

A NEW FEATURE BASED IMAGE REGISTRATION ALGORITHM

Karthik Krish
Stuart Heinrich
Wesley E. Snyder
Halil Cakir

Siamak Khorram

North Carolina State University
Raleigh, 27695

kkrish@ncsu.edu

sbheinri@ncsu.edu

wes@ncsu.edu

halil_cakir@ncsu.edu

khorram@ncsu.edu

ABSTRACT

This paper introduces a new feature- based image registration algorithm which registers images by finding rotation and scale invariant features and matches them using an evidence accumulation process based on the Generalized Hough Transform. Once feature correspondence has been established, the transformation parameters are then estimated using Non-linear least squares (NLLS) and the standard RANSAC (Random Sample Consensus) algorithm. The algorithm is evaluated under similarity transforms - translation, rotation and scale (zoom) and also under illumination changes.

INTRODUCTION

Image registration is the process of geometrically aligning two images taken at different times, orientation or sensors. Zitova et al.(Zitova , 2003) provides a comprehensive survey of existing image registration algorithms. According to them, there are essentially four steps to any image registration algorithm:

1. *Feature detection*: This involves finding salient features in the two images to be registered. An interest point detector is first employed to detect characteristic points in the image. These may include corner points, edges etc. Ideally, we want these points to be invariant to geometric and photometric transformations. A survey of the literature shows many approaches including the Harris Corner detector(Harris, 1988) and Scale Invariant Feature Transforms(SIFT) (Lowe, 1999),(Lowe, 2003). A complete review of affine region detectors can be found in (Mikolajczyk, 2005a).
2. *Feature matching* : Once features have been detected in the two images, the next step is to match them or establish correspondence. The common approach to feature matching is to build local descriptors around the feature point and then matching the descriptors. An exhaustive review of local descriptors for feature matching can be found in (Mikolajczyk, 2005b)
This is an important step because the percentage of correct matches identified here determines the how well the transformation can be estimated in the next step. Common matching methods include simple Euclidean distance matching, invariant moments and nearest neighbor based matching.
3. *Transform model estimation*: Once feature correspondence has been established, the next step is to solve for the parameters of some global transformation. Usually this involves finding the translation, rotation and scale parameters to transform one image to another.
4. *Image re-sampling and transformation*: The final step of any registration algorithm involves the actual mapping of the one image to the other using the transform model estimated in step 3.

This paper introduces a new feature-based image registration algorithm, which we will henceforth refer to as the SKS feature matching algorithm, which registers a target (or sensed) image to a reference (or source) image by finding rotation and scale invariant features and matching them using an evidence accumulation process based on the Generalized Hough Transform (Ballard, 1981).

FEATURE EXTRACTION

The first step in the image registration process is feature extraction. Features include salient points and distinctive objects in the image like closed contours, corners etc. We use the Harris-Laplace interest point detector in our algorithm. The Harris-Laplace detector uses the multi-scale Harris corner detector with the Laplacian operator for characteristic scale selection. Consider an image $I(\mathbf{x})$ where $\mathbf{x}=(x,y)$. The multi-scale Harris function (Mikolajczyk, 2001) is given by:

$$R(\mathbf{x}) = \text{Det}(C(\mathbf{x}, \sigma_D, \sigma_I)) - \alpha \text{Trace}(C(\mathbf{x}, \sigma_D, \sigma_I))$$

The matrix $C(\mathbf{x}, \sigma_D, \sigma_I)$ is defined as:

$$C(\mathbf{x}, \sigma_D, \sigma_I) = \sigma_D^2 G(\sigma_I) * \begin{bmatrix} I_x^2(\mathbf{x}, \sigma_D) & I_x(\mathbf{x}, \sigma_D) I_y(\mathbf{x}, \sigma_D) \\ I_x(\mathbf{x}, \sigma_D) I_y(\mathbf{x}, \sigma_D) & I_y^2(\mathbf{x}, \sigma_D) \end{bmatrix}$$

where $I_x(\mathbf{x}, \sigma_D)$ and $I_y(\mathbf{x}, \sigma_D)$ are the first derivatives of the image $I(\mathbf{x})$ along x and y and $G(\sigma)$ is a Gaussian kernel with standard deviation σ .

$$I_x(\mathbf{x}, \sigma_D) = I(\mathbf{x}) * \frac{\partial G(\sigma_D)}{\partial x}$$

$$I_y(\mathbf{x}, \sigma_D) = I(\mathbf{x}) * \frac{\partial G(\sigma_D)}{\partial y}$$

The Multi-scale Harris function $R(\mathbf{x})$ is positive in the corner regions, negative in the edge regions and small in flat or uniform regions. α is usually set in the range of 0.04 - 0.06. The spatial derivatives at a point will in general decrease with increasing scale. Therefore, the Harris function is normalized for scale to achieve scale invariance. This is done by normalizing the derivatives with the derivative scale σ_D . More information on scale normalization can be found in (Lindeberg, 1998).

Once the corner points are determined at a particular scale σ_I , we check if the point is at a characteristic scale. The characteristic scale is the scale at which a interest point is a local extremum across scale. Characteristic scales were extensively studied in (Lindeberg, 1998). Theoretically, the ratio of the scales for two corresponding points at the characteristic scales is equal to the scale factor of the two images. (Mikolajczyk, 2001) analyzed four different functions to determine the local maxima of a feature point across scales. These include the Laplacian, Difference of Gaussian(DOG), simple gradient and the Harris function. They concluded that the Laplacian detects the highest percentage of correct points which are at the characteristic scale. Therefore, we only retain corner points \mathbf{x} whose Laplacians are an extremum across scales.

$$L(\mathbf{x}, \sigma_n) > L(\mathbf{x}, \sigma_{n+1}) \wedge L(\mathbf{x}, \sigma_n) > L(\mathbf{x}, \sigma_{n-1})$$

where $L(\mathbf{x}, \sigma)$ is the Laplacian defined as:

$$L(\mathbf{x}, \sigma) = \sigma^2 |I_{xx}(\mathbf{x}, \sigma) + I_{yy}(\mathbf{x}, \sigma)|$$

By reducing the set of interest points to characteristic scale Harris corner points, we reduce the computational complexity of the problem without loss of scale invariance.

FEATURE MATCHING

Model Building

Once interest points have been detected in the images, the next step is to match them by building a descriptor. Consider a set of J interest points in the source image ${}^S F = \{{}^S F_1, {}^S F_2, \dots, {}^S F_j, \dots, {}^S F_J\}$, ${}^S F_j = (x, y)$. We first establish a rotation invariant coordinate system at each interest point ${}^S F_j$ by finding the dominant orientation (ϕ_j) in the circular neighborhood of radius corresponding to σ_j where σ_j is the scale at which the interest point was found.

We now build a model at each and every feature point $({}^S F_j)$ as follows. We formulate a set of K feature vectors $\mathbf{v}_{jk} = (r_{jk}, \theta_{jk}, \phi_{jk})$ at each point k in a circular radius corresponding to σ_j around ${}^S F_j$. (r_{jk}, θ_{jk}) are the polar coordinates of the point with respect to the invariant coordinate system at the interest point ${}^S F_j$. ϕ_{jk} is the difference between the gradient direction at that point k and the dominant orientation (ϕ_j) .

$$\phi_{jk} = \phi_k - \phi_j$$

The model at the feature point ${}^S F_j$ is given by:

$${}^S M_j(\mathbf{v}) = \max_{k=1}^K \exp\left(-\frac{\|\mathbf{v} - \mathbf{v}_{jk}\|^2}{2\sigma^2}\right)$$

The model function $({}^S M_j(\mathbf{v}))$ can be viewed as a function which estimates the closeness of a given feature vector (\mathbf{v}) to all the feature vectors around the interest point ${}^S F_j$. The model function can also be pre-computed and stored as a look up table which considerably speeds up the matching process.

Matching

The matching process uses evidence accumulation similar to the generalized Hough transform to determine the correspondence between two sets of interest points. Assume the source image(or reference image) has a set of J interest points ${}^S F = \{{}^S F_1, {}^S F_2, \dots, {}^S F_j, \dots, {}^S F_J\}$, ${}^S F_j = (x, y)$ and a corresponding set of models for each interest point $({}^S F_j)$. ${}^S M_j, j = 1 \dots J$. Let the target image(or sensed image) have M interest points ${}^T F = \{{}^T F_1, {}^T F_2, \dots, {}^T F_m, \dots, {}^T F_M\}$, ${}^T F_m = (x, y)$

Consider an interest point ${}^T F_m$ in the target image. To find the best point in the source similar to ${}^T F_m$, we first compute, similar to the model building process, L feature vectors $\mathbf{v}_{ml} = (r_{ml}, \theta_{ml}, \phi_{ml})$ in a circular radius corresponding to σ_m around the interest point. (r_{ml}, θ_{ml}) is the polar coordinates of the point with respect to the invariant coordinate system at the interest point ${}^T F_m$. ϕ_{ml} is the difference between the gradient direction at that point l and the dominant orientation (ϕ_m) at ${}^T F_m$.

Now, we compute a match between the feature interest ${}^T F_m$ in the target image and the model $({}^S M_j(\mathbf{v}))$ of the interest point ${}^S F_j$ in the source image,

$$A_{mj} = \frac{1}{L} \sum_{l=1}^L {}^S M_j(\mathbf{v}_{ml})$$

The best match for the point ${}^T F_m$ in the source image is given by:

$${}^S F_i = \arg \max_j A_{mj} \forall A_{mj} > T$$

where T is a user-defined threshold which determines the minimum match value between two points to consider them to be similar.

If no match is found, then it is assumed the interest point(${}^T F_m$) does not exist in the source image. Once we find a match, we do a reverse match by matching the source interest point(${}^S F_i$) to the target model built using ${}^T F_m$ to confirm correspondence.

Dominant Orientation Estimation

We find the dominant orientation at each interest point ${}^S F_j$ by building a histogram of the gradient directions in a circular neighborhood of radius corresponding to σ_j . σ_j is the scale of the interest point ${}^S F_j$.

Assume that the histogram has N bins and that the histogram is smoothed with a Gaussian window of width $N/3$. The dominant orientation (ϕ_j) is then taken as the highest peak of the histogram, which is the kernel density estimate of the maximum of the PDF of the local gradient orientations.

$$\phi_j = \arg \max_{n=1}^N h(\phi_n)$$

TRANSFORMATION MODEL

Once feature correspondence has been established, the next step is to transform the sensed image to the reference image. The transformation should ideally minimize some error function between all correspondence points. Since, the images must be overlapping for the algorithm to work, the focus point of the camera when taking both images can be assumed to be close by (globally), so distortion differences between the two images can be assumed to be low. Therefore, we assume a similarity transform between the two images - the model parameters are translation(tx, ty), rotation(θ) and scale(s) only.

Let $({}^S F_1, {}^T F_1), ({}^S F_2, {}^T F_2), \dots, ({}^S F_m, {}^T F_m)$ be a set of m point correspondences between the source and the target image. ${}^S F_m$ and ${}^T F_m$ are feature points in the source and target images respectively and are tuples of x and y . Let H be the transformation which maps $H : {}^T F_m \mapsto {}^S F_m$. We look for the model which minimizes the sum of squared error between the correspondences.

$$H(\theta, tx, ty, s) = \arg \min_{\theta, tx, ty, s} \sum_{i=1}^m ||H({}^T F_m) - {}^S F_m||$$

In the 1D case, we have a set of m equations from all the correspondences which can be solved which using Nonlinear Least Squares Regression(NLLS) (Weisstein, 2007). In the 2D case, we can solve this as a set of $2m$ equations, where each correspondence satisfies two separate equations in the x and y dimensions but uses the same model parameters.

To reduce the influence of outliers in the accuracy of the solution, we use RANSAC (Fischler, 1981) which involves finding the best model parameters by using random subsets from the set of correspondences instead of using the entire set. The model parameters estimated from the subset, which gives the least sum of squared error is taken as the best fit.

RESULTS

We test the performance of the algorithm on registering images by first using the feature matcher to determine correspondence points and then estimating the similarity transform using non-linear least squares regression (NLLS). We use a database of 8 synthetic image pairs (and some real pairs). Each image pair contains one sensed image with extreme rotation, scale and brightness changes, which we attempt to register with a reference image.

Some of the results of the registration are shown in Figures 1 and 2. Each figure shows the reference image, the sensed image and the difference image after registration. The results are visually perfect for all the tested images.

Notice that the difference image is not all black. This is because of the extreme brightness changes present in the sensed image.

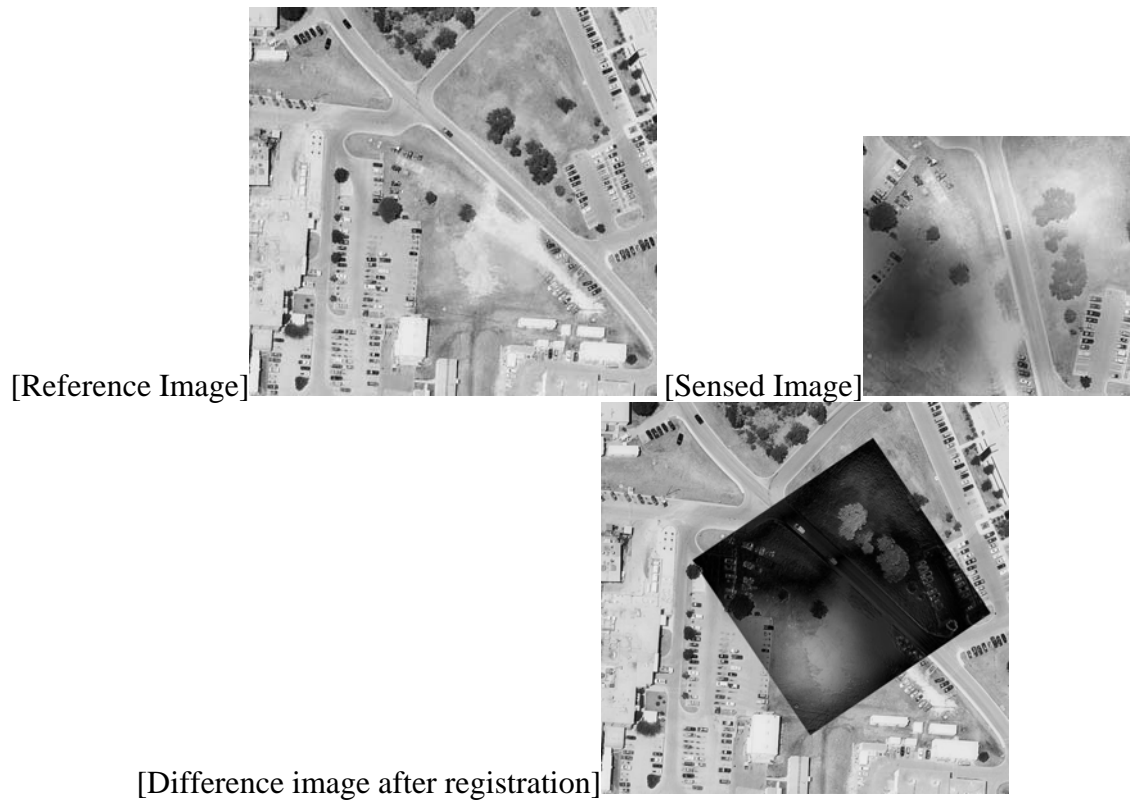


Figure 1. Registration Results.

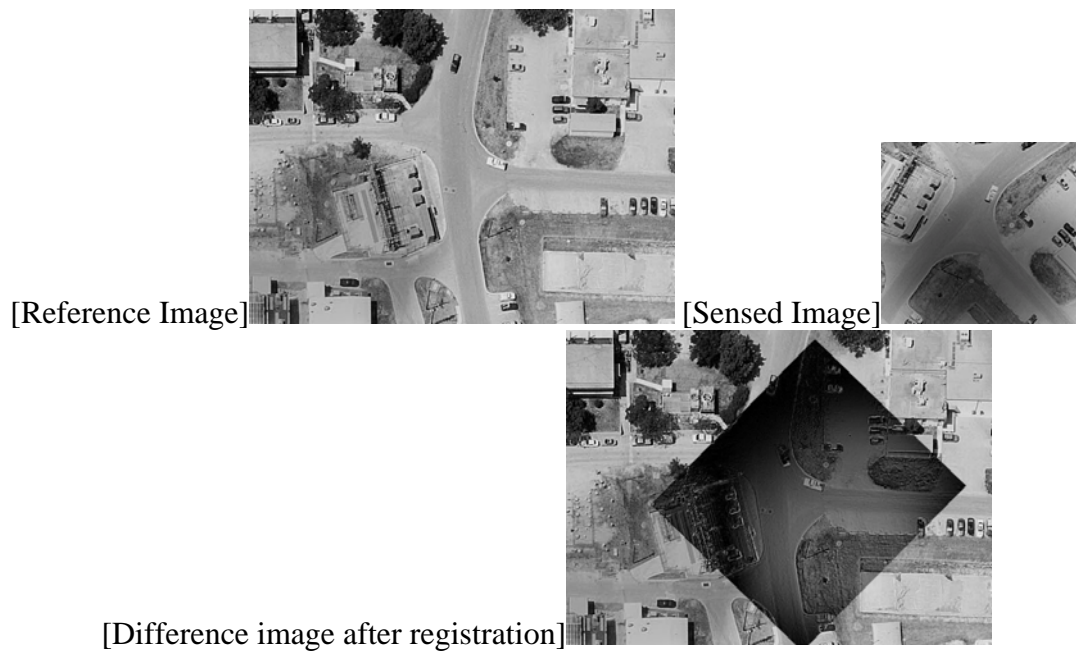


Figure 2. Registration Results.

CONCLUSION AND FUTURE WORK

We have introduced a new powerful feature matching algorithm which matches features using an evidence accumulation process and is invariant to translation, scale, rotation and illumination changes. The algorithm also outperforms conventional algorithms such as SIFT. The correspondence points provided by the feature matcher, was then used to estimate a similarity transformation to register two images with impressive results. Future work would involve extending the algorithm to model more complex transformations.

ACKNOWLEDGEMENT

This work was supported by the United States Air Force Research Office under grant no. FA9550-07-1-0176.

REFERENCES

- Barbara Zitova and Jan Flusser, 2003. Image registration methods: a survey, *Image and Vision Computing*, 21(11):977–1000.
- C. Harris and M.J. Stephens, 1988. A combined corner and edge detector, In *Proceedings of the 4th Alvey Vision Conference*, pages 147–152.
- C. Mikolajczyk, K.; Schmid, 2001. Indexing based on scale invariant interest points, *Computer Vision, 2001. ICCV 2001, Proceedings, Eighth IEEE International Conference on*, 1:525–531 vol.1.
- D. Lowe, 2003. Distinctive image features from scale-invariant keypoints, In *International Journal of Computer Vision*, volume 20, pages 91–110.
- D.H Ballard, 1981. Generalizing the hough transform to detect arbitrary shapes, *Pattern Recognition*, 13(2):111–122.
- David G. Lowe, 1999. Object recognition from local scale-invariant features, In *Proc. of the International Conference on Computer Vision ICCV, Corfu*, pages 1150–1157.
- Eric W. Weisstein, 2007. Nonlinear least squares fitting, <http://mathworld.wolfram.com/NonlinearLeastSquaresFitting.html>.
- K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, 2005a. A comparison of affine region detectors, *Int. J. Comput. Vision*, 65(1-2):43–72.
- Krystian Mikolajczyk and Cordelia Schmid, 2005b. A performance evaluation of local descriptors, *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630.
- M.A. Fischler and R.C. Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *ACM*, 24:381–395.
- Tony Lindeberg, 1998. Feature detection with automatic scale selection, *International Journal of Computer Vision*, 30(2):79–116.