Probabilistic image registration and anomaly detection by nonlinear warping

Author(s):
Kaynig, Verena; Fischer, Bernd; Buhmann, Joachim M.

Publication Date:
2012

Permanent Link:
https://doi.org/10.3929/ethz-a-006820379

Rights / License:
In Copyright - Non-Commercial Use Permitted
Probabilistic Image Registration and Anomaly Detection by Nonlinear Warping

Verena Kaynig Bernd Fischer Joachim M. Buhmann

Institute of Computational Science
ETH Zurich
8092 Zurich, Switzerland

Abstract

Automatic, defect tolerant registration of transmission electron microscopy (TEM) images poses an important and challenging problem for biomedical image analysis, e.g. in computational neuroanatomy. In this paper we demonstrate a fully automatic stitching and distortion correction method for TEM images and propose a probabilistic approach for image registration that implicitly detects image defects due to sample preparation and image acquisition. The approach uses a polynomial kernel expansion to estimate a non-linear image transformation based on intensities and spatial features. Corresponding points in the images are not determined beforehand, but they are estimated via an EM-algorithm during the registration process which is preferable in the case of (noisy) TEM images. Our registration model is successfully applied to two large image stacks of serial section TEM images acquired from brain tissue samples in a computational neuroanatomy project and shows significant improvement over existing image registration methods on these large datasets.

1 Introduction

The registration of images models the geometrical back transformation that maps each image to the next and to the previous image in a sequence, in order to resolve correspondences between images. In medical image analysis for example, image registration is used to identify anatomical structures of a patient according to a medical atlas, or to compare images from one patient taken at different times. In computer vision registration is required as a preprocessing step for motion tracking and 3D reconstruction. In this paper we address the registration of electron microscopy images for geometry extraction, which poses a challenging new task due to a low signal to noise ratio and structural changes between images.

In order to record a specimen with a transmission electron microscope (TEM), the specimen (conserved in a paraffin block) has to be cut into ultra thin sections of about 70 nm thickness. These sections are recorded with the microscope, stitched together, registered to form a 3D image stack and eventually segmented to gain a 3D reconstruction of the specimen.

Today most of this work involves time consuming manual labor, which renders the process very tedious for the microscopist as well as quite prone to errors [7].

In this paper we present an automatic approach for the registration related part of the 3D reconstruction pipeline. The most challenging part consists of the non-linear registration of images from single sections. Transmission electron microscopy with its special preparation of biological samples causes problems that are not solved by existing image registration methods. The problems are related to three major error sources,

(i) The electron beam of the microscope causes tissue damage in the specimen and, therefore, they can only be exposed to a limited electron dose. The image quality is rather poor, i.e. they have low contrast and a low signal to noise ratio.

(ii) In addition the preparation process often causes artifacts in the images that prevented an automation of the image processing pipeline so far. Examples of such artifacts are cuts in the samples caused
Institute of Computational Science, Technical Report Nr. 585

2 Stitching, Calibration and Correction of Nonlinear Distortions

In electron microscopy it is common practice to record a large region by taking several translated overlapping images. But up to today the resulting stitching problem is solved manually [7]. We follow an automatic landmark extraction approach based on SIFT features [11] to extract correspondence points from overlapping areas and solve the stitching problem automatically. SIFT features are calculated for each image to be stitched. Then a nearest neighbor search is used to find corresponding pairs of feature vectors, between images from the same section, that do not exceed a given distance threshold.

As can be seen in Figure 1, an affine transformation is not sufficient to stitch images that contain distortions caused by the electron microscope. A non-linear correction is required to stitch the images neatly together. Here we introduce a new method to estimate this non-linear transformation that can be used to correct the images. Note that the same non-linear transformation is applied to every single image for correction. Thus, only the image acquisition process can be the cause for this distortion.

Our approach estimates a rotation matrix $R^{(i)}$ and a translation vector $T^{(i)}$ for each single image $i$ and a non-linear distortion $\alpha$ that globally corrects all images. For the non-linear transformation we use an explicit polynomial kernel expansion to map the correspondence points into a higher dimension and use the transformation matrix $\alpha$ to project the points back into 2D space. The non-linear transformation should correct for errors of all correspondence points that cannot be corrected by the individual rigid transformations alone. This leads to the following optimization term:
Figure 1: Two example regions of the stitching intersection. In the top row without distortion correction the image border is clearly visible. In the second row the distortion correction makes a seamless stitching possible.

\[
\min_{\alpha, R, T} \sum_{i} \sum_{j} \left( X_{i,j}^{(i,j)} \alpha R^{(i)} + T^{(i)} \right) - \left( X_{j,i}^{(j,i)} \alpha R^{(j)} + T^{(j)} \right) \right\|^2 \\
+ \lambda \sum_{i} \left\| X_{i}^{(i)} - X_{0}^{(i)} \right\|^2
\]  

(1)

\( X_{i,j} \) denotes the higher dimensional correspondence features between image \( i \) and image \( j \) and \( X_{0}^{(i)} \) denotes the original image points of image \( i \). If there is no distortion present in the images, the first term of the sum is invariant to \( \alpha \), and the preference for the natural choice of \( \alpha \) being the identity transformation is expressed by the second term. To avoid a strong bias, the regularization weight \( \lambda \) should be set to a rather small value.

In practice the term can be optimized by iteratively obtaining the rigid transformations \( T^{(i)} \) and \( R^{(i)} \) and then the distortion matrix \( \alpha \). So far however, one or two iterations were sufficient to obtain \( \alpha \) for all our image sets.

Without any distortion correction the median error of the affine stitching was 15.17 pixels. Our experiments show that with a distortion correction based on a 5 fold overlap the error drops below 2 pixels.

Figure 1 shows a stitching of distorted images before and after the calibration. Figure 2 depicts a needle diagram of the non-linear distortion applied to all images. As can be seen, the distortion is fairly small in the center of the image and becomes larger at the image borders.

3 Expectation-Maximization

After the preprocessing steps, the images now have to be aligned into one image stack. Due to cutting errors (cracks or folds), different tissue stress during the preparation and variations in slice thickness a non-linear transformation model is necessary to register the images in detail. Furthermore, due to the difficult preparation process, a lot of images show artefacts with high contrast caused by staining errors, that lack correspondences in the target image. Our aim is to exclude points located on these anomalous structures from the matching process.

The problem is modeled as a mixture model where we not only mark these non-relevant structures, but we also solve the whole correspondence problem during the registration process itself. The optimization is achieved by the expectation-maximization algorithm.

In a Bayesian framework the optimal transformation matrix \( \beta \) maximizes the posterior probability

\[
p(\beta|X, Y) = \sum_{M \in M} \frac{p(X, Y|\beta, M) \cdot p(\beta) \cdot p(M)}{p(X, Y)}
\]  

(2)
where $X$ and $Y$ correspond to the warp image and the target image. The variable $M$ denotes a binary correspondence matrix. $M_{ij}$ is one, if point $x_i$ in the warp image corresponds to point $y_j$ in the target image and zero otherwise. In addition a point $x_i$ can be assigned to an outlier class, denoted as $M_{ij}(n_2+1)$. Thus, the whole matrix $M$ is of size $n_1 \times (n_2+1)$ where $n_1$ is the number of chosen points in the warp image and $n_2$ is the number of possible correspondences for each of these points in the target image. The set $\mathcal{M} = \{(0,1)^{n_1(n_2+1)}\}$ denotes all admissible assignment matrices $M$.

Application specific constraints on the assignments can be modeled by an appropriate definition of $\mathcal{M}$. As each point $x_i$ should be assigned to only one correspondence point $y_j$, we define $p(M)$ to be zero for configurations that assign more than one correspondence point to $x_i$ and assume all valid configurations to be uniformly distributed:

$$p(M) \sim \begin{cases} 1 & \text{if } \sum_{j=1}^{n_2+1} M_{ij} = 1 \text{ for all } i = 1, \ldots, n_1 \\ 0 & \text{else} \end{cases}$$

This definition of $p(M)$ ensures that each point $x_i$ is assigned exactly to one correspondence point or is marked as not relevant.

To define $p(\beta)$ the dimensions of the solution vector $\beta$ are assumed to be normally distributed. Therefore, we introduce a ridge penalty which is described by the normal distribution $\mathcal{N}(\mu,\sigma)$ with location parameter $\mu$ and variance $\sigma^2$ as prior distribution

$$p(\beta) = \prod_{i=1}^{n_\beta} \mathcal{N}(\beta_i; \mu, \sigma^2)$$

for $\beta$. The parameter $\lambda$ is the ridge penalty that controls the complexity of the regression function.

The distribution of $p(X,Y|\beta,M)$ should depend on the similarity of the correspondence points based on gray values, as well as on the quality of the geometric fit. Furthermore, we need to take care about the image anomalies that are assigned to the outlier class. This outlier class is modeled as a uniform distribution. The complete data likelihood can be written as

$$p(X,Y|\beta,M) \sim \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} \left( \varphi_{0,\sigma_1}(v(x_i) - v(y_j)) \right) \cdot 
\varphi_{0,\sigma_2}(\phi(x_i) \beta - y_j) \cdot M_{ij} \cdot \prod_{i=1}^{n_1} \left( \varphi_{0,\sigma_1}(c \sigma_1) \right)^{M_i(n_2+1)}$$

Here $v(x_i)$ is a vector of the pixel values of a small patch centered at $x_i$. The difference of two such patches serves as a dissimilarity measure that is easy to compute and takes context information about a small area around the points into account.

The function $\phi(x_i)$ describes the explicit polynomial kernel expansion of the points $x_i$:

$$\phi(x_i) = [1, x_{i1}, x_{i2}, x_{i1}^2, x_{i1} x_{i2}, x_{i2}^2, \ldots, x_{i2d}]^T$$

The transformation matrix $\beta$ then projects these points back into the image plane, leading to a nonlinear transformation of the image $X$.

The second factor provides a penalty for points that are marked as not relevant. The constant $c$ is given as the 0.95 quantile of the cumulative chi square distribution, where the degrees of freedom in principle correspond to the number of pixels in $v(x_i) - v(y_j)$. To reduce the influence of noise in the similarity measure, we perform a principal component analysis and project the high dimensional difference vectors down to the eigenvectors that correspond to the largest 98% of the eigenvalues. This also reduces the degree of freedom of the cumulative chi square distribution. Thus, the outlier factor transforms the assumed normal distribution of the gray value similarity into a heavy tailed distribution, providing a robust solution for outliers caused by not relevant elements.

In order to maximize $p(\beta|X,Y)$ which yields an optimized transformation, we maximize the logarithm of $p(X,Y|\beta,M) \cdot p(\beta)$ under the condition that $M$ is a valid matrix, i.e. $\sum_{j=1}^{n_2+1} M_{ij} = 1$ for all $i = 1, \ldots, n_1$. Since the assignment variables $M$ are unobservable, we use the EM-algorithm to maximize the joint log-posterior. The algorithm iterates between estimating the expectation of the latent variables $M_{ij}$ while keeping $\beta$ fix and maximizing the joint log-posterior while keeping the expectation values of $M$ constant. The variances for the normal distributions are also calculated during
the maximization step. The log-posterior is maximized with respect to the transformation $\beta$ as well as the variances of the normal distributions $\sigma_{1}$ and $\sigma_{2}$.

**E-step:** In each iteration the variables $M_{ij}$ are replaced by their conditional expectation given $\beta$. The expectation values are calculated using the current optimized $\beta$. Under the condition that $M$ is a valid assignment matrix, we derive the following result:

$$
\gamma_{ij} = \mathbb{E}[M_{ij}|X, Y, \beta] = \frac{p(X, Y|\beta, M_{ij} = 1)}{\sum_{l=1}^{n_{2}+1} p(X, Y|\beta, M_{il} = 1)}
$$

(7)

**M-step:** The expectation of the joint log posterior has the same form as the joint log posterior itself, but with $M_{ij}$ replaced by $\gamma_{ij}$. The parameter $\beta, \sigma_{1}, \sigma_{2}$ are then computed by maximizing the expectation of the joint log posterior.

For the transformation $\beta$ this MAP approach yields a ridge regression solution [5] with weights $\gamma_{ij}$

$$
\beta \leftarrow (\phi(\bar{X})^{T}\Gamma\phi(\bar{X}) + 2\lambda I)^{-1}\phi(\bar{X})^{T}\Gamma Y
$$

(8)

where $\Gamma$ is a $(n_{1} \cdot n_{2}) \times (n_{1} \cdot n_{2})$-dimensional diagonal matrix of the weights $\gamma_{ij}$. The $(n_{1} \cdot n_{2}) \times 2$ matrix $\bar{X}$ contains $n_{2}$ copies of each position vector $x_{i}$ and the $(n_{1} \cdot n_{2}) \times 2$ matrix $Y$ contains $n_{2}$ possible correspondence points for each position $x_{i}$. Note that the parameter $\lambda$ in Eq. 8 also appears in the prior distribution $p(\beta)$ (Eq. 4). In our experiments $\lambda = 0.001$ sufficiently regularized the assignments.

The standard deviations are updated by

$$
\sigma_{1} \leftarrow \frac{\sum_{i=1}^{n_{1}} \sum_{j=1}^{n_{2}} \gamma_{ij} \cdot p(x_{i}, y_{j})^{2}}{\sum_{i=1}^{n_{1}} \sum_{j=1}^{n_{2}} \gamma_{ij}}
$$

(9)

$$
\sigma_{2} \leftarrow \frac{\sum_{i=1}^{n_{1}} \sum_{j=1}^{n_{2}} \gamma_{ij} \cdot ||\phi(x_{i})\beta - y_{j}||^{2}}{\sum_{i=1}^{n_{1}} \sum_{j=1}^{n_{2}} \gamma_{ij}}
$$

(10)

where $\sigma_{1}$ and $\sigma_{2}$ are invariant to outliers since $\gamma_{i,n_{2}+1} \approx 1$ for these points.

**Choice of initial points $x_{i}$** So far all warp points $x_{i}$ are assumed to be arranged on a regular grid. While this design ensures that all interesting structures in the image are covered by a warp point nearby, interest points are often placed in background areas. To increase precision we would like to position each point directly in content rich parts of the image while still covering all biologically relevant structure in the image. For this purpose we calculate the entropy of the intensity value in a $13^{2}$ neighborhood around each point in the image. The entropy is high for pixels along structures with a high contrast. Now we shift each warp point $x_{i}$ of the regular grid to the position with the highest entropy value in its neighborhood. This local adaptation method preserves the coverage of the whole image while emphasizing areas with rich image content.

### 4 Data and Experiments

For our experiments we used images gathered in a computational neuroscience project. When imaging with a TEM it must be possible for single electrons to penetrate the probe. Therefore the specimen is first stained, then embedded into resin and cut into ultra thin sections of $70 \text{ nm}$ thickness.

During this whole process the three major steps that may cause artefacts in the image are the staining, the cutting and the recording with the electron beam. Staining may produce additional dark areas in the image that do not correspond to original biological structures. During the cutting process the slice is exposed to significant stress and it may be non-linearly transformed or it even can encounter fractures. Finally, exposure to an electron beam causes a mass loss of the specimen and leads to additional transformations.

Figure 3 depicts an example image. On the left side is an image of typical quality. The structure in the left bottom quarter of the image with the pike on the right hand side shows a dendrite. The smaller ellipse like structures over the image are myelinated axons. The four smaller images on the right show examples of image artefacts, that are caused by the sample preparation: In the lower right corner we see an example of a crack in the specimen caused by the cutting process. The dark spots in the upper left image and the dark stripes in the upper right and lower left image are artefacts of the staining procedure.

We tested our approach on two series of electron microscopy images. The first series contains 97 images that were taken at $3400x$ magnification with a resolution of $1032x1376$ pixels, one image per section. The second series consists of 284 images taken at $13500x$ magnification that were distortion corrected and stitched into 71 section images. The resolution of the second stack is $2672x4008$ pixels.
Figure 3: Examples of experimental data. On the left an image of typical contrast without major artefacts. The scale bar corresponds to 4µm. On the right side some examples of artefacts caused by the preparation process.

Figure 4: Difference images for (left) affine transformation, (middle) least squares matching with polynomial basis functions, (right) expectation-maximization including visibility estimation. The original images are shown in Figure 5.
In a preprocessing step we correct the radial illumination gradient visible in the images by dividing each image with a smoothed version of itself ($\sigma = 30$ pixel). Then an initial affine transformation based on SIFT features is estimated for each image.

This transformation is then refined with our warping approach based on expectation maximization. For comparison we also provide the refinements for ridge regression with the ordinary least squares error

$$\min_{\beta} E(\beta) = \min_{\beta} \sum_{i=1}^{n} \|\phi(x_i)^T \beta - y_i\|^2 + \lambda \|\beta\|^2. \quad (11)$$

and the robust version, where the Huber loss function is used instead of the least squares error

$$L_c(\xi) = \begin{cases} c|\xi| - \frac{c^2}{2} & \text{for } |\xi| > c, \\ \frac{\xi^2}{2} & \text{for } |\xi| \leq c. \end{cases}$$

As these methods need fixed corresponding points, we calculate the correlation coefficient of a patch around $x_i$ and the appropriate patches of the target image. The result is then weighted with a Gaussian density centered at the position of $x_i$ and a standard deviation of two times the standard error of the affine match. Each mapping point $x_i$ is then assigned its correspondence point $y_i$ according to the maximum of the obtained function.

To provide a visual impression of the obtained warpings we first show difference images for the affine initialization and two warpings (Figure 4). The original images are shown in Figure 5. The darker the color in the difference image, the larger the absolute difference in gray values between the warped image and the target image. The images in the left column show the difference map for the robust affine transformation, the images in the middle the result for ridge regression with polynomial basis functions. The third column shows the difference maps for our new expectation-maximization method. If one focuses on the images in the upper row on the top left region as well as on the right and bottom border, one can observe dark stripes in the affine transformation image that are getting thinner for the least squares and even more thinner for the expectation maximization method. This error measure shows clearly that important edges are not matched very well by the affine transformation, but for the expectation maximization solution, there are only very small differences left over. Edges in the image are fitting significantly better than for the standard technique. In addition we tried to compare our method to optical flow methods [1, 2]. We could not find parameter settings that were able to cope with the significant structural changes between sections.

Since the detection and localization of image anomalies is important for our method, we show examples of the estimated image anomaly regions in Figure 5. The darker the color in this image, the more likely the region belongs to an artefact. The dark strip over the upper image as well as the dark blobs in the lower image are clearly detected as not relevant structures which is our desired goal. The information about these anomalies can now be used for further processing steps in computational neuroanatomy, e.g. in the 3D reconstruction of the neural connectivity structure. To visualize the estimated non-linear transformation of the image, we show a needle diagram of the transformation in Figure 6.

To demonstrate the improvement of the new method, we have registered both stacks of TEM images from the above described neuroanatomy project and measured the cross-correlation between the target image and the warped image. Figure 7 shows the results for the two image stacks. Our EM approach has been initialized with the robust ridge regression solution. Therefore, improvements in cross-correlation values over the ro-
Figure 5: Left: The first image, Middle: The visibility map (Detected image anomalies). Right: The second image.

Figure 7: Comparison of EM warping (solid line) to ridge regression (dotted line) with respect to robust Huber loss estimation (normalized to zero). Above the distortion corrected and stitched image stack containing 71 images, below the single image stack containing 97 images.
bust ridge regression solution serves as a measure of success for our model of image registration, i.e., large differences in cross-correlations denote a significantly better registration of the TEM images than with the robust Huber loss. The EM method clearly outperforms the other techniques.

5 Conclusion

Registration of images is an important step in the 3D reconstruction. Especially in Transmission Electron Microscopy of biological samples, image anomalies occur frequently caused by the sample preparation process and by the image acquisition process. In this paper we propose a novel method for image registration that jointly estimates image anomalies and an image matching in a Bayesian model. The mixture model enables us to estimate assignment probabilities as well as probabilities for damages. The method performs superior to standard methods like linear affine transformation, and non-linear transformation. Even state-of-the-art outlier detection methods are inferior in performance compared to our mixture model. The expectation-maximization algorithm optimizes the model efficiently and is straightforward to implement. The experiments convincingly demonstrate that the model not only improves the image registration process, but that it also detects image anomalies. This quality of the algorithm will prove to be important for further steps in the automation of the microscopists tasks in image processing, like the 3D reconstruction of the neural cortex, including all synaptic connections between neurons.

Acknowledgement

We like to thank Nuno Maçarico da Costa and Kevan Martin, Institute of Neuroinformatics UNI-ETH Zurich, for valuable discussions on neuroanatomy and for providing the TEM images.

References