ADAPTIVE IMAGE REGISTRATION

Line Eikvil[1], Per Ove Husøy[2] and Alessandro Ciarlo[3]

[1]Norsk Regnesentral, P.O. Box 114, Blindern, N-0314 Oslo, Norway, Email:Line.Eikvil@nr.no
[2]Norsk Regnesentral, P.O. Box 114, Blindern, N-0314 Oslo, Norway, Email:Per.Ove.Husoy@nr.no
[3]ESA-ESRIN, C.P. 64, 00044 Frascati (RM), Italy, Email:Alessandro.Ciarlo@esa.int

ABSTRACT/RESUME

The objective of the study described in this paper has been to develop a registration approach that can automatically choose the most appropriate registration methods based on image characteristics. Methodology for intelligent selection of methods have been developed and combined with existing methods and tools for image matching and registration.

The approach works by dividing the pair of images to be co-registered into smaller sub-regions and extracting features from each region. Based on the extracted features the performance of each of the available methods is predicted by using a neural net. For regions with sufficiently high scores, the method with the best rating is used to perform a local co-registration. This results in a set of local transformations, which is used to find the global transformation. The methods have been implemented in a software-tool that has demonstrated satisfactory results for different types of image sequences.

1.   INTRODUCTION

The study of time series of satellite images is an important task in many remote sensing applications where the objective is to study different environmental phenomena. For such applications a co-registration of the satellite images acquired at different times is important. This co-registration is often performed using a combination of manual and automatic registration techniques. However, for a multi-temporal problem where the number of images becomes large, manual correction of images is often not feasible. Hence, a fully automatic procedure would be desirable.

Automatic techniques for performing each step of the process do exist, but selection of the appropriate method depends on the application and the image specifics. Hence, a single registration scheme will generally not work for all different applications. Consequently there exists a large number of different automatic registration techniques, all of which essentially perform the same task, the only difference being that they are constrained to working with a very small range of images.

For a user that needs to work on different types of time series, it would be useful to have a more general tool for image registration that could be used for several applications. It has been estimated that more than 90% of the studies in remote sensing that could have used automated approaches for registration of images, did not use it [7]. The lack of a more general tool for helping in this process may be one of the reasons for this.

In this study we propose an approach for obtaining a more general registration scheme. The idea is to provide a selection of registration methods, and to develop an approach for intelligently choosing between them. Hence, a methodology for intelligent selection of methods based on image characteristics has been developed and combined with existing methods and tools for image matching and registration.

The idea of integrating several methods for image registration into one tool and provide the system with some form of intelligence to automatically select the method best suited for each set of images is not entirely new. A similar idea was for instance proposed already in 1991 by Rignot et al [9]. Later such an approach has also been suggested as a viable solution by Fonseca & Costa [3]. However, no methods for doing this have been presented, and solutions have yet to appear.

A few papers present work towards the development of systems that provide a selection of different approaches to choose from [1, 4]. They do however not provide the intelligence that makes the system automatically choose the best method for the image to be processed.

2.   IMAGE REGISTRATION

Image registration is performed on a series of at least two images, where one of these images is the *reference image* (or *fixed image*) to which all the others will be registered. The other images are referred to as *sensed images* (or *moving images*).

The steps in an image registration process generally include feature extraction, feature matching, transformation selection and image resampling. The first two tasks go into the identification of tie-points,

which are the points that mainly determines the quality of the image registration and that are most dependent on the image characteristics.

Two main approaches exist; *area-based* and *feature-based*. When reliable, the area-based methods can be very accurate. Mutual information is a more recent and interesting similarity measure, which is more robust to differences in intensity values than the more traditional correlation-based measures. Feature-based methods can be needed when images do not contain enough texture for the area-based methods to work well, or if there are larger differences between the sensed image and the reference image. Region-based features can then be suited for images with several homogeneous regions, while edge-based features can work well for images with more detailed information.

3. OVERVIEW OF THE APPROACH

The objective of the study was to address co-registration of multi-temporal series of images resulting from the same sensor, disregarding – at least in an initial phase - multi-sensor registration. The approach was however intended to be general in that it should be able to handle sequences containing images acquired under different conditions. To be able to provide this, methods with different characteristics would need to be included.

Existing methods and tools for the different steps in the registration process were reviewed and evaluated. Based on this evaluation the ITK/Insight library [6] was selected to provide the basic registration methods. This is a C++ library, originally developed for use in medical imaging. The library contains, among other things, similarity metrics based on normalized correlation, mean squares and varieties of mutual information in combination with a selection of optimizers for searching through the space of transform parameters.

A suitable wrapper was built around the library in addition to a user interface, flow control, and, most importantly, a method for the intelligent selection and application of the tools provided by the library.

In summary, the proposed approach starts by diving the reference image and the sensed image(s) into rectangular sub-regions. Features are then extracted from each region. From the set of matching methods available (from the ITK/Insight toolkit), and based on the features extracted from each region, the expected performance of each of these methods is predicted using a suitably trained neural network. Based on this prediction a final selection of regions and methods is performed, and for each selected region a local co-registration is performed. The final global transformation is then found by combining the set of local transforms.

An illustration of the steps involved in this approach is given in Fig. 1. The details of each step will be treated in the next sections.
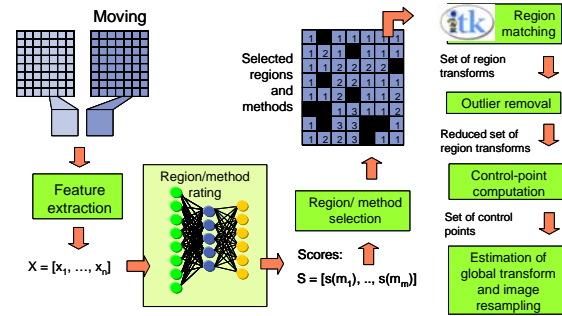

Figure 1. *Overview of the approach.*

4. FEATURE EXTRACTION

Our approach is based on dividing the images in smaller regions and performing a number of local registrations rather than trying to handle the entire image at once.

There are several reasons for this. First, in remote sensing the images are often quite large, which means that using the entire image for matching will be computationally very expensive. Also, some of the areas in the image may not match very well and may not be suited for registration, e.g. because of clouds, and it would be useful to be able to identify and discard such regions. Finally, when using an approach where the method to be used for matching is selected automatically based on image characteristics, it can be useful to permit different methods to be used for different regions based on local characteristics. Hence, prior to the feature extraction, the images are divided into regions. Features are the extracted for each region.

4.1. Definition of regions
It is important that the division into regions is robust so that the regions from the reference image and the sensed image are comparable. We have therefore chosen a simple and robust approach that divides both the reference image and the sensed image(s) into a grid of rectangular sub-regions. The images (and thereby the sub-regions) are required to have a reasonably small relative distortion and cover approximately the same area. Using sub-regions much larger than the expected distortion, corresponding regions in the reference and sensed images will overlap enough to ensure capture by the registration algorithm.

Dividing the image into sub-regions could also be achieved by using image segmentation techniques, but this was not considered to be sufficiently robust. The chosen approach also has the additional advantage of direct control over the regions, both their size and their distribution over the image.

4.2. Feature extraction

The purpose of the feature extraction is to derive features that describe image characteristics that are relevant in a co-registration process and which can be used to select a subset of regions and choose an appropriate method for each. Which features that are relevant will depend on the metrics that are used in the matching. One could compute the actual metrics and use these directly to decide what to do. This would however be much too time-consuming, as it would require that all methods were tested for all regions. This was therefore not considered as a viable alternative. Instead we have selected features that can say something about the characteristics and correspondence between the reference image and the sensed image(s).

The image characteristics should say something about the information content of the image, e.g. whether the image is dominated by homogeneous areas, textured areas, edges and lines and may be something about the noise level. The regions that contain characteristic patterns or details will often provide more accurate matches and should be retained. At the same time the regions where there is no correspondence and those where the reference image and the sensed image have different contents should be discarded. For EO-images such a situation is often caused by clouds.

The result of these considerations was that a set of features based on differences between image statistics, image textures and difference between image textures was chosen.

4.3. Features

The *texture features* that are used are computed from the Grey Level Co-occurence Matrix (GLCM). These were originally introduced by Haralick et al [5] and are some of the most common texture measures. They are based on second-order statistics, and are computed by finding repeated occurrences of grey-level configurations in a texture. The method is general and not specialized for certain types of textures.

The *statistical features* that are included are mean, variance and entropy computed for each region. In addition to means for each region, zone means within the regions are computed. We also compute a measure of registrability, which was introduced by Chalermwat [1] and is said to be a measure of standard deviation of samplings of self-correlation. It will indicate whether

the subimage has strong features that can result in correlation peaks. Finally, the gradient magnitude computed by the Sobel operator was included.

The *difference features* are computed as differences between the features computed from the reference image and the features computed from the sensed image. Both differences between texture features and statistical features are computed. For the zone means the Euclidean distance between the means for each region is computed in addition to the variance over the differences between the zone means.

The result of the feature extraction is a feature vector $X=[x_1, ...x_n]$ of length n (n=number of features) for each region.

5. SELECTION OF REGIONS AND METHODS

The features that have been extracted from regions form the basis from which to determine both which regions to use in the registration and which methods to use for each region. Hence, what is needed is to establish a correspondence between extracted features and the registration methods' performance. 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 method.

- By using a training approach to establish the correspondence between image characteristics and method performance.

In practice, a combination is often used. In our approach we have used a priori knowledge in our choice of features, while we have chosen to use a training approach to establish the correspondence. We have viewed the problem as that of predicting the performance by estimating a score for each registration method from the features extracted from a region. The higher score, the better the method would be expected to perform. This can then be seen as a regression problem, where the objective is to predict a performance score based on the features. For this regression problem we have chosen to use a neural net.

5.1. Training

For the regression we define a neural network where the number of input nodes corresponds to the number of features and the number of output nodes corresponds to the number of methods. For the training of this network the feature vectors extracted from the regions will be used as input. For the target values a measure of each method's performance for the corresponding regions is needed. We therefore needed to define a measure of this performance.

One possible choice was to do the matching for a series of regions for all the methods and use the metric value as the performance measure. But as the registration methods may end up in a local minimum (or maximum) that does not correspond to the true minimum (maximum), this would not necessarily say anything about the correctness of the match that was found.

We therefore decided instead to use a performance measure based on the distance from the true transformation. This was obtained by using pairs of images for which the true distortion was known. All the available registration methods were then applied to all regions, and then the distance from the estimated transform obtained for each registration method and the true transform was computed. (The selection of registration methods that were used will be described in Section 6).

A potential problem with this approach is that if a matching method fails completely for a region, the distance from the transformation might be somewhat unpredictable. To overcome this, a truncation of the distances was applied to reduce the variability in the target values. Finally, as a distance equal zero would correspond to the highest performance, we changed the sign of the target value to achieve a performance measure that increased with increasing matching quality.

5.2. Performance prediction
When the neural net has been trained it is used to predict the performance for the available set of methods based on the features extracted for a region. This prediction will result in a score vector for each region, containing the predicted performance for each of the n methods, $S=[s(m_1), ...s(m_n)]$.

5.3. Selection of regions and methods
The predicted performance is then used to select a subset of regions and to select the registration method to be applied for each of these regions.

The selection of regions is performed in order to avoid regions that can reduce the quality of the registration and also to reduce the number of regions to be matched and thereby the computational load of the process. An overall low score for a region will indicate that this region is not suited for registration. Hence, the maximum score for each region is investigated in the region selection process. In addition, care is taken to ensure that a sufficient spatial distribution of regions over the image is retained.

Going through the regions sequentially, the following strategy is followed for the selection of regions. The regions with the highest score are selected, and a

rectangular neighbourhood around the selected region is (temporarily) blocked for selection. If there is an insufficient number of regions remaining for selection, the size of the blocking area around the selected regions is reduced to free more regions for selection. By following such an approach iteratively, a gradual increase in region density is obtained.

When the selection of regions is finished, the method to be applied for each region is chosen by simply taking the one with the highest score.

6. TRANSFORM ESTIMATION

During the transform estimation, the regions are first co-registered one-by-one with the method that has been selected for that region. This results in a set of locally estimated transformations. From this set, obvious outliers are removed, and then the remaining set of transformations is used to estimate the global transformation.

6.1. Co-registration of regions
The co-registration of regions is performed between corresponding regions from the referenced and sensed images using the method determined through the performance prediction. The registration is carried by using methods provided by the ITK/Insight library [6].

Registration in ITK/Insight is performed by optimizing the parameters of a transform T. The optimality criterion is provided by image metrics that compare the moving image with the fixed image. Hence, the combination of metric and optimizer defines the registration method.
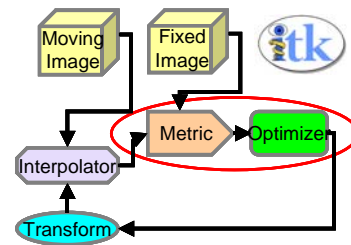
Figure 2. *Overview of the ITK/Insight registration.*

For our application a set of metrics and optimizers designed to handle different problems were selected from the ITK/Insight library. A subset of combinations of metrics and optimizers was then selected as the set of registration methods.

6.2. Metrics
The metrics are used to measure the correspondence between images (or regions). Four different metrics from the Insight toolkit were selected:

- *Mean Squares Metric*. 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. It relies on the assumption that intensity representing homologous points must be the same in both images, and any linear changes in the intensity result in a poor match value.

- *Normalized Correlation Metric*. This metric computes pixel-wise cross-correlation and normalizes it by the square root of the auto-correlation of the images. Misalignment between the images results in small measure values. The metric is insensitive to multiplicative factors between the images and produces a cost function with sharp peaks and well-defined minima. On the other hand, it has a relatively small capture radius.

- *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). The major advantage of using MI is that the actual form of the dependency does not have to be specified. Therefore, a complex correspondence between image values can be modeled. Two different variations have been included: Viola-Wells [11] and Mattes [8] Mutual Information.

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

- *Regular Step Gradient Descent:* This optimizer advances parameters in the direction of the gradient where a bipartition scheme is used to compute the step size. The Regular Step Gradient Descent will advance at a more stable rate than the other two optimizers.

- *Gradient Descent:* This optimizer advances parameters in the direction of the gradient where the step size is governed by a learning rate. The drawback of the Gradient Descent 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*: 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 metrics are noisy.

6.4. Metric and optimizer combinations
A registration method is defined as a combination of a metric and an optimizer. From the set of metrics and optimizers described above, a set of 10 combinations was selected. The set of combinations that was selected is summarized in Table 1.

Method	Metric	Optimizer
M1	Mattes Mutual Information	Regular Step Gradient Descent
M2	Normalized Correlation	Gradient Descent
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	Regular Step Gradient Descent
M8	Mean Squares	Gradient Descent
M9	Mean Squares	One Plus One Evolutionary
M10	Normalized Viola-Wells Mutual Information	Regular Step Gradient Descent

Table 1. *Selection of registration methods*.

6.5. Global transform estimation
The result of the matching of regions is a set of local transformations, one transformation for each region. From this set obvious outliers are removed prior to the final transform estimation. The final transform estimation is then performed based on the remaining set of local transformations. A quite simple approach was used for this, computing local control points from the local transforms and using these to estimate the global transform.

7. EXPERIMENTAL RESULTS

For our experiments we have used a data set consisting of image sequences from several sensors presenting different types of challenges:
- Sequences of NOAA-AVHRR data covering Norway and acquired during the melting season. This image set presents challenges such as clouds and varying snow cover.
- Sequences of Landsat data over mountainous areas in Norway, with clouds and varying phenology.
- Sequences of ERS1 data over agricultural areas in France with variation in soil moisture and crop maturity.

Some of the data were selected for training, while others were used for testing.

A standard neural network with one input layer, one hidden layer and one output layer was defined. From the features described in Section 4.3, a set of 26 features were selected. These features constituted the input layer, while scores for the 10 methods 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) to be estimated.

Features and scores were computed for a large set of regions from image pairs with a known distortion. Then both the input and the output parameters were normalized (mean 0 and standard deviation 1), and the parameters of the net were estimated with Splus (statistical software package) using least squares fitting and weighting decay.

In Figure 3 a pair of NOAA-AVHRR images taken over the Southern part of Norway is shown. The image to the left is from May 31, 2003. Here the mountain areas in the south are still covered with snow (red), while there are a lot of clouds further north. The image to the right is from July 7 the same year, and at this time the snow has melted. There is also much less clouds in this image. Hence, the region selection should here need to select areas that are not covered by clouds. Also the regions over the sea are not well suited for registration and should not be selected. Finally, many of the areas that can be used for registration will have quite different spectral signatures due to the differences in snow cover, which means that the correlation-based techniques are expected to perform poorly for these areas.
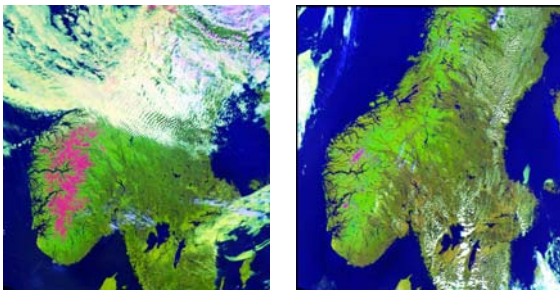

Figure 3. *A pair of NOAA-AVHRR images over Norway.*

In figure 4 the regions selected for the image pair in Figure 3 when using the adaptive registration is showed. The superimposed grid shows the division into sub-regions, and the regions marked with yellow shows the regions that were selected using two different parameter settings. The figure to the left shows regions selected without any blocking, while the figure to the right shows the regions selected with a degree of blocking. As can be seen from this result, the adaptive registration has selected regions over the parts that are not covered by clouds and it has also ignored sub-regions only containing sea.
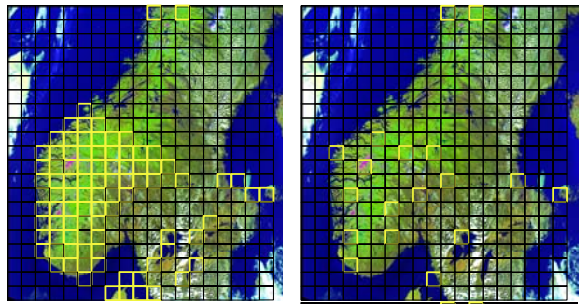

Figure 4. *Selection of regions with different parameters.*

Figure 5 illustrates which method that was selected for each region, where each sub-region is colour-coded according to which method that was applied for that region. As can be seen from this illustration, the method selected most often is method number 4. This method consists of Mattes mutual information metric in combination with the evolutionary-based optimizer. This method combination is able to handle complex correspondence between image values and is also the one that is most tolerant to noisy metrics. Hence, it is well suited for an image pair like this where the differences are quite large. The final registration of this image pair, based on the resulting local transforms was very accurate.
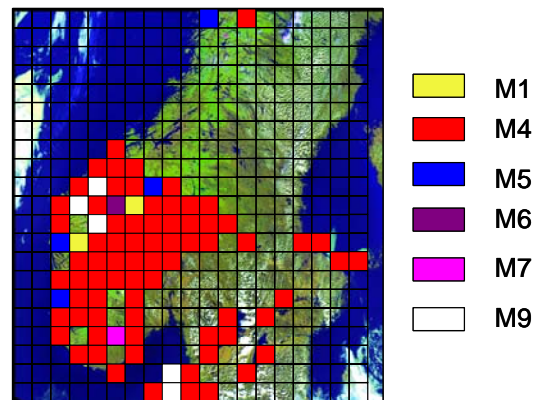

Figure 5. *Selection of methods.*

The approach has also been tested for sequences of Landsat images and ERS1 images. All these images had quite large variations and for most of the test examples, one of the methods based on Mattes mutual information (M1 or M4) were in majority. The registration results are in general quite good (within a pixel) when the distortions are not too large.

8. DISCUSSION AND CONCLUSION

In summary, an adaptive approach for image registration has been presented. Methodology for intelligent selection of methods have been developed and combined with existing methods and tools for image matching and registration.

The current approach has been implemented in a software tool, which is developed in ENVI/IDL and C/C++. It has a simple graphical user interface and is intended for expert users, with a technical background and good understanding of remote sensing imagery and of the problem of image registration.

The approach has been tested on time series of optical and radar EO images and results are promising. The methods have demonstrated ability to automatically select the areas in the images that are best suited for registration, discarding regions covered by snow and very homogeneous areas with little information. At the same time appropriate methods are selected for handling the remaining areas.

On average the registration results are satisfactory. The registration does however at times fail for some of the local regions, leading to an erroneous transform locally and a less accurate transform globally. Currently, a very crude approach is used for detecting such outliers, and the final registration is not very robust to local errors. Hence, more sophisticated methods for detection of these failures could improve the performance and make the registration more robust. Detection of large inconsistencies in the local transforms could also be used to indicate probable failure.

The neural network applied here has been trained on image pairs with quite large differences in contents caused by variations in cloud cover, snow cover, soil moisture, phenology etc. Hence, for images pairs with minor differences, it may currently not choose the most optimal method combination. The training set should therefore be extended to also include these types of examples.

The approach is currently not designed to handle larger distortions between images. Large distortions will especially present problems for the more complex transforms that require optimization of a larger set of parameters. In these cases a large number of iterations are needed for the optimizer to converge, in addition the risk of ending in a local minimum (maximum) increases. Hence, for improved performance for images with larger distortions a multi-resolution strategy would need to be included.

Future work may also investigate approaches for automatic selection of bands for registration of multispectral images. It would also be interesting to look into multisensor registration using this approach.

REFERENCES

1. P. Chalermvat, *«High performance automatic image registration»*. PhD thesis, George Mason University, 1999.
2. L.Eikvil, P.O. Husøy, *«Multiple Image Registration: Review of problem and methods»*. Technical Note. July 2004.
3. D. Fedorov, L.M.G. Fonseca, C. Kenney, B.S. Manjunath. *«Automatic Registration and Mosaicking System for Remotely Sensed Imagery»*. 9th International Symposium on Remote Sensing, 22-27 September 2002, Crete, Greece.
4. Fonseca, L.M.G, and Costa, M.H.M; *«Automatic Registration of Satellite Images»*, X Brazilian Symposium of Computer Graphic and Image Processing, Campos de Jordao, SP, p.219-226, October 13-16, 1997.
5. R. Haralick, K. Shanmugam, I. Dinstein. *«Textural Features for Image Classification»*. IEEE Trans. Systems, Man, Cybernetics, Vol. 3, 610-621, Nov. 1973.
6. L. Ibanez, W. Schroeder, L. Ng, J. Cates, *«The ITK Software Guide»*. Insight Software Consortium, August 2003.
7. R. E. Kennedy and W.B. Cohen. *«Automated designation of tie-points for image-to-image coregistration»*. Int. Journal of Remote Sensing, Vol. 24, No. 17, pp. 3467-3490. 2003.
8. D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. *«Non-rigid multimodality image registration»*. In Medical Imaging 2001: Image Processing, pages 1609–1620, 2001.
9. J. Le Moigne, J. Morisette, P. Jain, H. Stone, I. Zavorin, A. Cole-Rhodes, K. Johnson, R. Eastman, N. Netanyahu. *«Registration of Multiple Sensor Earth Science Data»*. NASA's Earth Science Technology Conference, June 22-24, Palo Alto, California. 2004.
10. E.J.M. Rignot, R. Kowk, J.C. Curlander, S.S. Pang. *«Automated Multisensor Registration: Requirements and Techniques»*. Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 8, August 1991. pp. 1029-1038.
11. P. Viola and W. M. Wells III. *«Alignment by maximization of mutual information»*. IJCV, 24(2):137–154, 1997.