar sensors. This registration is typically performed after a model-based geo-referencing based on satellite parameters. One challenge in the registration of time series of images is that a scene may change considerably with time, due to changing seasons and weather conditions. Hence, we have chosen to evaluate our registration approach on image pairs with varying degrees of difficulty in terms of differences between the images to be registered. In the next sections, our data set, the evaluation procedure, the procedure used for training and the different experiments are described. In addition to an evaluation of the performance, some additional experiments have been performed to evaluate the effects both of using a region-based strategy and of the size of the regions that are used.

3.1. Data set

As we wanted to evaluate the performance for several degrees of difficulty in terms of changes in scene appearance, we have selected three classes of images with *no differences* in content, *moderate differences* and *large differences*:

- *No differences* in content:
 - One *Quickbird image* (Q, Figure 2) with resolution 0.6 m, of the centre of Oslo. Image pairs were created from the same image.
- *Moderate differences* in content:

¹ The ENVI® (Environment for Visualizing Images) and IDL® (Interactive Data Language) software products from ITT Visual Information Solutions.

- A pair of *Envisat ASAR images* (E_1 and E_2 , Figure 3) with resolution 100 m, covering an area around the Oslo fjord in Norway. Challenges are here related to differences in soil moisture and wet-snow cover.
- A pair of *Landsat TM images* (L_1 and L_2 , Figure 4) with resolution 25 m, covering mountainous areas in Norway. The images have varying snow cover. In L_1 large areas consist solely of snow and contain no texture.
- A pair of *NOAA-AVHRR images* (N_1 , and N_2 , Figure 5) with resolution 1 km, covering Norway. Challenges are related to clouds and differences in snow cover.
- *Large differences in content:*
 - Another pair of *NOAA-AVHRR images* (N_2 and N_3 , Figure 5) with resolution 1 km, covering Norway. Challenges are related to substantial differences in both cloud and snow coverage.

To be able to measure the registration performance in each case, our image pairs were chosen from a set that was already manually co-registered, and then known geometric distortions were applied to one image from each pair. Hence, we could later evaluate performance by comparing the estimated distortion with the true applied distortion (see Section 3.2 for more details).

In addition, we tested the approach on an image pair that was not initially co-registered, and where the geometric distortion was unknown and caused by differences in satellite paths:

- A pair of *MODIS images* (M_1 and M_2 , Figure 6) from different dates with resolution 250 m, covering an area around Stockholm in Sweden.

The size of all the images was approximately 1000×1000 pixels.



Figure 2: A Quickbird image (Q) of the centre of Oslo. QuickBird image © DigitalGlobe, 2006, distributed by courtesy of Eurimage.

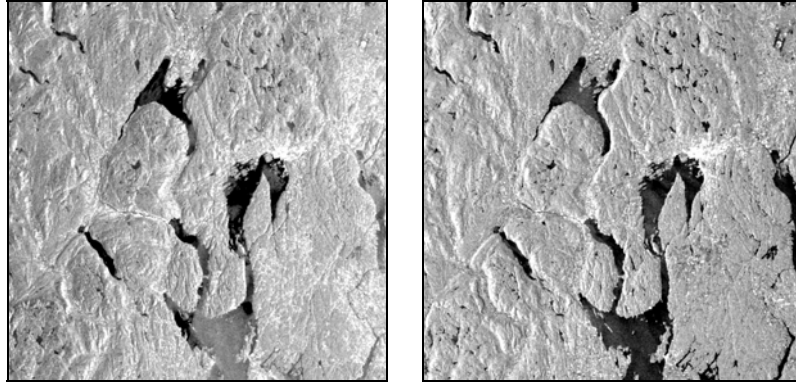


Figure 3: The Envisat ASAR image pair (E_1 and E_2), © 2005 ESA. The images cover an area around the Oslo fjord with variations in soil moisture and wet-snow cover.

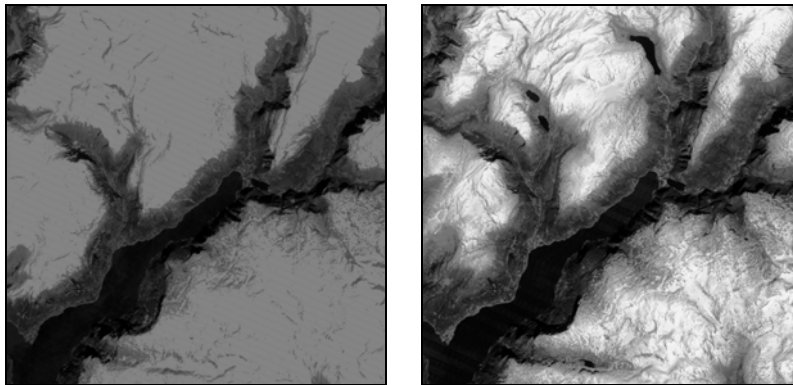


Figure 4: The Landsat image pair (L_1 and L_2). The images cover mountainous areas in Norway and are acquired during the snow melting season.

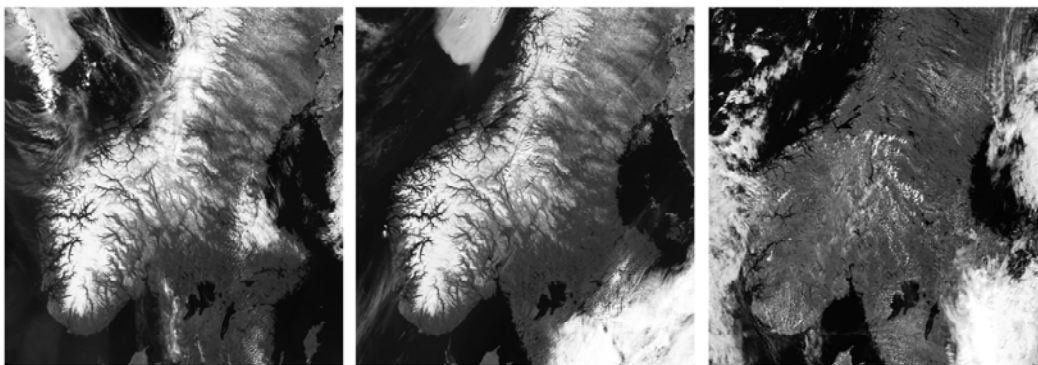


Figure 5: The NOAA-AVHRR image triplet (N_1 , N_2 and N_3). The images cover Norway and were acquired at different times during the snow melting season.

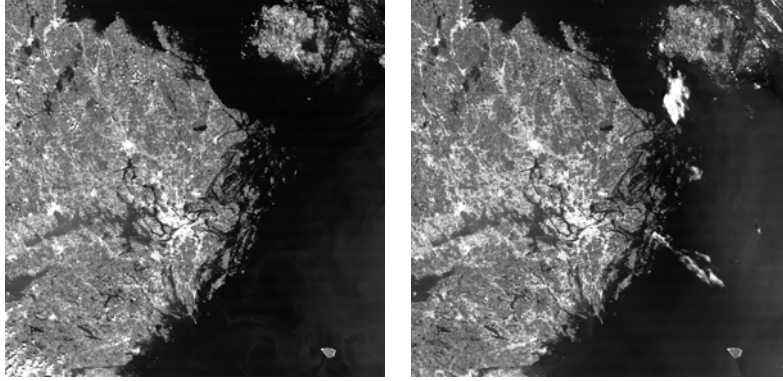


Figure 6: The MODIS image pair (M_1 and M_2). The images cover an area around Stockholm in Sweden. The images are acquired at two different dates during the summer season.

3.2. Evaluation procedure

To be able to properly evaluate and compare registration accuracies, we have chosen to perform our experiments on images where the distortion is known, as this enables computation of objective measurements in the form of RMS (root mean square) errors. During operative use of the system, however, the distortion will not be known.

The evaluation procedure makes use of two images, A and B, which are already registered to each other. Image B is then transformed using some known transformation U_0 to obtain image B'. The registration approach to be evaluated is then used to register image B' back to image A, resulting in the transformation U_R . Ideally the estimated transform for this registration should equal U_0^{-1} . The quality of the registration can thus be measured by the RMS residual differences between $U_0^{-1}(x)$ and $U_R(x)$, or, equivalently, their corresponding displacements.

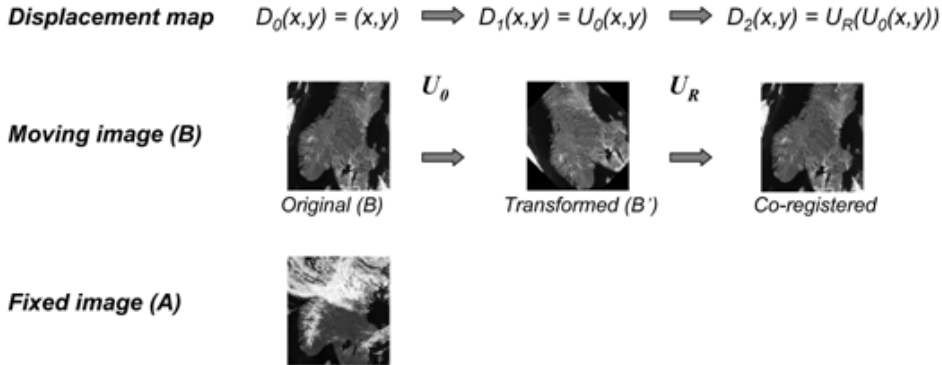


Figure 7: Transforms and displacement maps involved in the computation of the RMS error.

To do this we create a displacement map, $D_0(x,y) = (x,y)$, where all transformations applied to image B (the moving image) are also applied to this displacement map (see Figure 7). From such a displacement map, we can then compute the RMS error as follows:

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n error_i^2} , \quad (1)$$

where n is the number of pixels in the image and, for each pixel $i=(x,y)$, the $error_i$ is the Euclidean distance (in pixels) between $D_0(x,y)$ and $D_2(x,y)$. D_0 and D_2 are the displacement maps associated with the original and the registered image, respectively. This means that for each pixel i , $error_i$ expresses how far pixel i has been moved from its true position.

Image pairs to be used in the evaluation were produced from the data set described in Section 3.1, by keeping one image unchanged and transforming the other one. The following transformations were applied, where the parenthesis give the abbreviations used in the tables:

- *Identity*,
- *Translation* by 2, 4 and 8 pixels (T2, T4, T8),
- *Rotation* by 0.25, 0.5 and 1.0 degrees (R .25°, R .5°, R 1°),
- *Scaling* of 0.5%, 1.0% and 2.0% (enlargements) of the original (S .5%, S 1%, S 2%),
- A composition of a *translation* by 2 pixels, a *rotation* by 0.25 degrees and a *scaling* of 0.5% (T 2/S .5%/R .25°).
- A composition of a *translation* by 4 pixels, a *rotation* by 0.5 degrees and a *scaling* of 1.0% (T 4/S 1%/R .5°).

The magnitude of these relative distortions corresponds to displacements from 0 to 10 pixels. Currently, the approach is not designed to handle much larger distortions than this.

3.3. Training

A standard neural network with one input layer, one hidden layer and one output layer was defined. The 26 features described in Section 2.2.1 constituted the input layer, while performance values for the 10 registration algorithms (Table 1) constituted the output layer. The number of hidden nodes was chosen to be 15. Direct connections from the input layer to the output layer were not allowed. This gave a total of 565 parameters (weights for the connections in the network) that needed to be estimated in the training phase.

The training was performed as described in Section 2.3.1, using a set of approximately 4000 different training samples (pairs of image regions) fetched from images originating from different types of sensors. The input (image features) and the target values (performance in terms of distance from true distortion) were first normalized, and then the parameters (weights) of the net were estimated with Splus (statistical software package) using least squares fitting and weighting decay (Venables and Ripley 1994).

3.4. Evaluation of performance accuracy

To get objective measurements of the performance accuracy, we have followed the evaluation procedure described in Section 3.2. The experiments were performed on image pairs from the three classes of images with no differences, moderate differences and large differences, as described in Section 3.1.

Table 2 gives a summary of the results obtained for the different classes of images. Different parameter settings with combinations of local transforms and region size (50×50 and 100×100 pixels) were tested. For image pairs with no differences and moderate differences, only small variations in RMS errors were observed for different parameter settings. In Table 2 we have therefore presented the mean of the RMS errors for these experiments. For the image pair with large differences, there were larger variations in the results. Hence, we have here not given the means, but instead included the RMS error obtained for two of the experiments to show the variation.

<i>Distortion</i>	No diff	Moderate differences			Large diff	
	<i>QB</i>	<i>E1, E2</i>	<i>L1, L2</i>	<i>N1, N2</i>	<i>N2, N3</i>	
Identity	0.00	0.54	0.92	0.51	0.47	0.25
T2	0.04	0.54	0.97	0.53	2.00	0.60
T4	0.05	0.57	0.94	0.56	3.05	0.85
T8	0.04	0.53	1.03	0.54	10.24	0.53
S .5%	0.05	0.53	0.99	0.54	1.00	0.45
S 1%	0.04	0.46	1.01	0.58	2.76	0.76
S 2%	0.12	0.50	0.90	0.67	5.19	0.85
R .25°	0.03	0.51	1.01	0.52	1.65	0.47
R .5°	0.05	0.51	0.99	0.58	2.41	0.53
R 1°	0.09	0.48	0.93	0.63	7.55	1.38
T 2/S .5%/R .25°	0.04	0.49	0.98	0.56	2.06	0.75
T 4/S 1%/R .5°	0.07	0.53	1.01	0.60	5.03	5.78
RMS mean	0.06	0.52	0.97	0.57	4.55	1.81

Table 2: Results from the experiments with the adaptive registration approach. The table presents the RMS error in pixels obtained for the different images and distortions. For the image pairs with large differences, both results for a region size of 50x50 (left) and 100x100 pixels (right), are shown. For the other images the variance over the different experiments was small and the RMS mean is shown.

The image pair with no differences was mainly included to verify that the approach works properly under perfect conditions, i.e. when there are no differences in content between the images. A mean RMS error of 0.06 for these cases, verifies this.

The results for the image pairs with moderate differences are also presented in Table 2. As can be seen from this table, the mean RMS error is less than one pixel for each pair of images. However, we do not know the exact sub-pixel accuracy of the manual co-registration used as ground truth for these images as this is difficult to verify. We have for instance observed that through all our experiments we quite consistently get an RMS error of about 0.5 for the E_1 - E_2 image pair, 0.9-1.0 for the L_1 - L_2 image pair and 0.5-0.6 for the N_1 - N_2 image pair. This may indicate that we actually get better sub-pixel accuracy with our automatic methods, than what was the case for the initial manual co-registration.

For the image pair with large differences, the performance was more sensitive to the choice of region size and local transform. In Table 2 we have included the results obtained for two experiments with region sizes 50x50 and 100x100 pixels. As can be seen, the variation is larger here. The experiments with the largest region size gave the best performance. Here, the RMS error was below 1 for all but the two largest distortions. In the cases where the registration failed, the problem for this pair of images was typically that too few regions were selected for registration. Analyzing the images in more detail (Plate 1), we see that quite large areas in image N_3 are covered by fragmented clouds producing a special pattern. Hence, the main problems here may not be due so much to the differences between the images, as the large areas that are less suited for registration due to clouds and homogeneous sea regions.

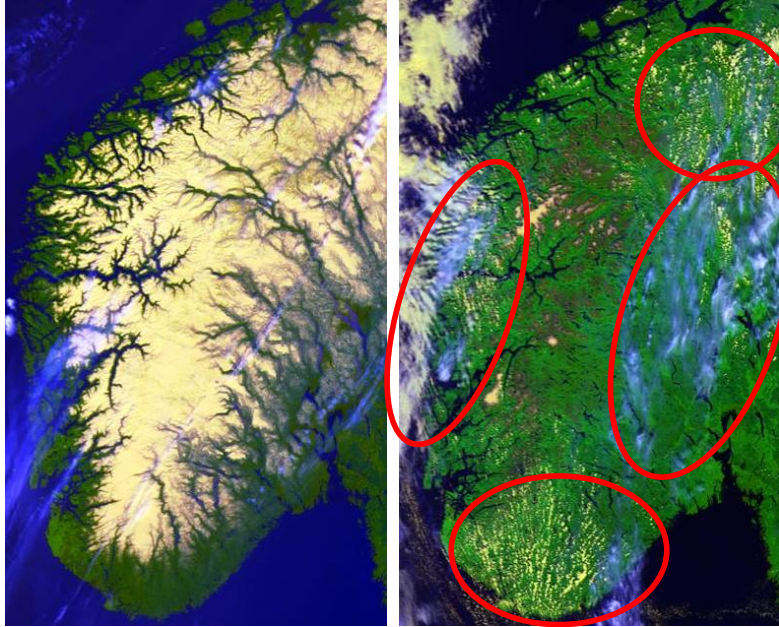


Plate 1: Detail of the N_2 , N_3 image pair. This detail shows that the image N_3 (on the right) contains a lot of fragmented clouds.

Finally, we ran our registration approach on a MODIS image pair (see Figure 6) where the relative distortion was caused by differences in satellite paths rather than a known geometric distortion. For this image pair no manually co-registered result was available, and as the geometric distortion was not known, the RMS error as described in Section 8.2 could not be computed. Instead, in Plate 2, image mosaics, created by alternately fetching tiles from the fixed and moving images, are included to illustrate the quality of the registration. The image mosaics show the island Yxlan in the Stockholm archipelago. The first mosaic is created from the fixed image and the original moving image. Here we see a broken coastline indicating the original distortion. In the second mosaic, created from the fixed and the moving image resulting from the automatic registration process, the same coastline is continuous. This indicates that the registration has been successful.

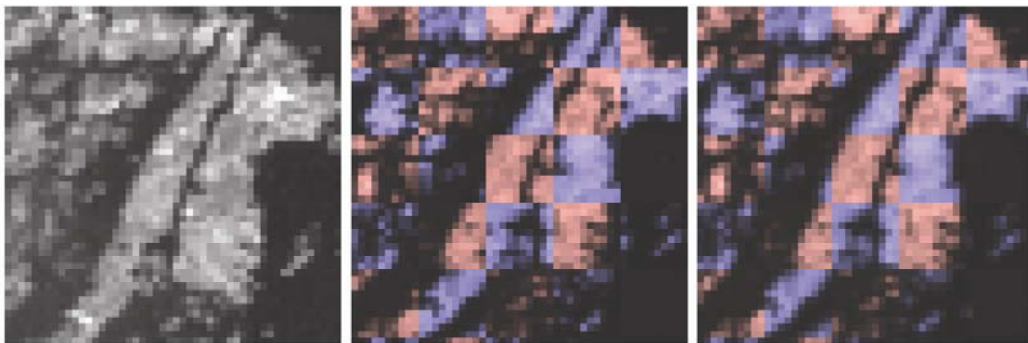


Plate 2: Illustration of the results from the registration of the MODIS image pair for an image detail from the island Yxlan in the Stockholm archipelago. The first image shows the detail in the fixed image. The next image shows a mosaic created from the fixed image and the original moving image before the registration. The red pixels come from the fixed image and the blue pixels from the moving image. The last image shows the corresponding mosaic created from the fixed image and the moving image after registration.

3.5. Examining the effect of region selection

When using a region-based strategy, a necessary condition for successful registration is that regions are selected in areas containing the most informative structures. We present two examples to demonstrate this. Plate 3 shows the regions that have been selected for registration of two *Envisat ASAR* images and two *Landsat TM* images. For the Landsat image pair, we see that regions completely covered by snow, containing no characteristic details, are not selected. For both image pairs, we see that the regions that *are* selected are typically found in areas containing edges of water bodies, and these are indeed the most informative structures in these images. In the illustration, the algorithm selected for each region is also indicated, showing that algorithm M4 (a combination of Mattes Mutual Information and One Plus One Evolutionary) appears to be the most frequently selected registration algorithm. However, the distributions of the selected algorithms are somewhat different for the two image pairs.

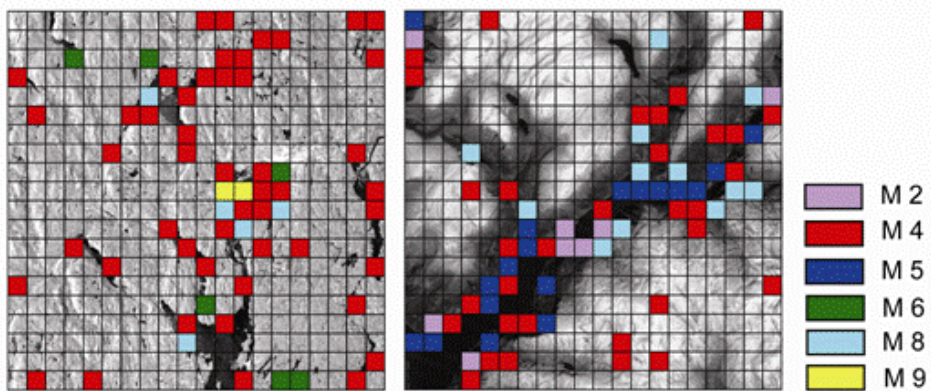


Plate 3: Regions and algorithms that have been selected for registration of two Envisat ASAR images (left) and two Landsat TM images (right). The colour coded regions were the ones selected for registration, and the colour indicates which algorithm that was chosen for each region. (An overview of the available algorithms is given in Table 1).

To evaluate the adaptive region selection process, we have also compared this to a traditional registration approach where the whole image is used in the registration. To test this, we used our adaptive approach only to select regions, using the same registration algorithm for all regions. The results of this process were compared to those obtained when applying the corresponding registration algorithms to the entire image as a whole. Experiments were performed on the Landsat image pair (L_1, L_2) and the NOAA-AVHRR pair (N_1, N_2) with the four largest distortions from Section 3.2 (T8, R1, S2 and T4-S1-R.5). In Table 3 we have presented the results obtained with the registration algorithm with the best performance (M4). The table gives the RMS errors obtained with this algorithm both with and without the use of regions. From this table it can be seen that the use of regions generally results in much better registration accuracies.

<i>Distortion</i>	Using regions (100x100)		Using entire image	
	<i>L1, L2</i>	<i>N1, N2</i>	<i>L1, L2</i>	<i>N1, N2</i>
T8	0.95	0.52	1.37	1.10
S 2%	0.96	0.77	8.69	8.77
R 1°	1.10	0.68	7.58	7.23
T 4/S 1%/R .5°	1.01	0.58	5.73	5.55

Table 3: Comparison of registration results with and without the use of regions.

3.6. Examining the effect of region size

As demonstrated above, the use of regions is fruitful in our registration approach. In our system, the user will be able to choose the size of regions. Although the method is not very sensitive to the choice of region size, there is a relationship between the ideal region size, and the contents and the resolution of the image. Typically, we would expect that for images containing large structures a larger region size would be better suited. Through our experiments we have found that region sizes of 50×50 and 100×100 pixels both work well. Smaller regions should be avoided, as these will generally not contain enough information.

For our experiments with an image size of 1000×1000 pixels, larger regions were not so well suited, as this will result in a too low total number of regions. In terms of registration quality there is however no dependency between image size and region size as long as the region size allows for a division of the image into a sufficient number of regions. Finally, a user will often have some knowledge of the order of magnitude of the distortion, and should then also select a region size that ensures that corresponding regions in the fixed and moving images will overlap enough to give a match.

4. CONCLUSION

The approach presented in this paper, is intended for co-registration of time series of remote sensing images. For a user who needs to handle different types of time series, it can be difficult to choose the most appropriate registration algorithm for each case. Our system aims at simplifying this process for the user, by offering a tool that automatically chooses a registration algorithm based on image characteristics. During a training phase the system learns the relationship between image features and performance for a set of registration algorithms. At run-time, the system can then predict the performance for each algorithm from image features. This strategy is applied to image regions. Hence, the most appropriate algorithm can be selected for each region, and regions unsuited for registration can be discarded.

The approach has been tested on a selection of image pairs from different sensors, presenting different types of contents and different degrees of difficulty in terms of differences between the images to be registered. The results from these experiments have demonstrated that the adaptive approach works well and that the same approach can be applied to different types of time series with different types of contents without tedious testing and tuning. The approach is also able to handle images with at least moderate differences in contents, selecting different registration algorithms for different regions and discarding regions that are not suited for registration. In conclusion, the experiments show that both the learning-based and the region-based approaches are fruitful.

In addition to simplifying the registration process for the user, this approach may also open for new possibilities, like simple integration of predefined masks. Cloud masks could, for instance, easily be exploited by excluding regions covered by the mask from the local registration. Furthermore, our system can easily be extended to include more registration algorithms. In future work we will also consider inclusion of a multi-resolution strategy to extend the approach to work for larger distortions.

Acknowledgements: This work was supported by ESA/ESRIN and the Norwegian Research Council.

REFERENCES

- Chalermwat, P., 1999. *High performance automatic image registration*, Ph.D. dissertation, George Mason University, Fairfax, Virginia, 125 p.
- Eikvil, L., P.O. Husøy, and A. Ciarlo. 2005. *Adaptive Image Registration*, Proceedings of ESA-EUSC 2005 workshop on Image Information Mining – Theory and Application to Earth Observation, 5-7 October 2005, Frascati, Italy, URL: http://earth.esrin.esa.it/rtd/Events/ESA-EUSC_2005/Ar22_Eikvil.pdf, ESRIN (last date accessed: 27 January 2009).
- Fedorov, D., L.M.G. Fonseca, C. Kenney, and B.S. Manjunath, 2002. *Automatic Registration and Mosaicking System for Remotely Sensed Imagery*, Proceedings of SPIE 9th International Symposium on Remote Sensing, 22-27 September 2002, Crete, Greece, pp. 444-451.
- Fonseca, L.M.G., and M.H.M. Costa, 1997. *Automatic Registration of Satellite Images*, Proceedings of X Brazilian Symposium of Computer Graphic and Image Processing, October 13-16, 1997, Campos de Jordao, pp. 219-226,.
- Haralick, R., K. Shanmugam, and I. Dinstein, 1973. *Textural Features for Image Classification*, IEEE Transactions on Systems, Man and Cybernetics, **3**(6):610-621.
- Heiberger, R.M., and R.A. Becker, 1992. *Design of an S Function for Robust Regression Using Iteratively Reweighted Least Squares*. Journal of Computational and Graphical Statistics, **1**(3):181-196.
- Huber, P.J., 1972. *Robust Statistics: A Review*, The Annals of Mathematical Statistics, **43**(4):1041-1067.
- Huber, P.J., 1981. *Robust Statistics*, John Wiley and Sons, New York, 308 p.
- Ibanez, L., W. Schroeder, L. Ng, J. Cates and R. Hamming, 2005. *The ITK Software Guide*. Insight Software Consortium, URL: <http://www.itk.org/ItkSoftwareGuide.pdf> (last date accessed: 27 January 2009).
- Mattes, D., D.R. Haynor, H. Vesselle, T.K. Lewellyn, and W. Eubank, 2001. *Non-rigid multimodality image registration*, Proceedings of Medical Imaging 2001: Image Processing, 19 February 2001, San Diego, California, SPIE Vol. 4322, pp. 1609–1620.
- Le Moigne, J., J. Morisette, P. Jain, H. Stone, I. Zavorin, A. Cole-Rhodes, K. Johnson, R. Eastman, and N. Netanyahu, 2004. *Registration of Multiple Sensor Earth Science Data*, NASA's Earth Science Technology Conference, 22-24 June 2004, Palo Alto, California. URL: <http://esto.nasa.gov/conferences/estc2004/papers/a7p3.pdf>, Earth Science Technology Office, NASA (last date accessed: 27 January 2009).
- Rignot, E.J.M., R. Kowk, J.C. Curlander, and S.S. Pang, 1991. *Automated Multisensor Registration: Requirements and Techniques*, Photogrammetric Engineering & Remote Sensing, **57**(8):1029-1038.
- Styner, M., C. Brehbuhler, G. Szekely, and G. Gerig, 2000. *Parametric estimate of intensity homogeneities applied to MRI*, IEEE Transactions on Medical Imaging, **19**(3):153–165.
- Venables, W.N., and B.D. Ripley, 1994, *Modern Applied Statistics with S-Plus, First Edition*, Springer-Verlag, New York, 462 p.
- Viola, P., and W.M. Wells, 1997. *Alignment by maximization of mutual information*, International Journal of Computer Vision, **24**(2):137–154.