

Point-based registration of high-resolution histological slices for navigation purposes in virtual microscopy

Peter Leškovský¹, Alexander Alekseychuk¹, Adam Stanski¹, Olaf Hellwich¹, Karsten Schlüns², Norman Zerbe², Peter Hufnagl²

¹Computer Vision & Remote Sensing, Technical University Berlin, Berlin, Germany

²Dept. Digital Pathology and IT, Institute of Pathology, Charité - Universitätsmedizin Berlin, Berlin, Germany

Abstract

Automatic registration methods for histological images in virtual microscopy are presented. Point-based approaches with outlier filtering by RANSAC and multibody factorization are investigated. The algorithms are tested on high-resolution biopsy images of the gastrointestinal tract. Quantitative evaluation is performed at different resolution levels. The results reveal the superiority of Harris feature points over SIFT features, in our setting. At low resolutions or when only subsequent slices are aligned, favorable registration results are obtained using only raw feature matching and Delaunay-based interpolation of the transformation parameters. Outlier filtering is necessary to obtain stable results in more difficult scenarios, *i.e.* when registering nonconsecutive slices.

1 Introduction

Virtual microscopy describes the examination of microscopic structures on digitized images. With last generation optical scanners, taking less than two minutes for the acquisition of standard histological slides ($6 \times 2 \text{ cm}^2$) with 40-fold magnification, the digitization settles an important role in diagnostic pathology. It avoids problems with proper storage of tissue samples and promises easy sharing of histological samples, now in digitized form, among the medical doctors over long distances. It also allows for editing annotations and marking corresponding regions among several slides. Such placemarks make the navigation over or between histological slides possible. The navigation feature is important during detailed specimen diagnosis where examination between several sections is often unavoidable, *e.g.* when tracking the propagation of malignant tumors.

In histological pathology, the tissue samples are cut into thin slices during the preparation phase. These are then individually mounted on glass slides. At this stage, any correspondence information between slices is lost. Using optical microscope the navigation between slices becomes exhausting. Numberless manual shifts of the glass slides and visual verification at different magnifications are necessary in order to identify the corresponding cell groups in subsequent slices. In virtual microscopy, the use of digitized images suggests possible automation of the time demanding manual alignment of different slices in that the correspondence information is precomputed and stored. Navigating between slices is then done only by a button click. The main challenge here, however, is the generation of a dense set of correspondences across the tissue sections. In this work we therefore suggest and evaluate automatic ways to extract correspondence points between histological slides. Our approach is based on registration algorithms optimized for histological images of high-resolution.

2 Related Work

The main issue when solving a registration task is the proper selection of geometric transformation function and similarity measure. Large variety of transformation functions, ranging from rigid to deformable models, has been adopted for the registration of medical data. Although non-linear models provide more degrees of freedom, and thus can better describe the underlying tissue distortions, they easily introduce spurious deformations due to noise, initial misalignment or present outliers. Special care to regularization of the deformation field must be taken (see [Arganda-Carreras et al., 2006, Sorzano et al., 2005] for B-Spline, [du Bois d'Aische et al., 2005] for Finite Element Method and [Yang et al., 2011] for Thin Plate Splines regularization approaches). Another negative characteristics of non-rigid registrations is that the truth of the estimated deformation cannot be granted. In histological diagnosis, where the shape of fine features (*e.g.* cells or nuclei) plays an important role, this would pose a problem for future analysis. The non-linear models are therefore mostly used only for an informative visualization and not for histological diagnosis.

Deformations introduced during the sample preparation can be well described locally by rigid transformation models. During the specimen preparation, several samples are bound into one block by embedding them in paraffin. When paraffin is cool and hardened, the specimen is cut in slices. This procedure limits the extent of deformations applied on the tissue when cut. Only little stretching and bending is observed. The sections rather tear and break into small pieces which can be shifted arbitrarily on the glass slide. Therefore, the deformation can be locally well described by rigid models even at high resolutions. In [Cooper et al., 2009] the authors rigidly register small neighborhoods of histological structures, folices or blood vessels, on images captured at $40\times$ magnification. Polynomial refinement is applied posteriorly to interpolate the deformation over the whole image. Similarly, in [Pitiot et al., 2006] block matching with associated rigid transformation model is used to obtain initial registration which is then extended to a piecewise affine model.

The understanding of the underlying deformations being small has also been noted by histologists in [Pitiot et al., 2006]. They suggest to consider the whole slice as a set of independent components subject to linear transformation. We therefore apply registration schemes which assume piecewise rigid transformation and linear interpolation between the aligned regions. Such approach should handle cases where tissue parts are independently

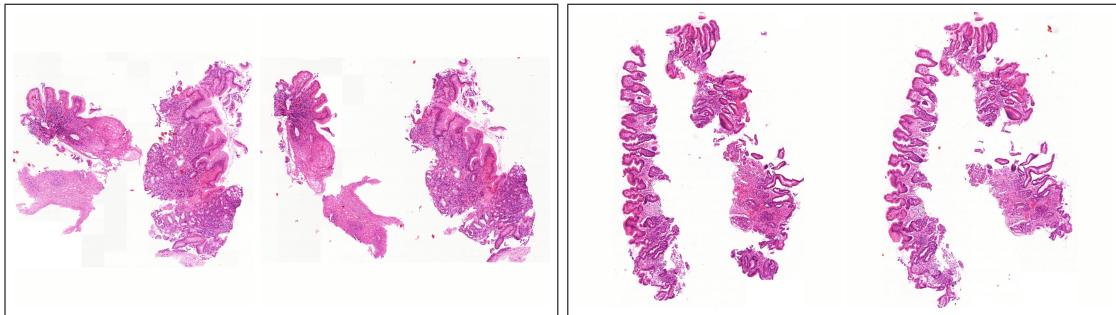


Figure 1: Examples of input images, presenting two cases with two slices each. In each slice, several independently displaced pieces of tissue, repeating across the sections, can be identified.

shifted within one slice (see *Fig. 1, left*) and also be able to partially compensate for local deformations of elongated parts (see *Fig. 1, right*).

The registration task can be driven by image or feature-based techniques. The former ones use differences in image intensities within overlapping regions to improve their alignment. Simple pixelwise mean-square difference was applied by [Arganda-Carreras et al., 2006]. [Pitiot et al., 2006] used constrained cross-correlation similarity measure for the optimization of the deformation field. A constrained term was added to compensate for the affine intensity changes between the slides. Note that their intensity-based registration was run only locally, within a block-matching scheme. Mutual information is a well applicable approach when registering multi-modal input data, such as CT with histological slices or even differently stained histological images. [du Bois d'Aische et al., 2005] use this measure to align H&E stained histological images with photographed macroscopic images of the specimen.

The disadvantage of using an image intensity-based registration method is that they require a good initial overlap of the region of interest. Otherwise the optimization part will hardly converge, as the similarity is always computed from the current alignment. In general it can not be assumed that such alignment, of all tissue parts, can be easily obtained for histological data. This is due to the arbitrary shifts of the torn pieces of tissue. When comparing different slices the separate tissue pieces often touch each other at different places or even overlap. This impedes their segmentation and straightforward identification based on their shape or regional statistics.

No such initialization is necessary for the feature-based methods. They align the images using salient point or local structure correspondences. The position of each feature is usually robustly extracted from the images and statistical description of only a small surrounding region is considered during matching. Salient points [Yang et al., 2011], contours [Cohen et al., 1998] or segmented histological structures of interest [Cooper et al., 2009] can be used as the registration features. In our implementation we concentrate on point-based methods. We test several general salient point detectors (see [Tuytelaars and Mikolajczyk, 2008]) as their sensitivity depends on the input image data. Nevertheless, for the matching task we only selected the SIFT descriptor [Lowe, 2004] and did not experiment with different options.

The point-based registration methods are sensitive to outliers arising during the matching of salient points. For a successful setup of such task it is therefore important to select a stable feature detector, a robust feature descriptor, to determine the best resolution at which

features should be extracted and to answer how or whether outliers in the correspondence set have to be detected. Our preliminary results have been presented in [Leškovský and Hellwich, 2011]. In this paper we additionally provide more detailed explanation of our registration technique, suggest possible automatic parameter selection approaches applied to our algorithms and extend the evaluation analysis to registrations of nonconsecutive slices.

3 Registration Methods

3.1 Feature Point Detectors

We experiment with SIFT [Lowe, 2004] and structure tensor based Harris [Tuytelaars and Mikolajczyk, 2008] feature detectors to extract salient points for the point-based registration task. While the first one detects round, blob-like structures, the latter one has better response to distinctive corner-like regions. For an informative illustration, the different behavior of the detectors can be followed in *Fig. 2*, where all feature points in the image have been extracted at a single resolution scale.

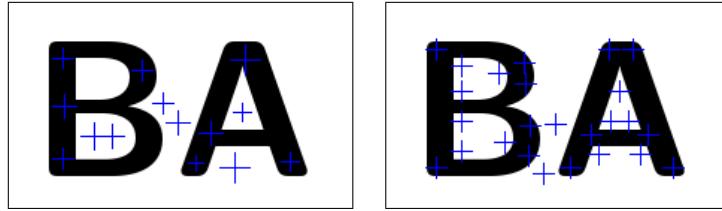


Figure 2: Salient point detection using SIFT (left) and Harris (right) feature points, for the image of letters "BA". Different response of the proposed detectors can be observed.

The SIFT detector is based on the Difference of Gaussians (DoG) operator:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y).$$

Here, $I(x, y)$ represents the 2D image which is convolved in both directions with a variable-scale Gaussian:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma} e^{-(x^2+y^2)/2\sigma^2}.$$

The parameter k is a multiplicative factor separating two scales. The resulting detector is run over multiple scales, and the maximum response is searched in the obtained scale space. Since the DoG operator is scale normalized [Lowe, 2004], no further adaptations are necessary.

Harris feature point detector (see also [Tuytelaars and Mikolajczyk, 2008]) is based on the local autocorrelation matrix, also known as the structure tensor:

$$\mathbf{A} = W(x, y) \begin{bmatrix} I_x(x, y)^2 & I_x(x, y)I_y(x, y) \\ I_x(x, y)I_y(x, y) & I_y(x, y)^2 \end{bmatrix},$$

where $I_x = \frac{\partial}{\partial x} I(x, y)$ and $I_y = \frac{\partial}{\partial y} I(x, y)$ represent the partial derivatives of the image intensities in the direction x and y , respectively, and W is a weighting window smoothing the results. We apply Gaussian-based weighting $W(x, y) = G(x, y, \sigma_W)$. To adapt the tensor

computation for the scale space the image data is convolved with the Gaussian and appropriately scaled:

$$\mathbf{A}_S = G(x, y, \sigma_W) \sigma_S^2 \begin{bmatrix} L_x(x, y, \sigma_S)^2 & L_x(x, y, \sigma_S)L_y(x, y, \sigma_S) \\ L_x(x, y)L_y(x, y, \sigma_S) & L_y(x, y, \sigma_S)^2 \end{bmatrix},$$

where L_x and L_y represent the derivatives of the image intensities convolved with a Gaussian kernel function:

$$\begin{aligned} L_x(x, y, \sigma) &= G(x, y, \sigma) * I_x(x, y) = \frac{\partial}{\partial x} G(x, y, \sigma) * I(x, y), \\ L_y(x, y, \sigma) &= G(x, y, \sigma) * I_y(x, y) = \frac{\partial}{\partial y} G(x, y, \sigma) * I(x, y), \end{aligned}$$

for $\frac{\partial}{\partial x} G(x, y, \sigma)$ and $\frac{\partial}{\partial y} G(x, y, \sigma)$ representing the derivative of Gaussian in x and y direction, respectively, and σ being the scale space specific constant. The cornerness of the local image structure can be described with respect to the eigenvalues λ_1 and λ_2 of the structure tensor. If both are high, large derivations in both principal directions are observed, corresponding thus to a corner like feature. Harris described the cornerness of a feature by a value

$$M_H = \lambda_1 \lambda_2 - \alpha (\lambda_1 + \lambda_2)^2 = \det(\mathbf{A}) - \alpha \text{trace}(\mathbf{A})^2.$$

The parameter α controls the detector's ability to discern between corners and edges. To extract the position of the Harris feature points we apply non-maximum suppression at each scale σ_S of our scale space.

3.2 Feature Point Description and Matching

We have chosen to use the SIFT descriptor [Lowe, 2004] to robustly characterize each feature point. It yields statistics based on the local gradient image. An 128 bin histogram, captured at the most responsive scale of the scale space, is obtained for each point.

We get the initial correspondences similar as in [Lowe, 2004], by coupling the nearest points in the descriptor's space and filtering the putative matches based on the ratio between the nearest neighbor and the second nearest one to the query point. This is known as the nearest neighbor distance ratio, NNDR, matching. Note that our images have been acquired at the same zoom factor. The corresponding image structures therefore have to be of the same absolute size. This allows us to restrict the matching algorithm to consider only descriptors of the same scale. At last, we enforce bijective matching between the point sets, to ensure consistency between the matches. Therefore, if the query point is not as well, respectively, the nearest neighbor of the inspected matching candidate, we discard the possible correspondence.

3.3 Outlier Detection

The NNDR matching process does not guarantee that all matches will be correct as it evaluates the matching similarity upon a local descriptor, *i.e.* comparing only a small neighborhood of the feature points. To exclude mismatches, the spacial relations among the feature points have to be taken into account.

We apply two outlier removal methods: RANSAC [Fischler and Bolles, 1981] and multi-body factorization (MBF) [Subbarao and Meer, 2009]. They both filter the matches by examining their correspondence to the object's geometry. With regard to our assumption of a

piecewise rigid transformation describing the deformations of the sample (see Sec. 2), we can expect that groups of corresponding points will follow single rigid transformation (rotation and translation). These should model the transformation of the torn pieces of tissue.

The RANSAC algorithm identifies the largest group of pair correspondences which rigidly map onto each other. In general, it first samples several models. Each of them is represented by a minimum number of points which uniquely define our elemental transformation, *i.e.* two correspondence pairs for rigid transformations. The points, defining each sampled model, are randomly selected from the correspondence set. The quality of each model is then assessed by the number of all matches which follow the model’s transformation. A maximum transformation error must be specified according to which the matches can be assigned to inliers and outliers. Considering one sample model we estimate its expected transformation error by a non-parametric probability density function estimator, CEPEL [Stanski, 2012]. It evaluates the error distribution of each correspondence pair, subject to the given transformation, and returns the characteristics of its first mode. All correspondences with errors smaller than the first mode adjusted by its variance are considered as inliers. The rest are outliers. Finally, we select the best model by choosing the smallest value of the corresponding transformation error as well as the largest number of assigned inliers.

As described so far, the RANSAC procedure outputs a single most prominent rigid transformation model. The outliers of this model contain wrong matches along with true correspondences which describe the transformation between other tissue parts. We thus recursively apply the RANSAC algorithm to the outliers set until a maximally admitted transformation error is reached. This termination parameter is the only constant which we have to specify in advance. With this setting, the extraction of the transformation models follows a greedy approach.

The random sampling as well as the greedy extraction of the transformation models is not optimal. The MBF method extracts all groups of rigidly matching point sets in parallel. No iterative extraction is thus necessary. A mean shift algorithm in the space of transformations is used. As data points we use a set of elemental rigid transformation functions. Similar to RANSAC, two correspondence pairs could be used to define the translation and rotation of each elemental transform. Nevertheless, we obtained better results when estimating the transformations from a single correspondence pair, using the SIFT angles to define rotation. Note that the transformation space, for rigid transformations, is not linear. We therefore follow the nonlinear version of the mean shift algorithm described in [Subbarao and Meer, 2009]. The algorithm is controlled by a bandwidth constant which defines the support size of the mean shift kernel function. A Gaussian kernel function is used. Finally, we select only those transformations for which the number of converging elemental models is bigger than five. This constant is set empirically.

3.4 Deformation Field Definition

Both, the matching and the outlier filtering algorithms output groups of point matches which are considered as true. We use these true matches to define the deformation field over the whole slice. Although for the outlier methods we also obtain a group of rigid transformations which represent the geometrical changes of the separate tissue parts, applying only this limited set of transformations is often not sufficient. It is because the rigid transformation assumption is not true in reality. Small deformations within each tissue piece exist. To compensate for these, we rather directly interpolate between the mappings (translation and

rotation) defined by each point match. Remember that each correspondence pair defines one rigid transformation when using the SIFT angle of the feature point to define rotation. Linear interpolation over a Delaunay triangle mesh, defined by the feature points, provides acceptable results. In addition, we apply a mask of the imaged tissue to avoid interpolation between mesh vertices which belong to clearly distinct tissue pieces. A direct interpolation between these points would lead to strong distortions at the tissue edges, especially because the tissue edges are never densely covered by the filtered feature points. The necessary mask is generated automatically using a simple triangle thresholding method [Yoo et al., 2002]. The Delaunay triangles which overlap with the excluded image part are first left out from the deformation field computation. Afterwards, at these places the deformation is extrapolated from the nearest feature points.

4 Evaluation

The evaluation has been run on a set of approximately 150 histological slices, containing biopsies of the gastrointestinal tract. 22 specimen dyed with three stains (H&E, GI and PAS) have been included in the study. Only intra-stain registrations have been examined.

The samples were optically scanned at resolution of $0.23\mu m$ per pixel, resulting in images of 500 mega-pixel size. After acquisition, an image pyramid of eight coarser resolutions was created by successively downsampling the images by a factor of two. The image levels range from values 0 (finest resolution) to level 7 where each pixel corresponds to a region of $2^7 \times 2^7$ pixels at the highest resolution.

4.1 Ground Truth Data

We use point correspondences as ground truth (GT) data. Approximately 25 point matches have been defined manually between every couple of subsequent histological slices. We selected them at image resolution level 4. Note that already at this resolution it is not possible to obtain true pixel to pixel correspondences. This fact is obvious if one considers that each tissue slice consists of different cells. Pixelwise correspondences are therefore not existing. Nevertheless, the true position of matches can be approximated better when considering larger histological structures. On subsequent slices they appear to be only slightly deformed. To increase the precision of the manually set GT points, but also to obtain the parameters of the rigid transformation, we run local image-based registration at these points. Patches of size 50×50 pixels, at image level 4, were extracted around each correspondence pair for this purpose. After the local registration the patches are mostly well aligned (see *Fig. 3*) and the precision of the GT point set is increased.

4.2 Parameter Setting

One single parameter has to be set in advance for each point matching method. It is the matching ratio threshold for NNDR matching, the maximum distance error threshold for RANSAC and the mean shift kernel bandwidth for MBF. We assessed the parameter selection for the outlier filtering methods on two representative virtual slides, shown in *Fig. 1*. The results for all methods, performed at the resolution levels 3 to 6, are shown in *Fig. 4*. Only subsequent slices have been registered for this evaluation. The mean distance between

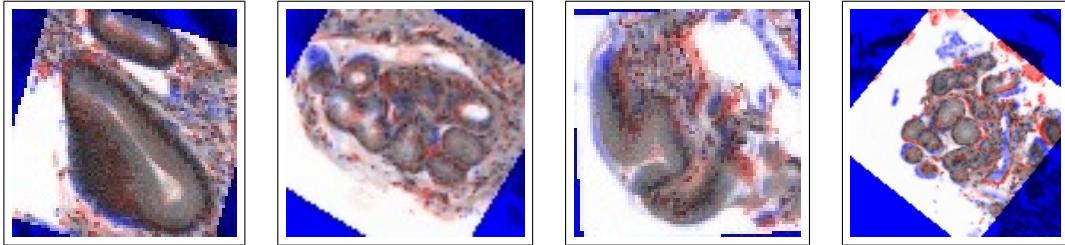


Figure 3: Locally registered patches ($50px \times 50px$) used during the fine tuning of the GT data. The images are shown overlaid, one in red the other in blue color. Grayish tones represent good alignment.

the true GT point positions and those obtained using our algorithms was used as the error measure.

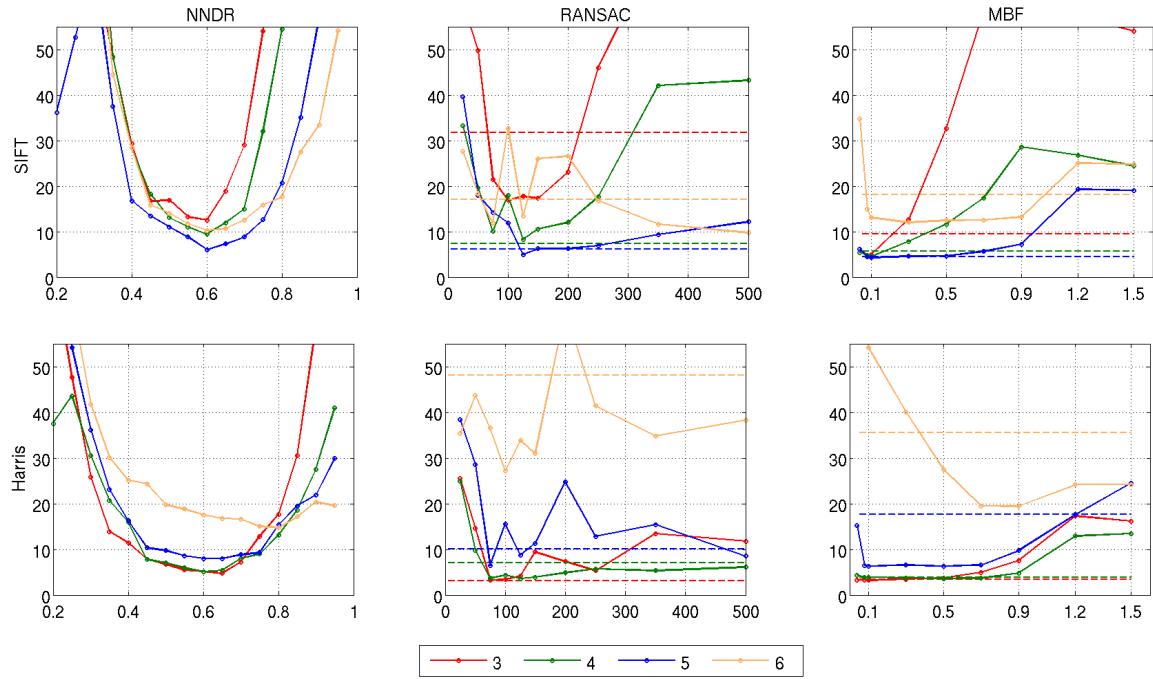


Figure 4: Evaluation of the registration specific parameters for each method: nearest neighbor distance ratio for the initial matching (NNDR), maximum distance error for RANSAC and mean shift kernel bandwidth for MBF. The mean GT point errors, in pixels at resolution level 4, are plotted. The results when using the SIFT (top) and the Harris (bottom) detector, examined for the methods NNDR, RANSAC and MBF (from left to right), are shown. The mean errors obtained at each resolution level while using the automatic parameter selection are depicted with dashed lines (see Sec. 4.2.1).

For the NNDR method, we observed the best results at value 0.6. Nevertheless, using this value the number of found correspondences is low and can lead to worse deformation field coverage. We therefore decided to use this tight value only in the case when NNDR matching is used with no subsequent outlier removal. For the other methods we obtain the initial

matches using the NNDR value of 0.8. This value represents a good trade-off between the number of obtained correspondences, and thus the final deformation field coverage, and the number of outliers, influencing the registration precision. We tend to allow for more outliers since they will be filtered out in the further steps. Following our experiments we set the maximum distance error threshold of the RANSAC method to 100 pixels (at resolution level 4) and the kernel bandwidth of the MBF method to 0.08. The selected values were observed to result in low errors while being also relatively stable across all evaluated resolution levels.

For the RANSAC-based outlier filtering, the threshold parameter defines the maximum acceptable error between the corresponding points for a given rigid transformation. Typically, at each iteration of our greedy approach, the RANSAC algorithm extracts the transformation with the least error. The set threshold is used to terminate the greedy extraction of the transformation models. It is therefore compared to the estimated error (see Sec. 3.3) of the best sampled model at each greedy iteration. Small parameter values would exclude the extraction of largely supported models and thus lead to few correspondences. The registration error would rise due to limited capabilities of our interpolation scheme. Too large values, on the other side, would give rise to outliers, also increasing the registration error. This behavior can be observed in *Fig. 4* (middle, solid lines). Note that the set threshold does not describe the final registration error, when compared to the ground truth. Possible outliers used during the rigid transformation estimation and local deformations of the tissue may result, locally, in larger real registration errors.

The bandwidth parameter of the MBF method describes the distribution of the parameters of the elemental transformations. Small values would partition the transformation space to many models with small support while large values would tend to extract only one common transformation. The dependency of the transformation error on the bandwidth parameter is shown in *Fig. 4* (right, solid lines).

4.2.1 Automatic Parameter Selection

A good value of the maximum error threshold for the RANSAC method can be estimated from the initial correspondence set. Using the least squares method, we first estimate one rigid transformation from the entire set of initial matches. We then set the necessary error threshold to the distance error computed for this transformation. The mean errors obtained when using the automatically selected maximum error threshold are depicted in *Fig. 4* (middle), drawn with dashed lines.

To automatically define a suitable parameter for the bandwidth of the MBF method we evaluate the distances among the elemental transformations using K-Nearest Neighbor approach (K-NN). We examined three techniques: using global mean bandwidth; computing and using single bandwidth for each elemental transform; computing the bandwidth at each iteration of the mean shift algorithm. We have set the bandwidth equal to the distance of the fifth nearest neighbor for all cases. Using the global mean method we have obtained the most promising results. The local approaches showed lower ability to separate the outlier transformations. The results obtained when using the automatically selected bandwidth parameters are shown in *Fig. 4* (right), drawn with dashed lines.

4.3 Results

We have evaluated our point-based registration method with three outlier removal settings and two salient point detectors. The NNDR matching was used with none, RANSAC and MBF outlier removal methods. For each method, SIFT and Harris feature points have been tested. We refer to these settings as to NS, NH, RS, RH, FS and FH, defining the applied outlier and the feature detector by the first and the second letter, respectively.

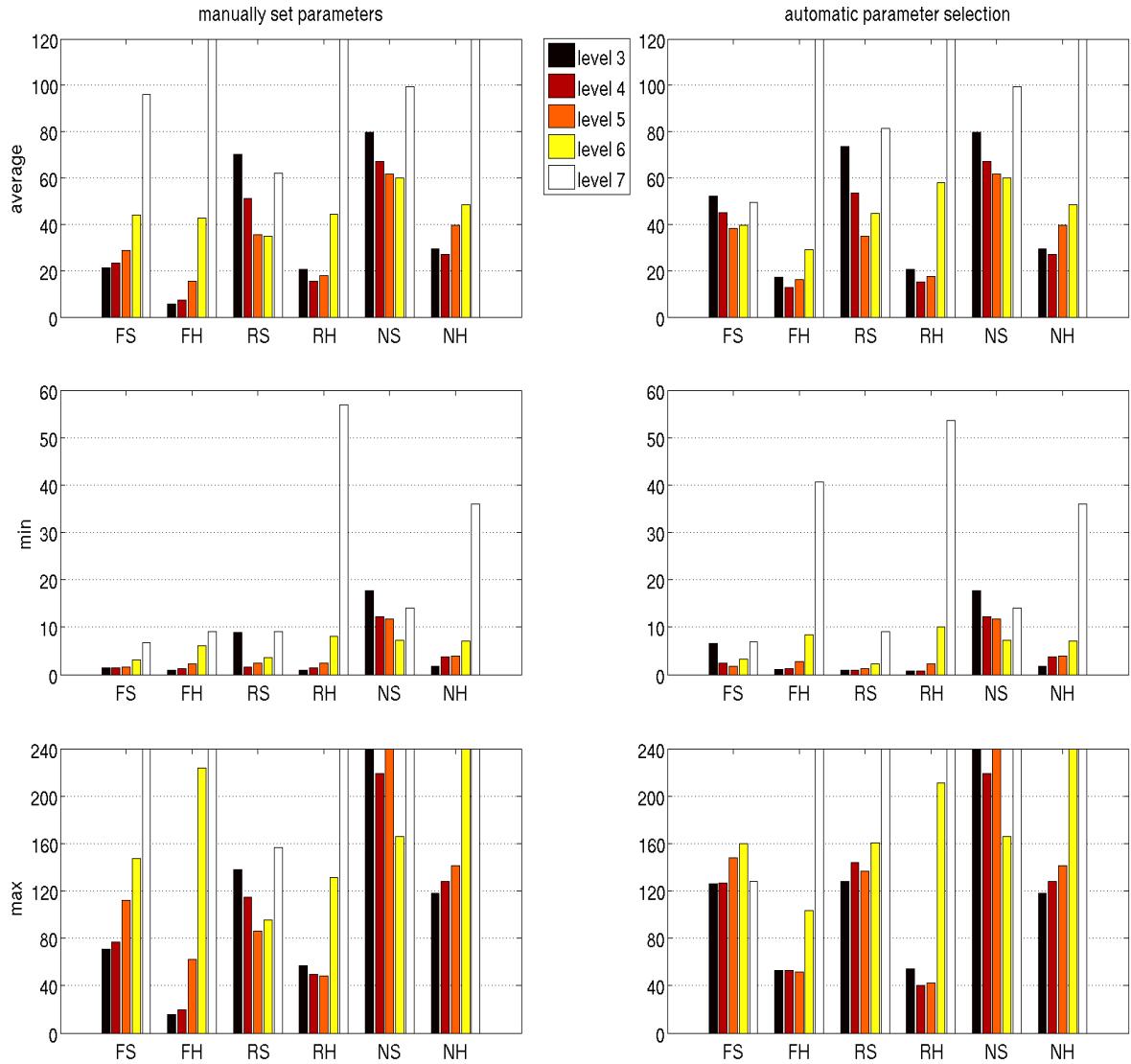


Figure 5: Evaluation results grouped by matching method, with registration algorithms run on all slice combinations. The root mean squared error, in pixels at resolution level 4, is plotted. The overall mean and the best and worst average results obtained for registrations across the specimen are shown from top to bottom. Note the different scale of the plots, adjusted for better visualization.

Using the selected parameters (see Sec. 4.2) we run the evaluation on our set of histological slides, excluding those used for the manual parameter selection. The registration quality

was measured by the distance between the true GT point positions and those obtained using our algorithms. The results, expressed by root mean square error (in pixels at resolution level 4) and evaluated for image resolution levels 3 to 7 are shown in *Fig. 5*. The left side of each image presents setting with manually selected parameters, on the right automatically selected parameters were used for the outlier filtering methods RANSAC and MBF. Registrations between all slice combinations per specimen have been considered. On the top, the overall average performance over all specimen (460 slice registrations) is presented. The minimum and maximum registration error across specimen are shown below. Considering that at level 4 the used images were of an average size 1200×1200 pixels, the best mean error is below 5%.

The automatic parameter selection for the outlier filtering methods shows comparable results to our manual selection (see *Fig. 5*) in the RANSAC setting. For the MBF approach, larger differences occur at finer resolutions. We believe that this deficit comes from the use of only five nearest neighbors for the bandwidth selection. Post-experimental tests indicate that increasing this number at higher resolutions helps.

More results are shown in *Fig. 6*. They provide comparison between simple scenario of registering only subsequent slices (left) and more difficult setting where three intermediate slices were skipped between each registered pair of histological sections (right). The similarity between the slices in the latter setting is therefore strongly reduced and the benefits of outlier filtering are visible. 130 and 63 cases were considered for the former and the latter scenario, respectively. Empirically selected parameters were used for this evaluation.

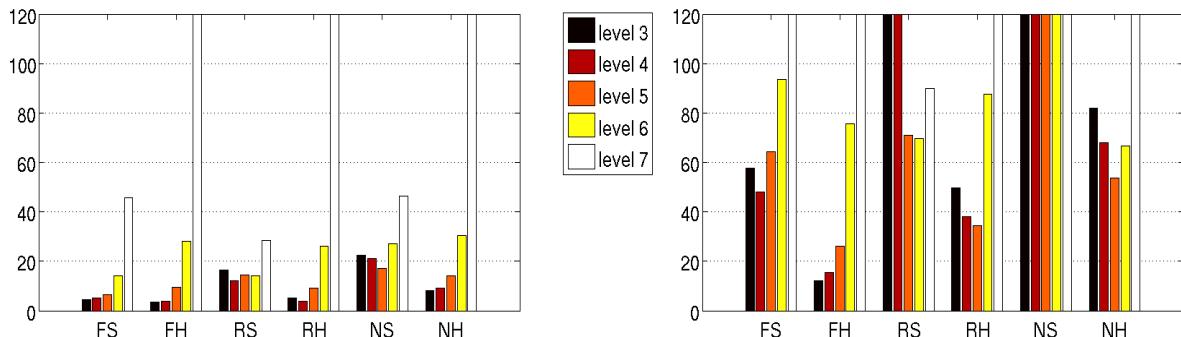


Figure 6: Evaluation results grouped by method, with registration algorithm run on successive slices only (left) and on slices separated by three intermediate sections (right). The root mean squared error, in pixels at resolution level 4, is plotted. Empirically set parameters were used.

The best results have been obtained when additional outlier filtering is used (RANSAC or MBF). The necessity of applying these filters is especially visible when registering slices which are not consecutive in the cutting order (see *Fig. 6*, right). Note that good results can already be obtained using the simple NNDR matching when run on subsequent slices. In contrast to our preliminary study [Leškovský and Hellwich, 2011] we attribute this significantly better performance of the NNDR method to the tighter parameter setting of the nearest neighbor distance ratio and also to the enhanced interpolation method defining the final deformation field.

At further inspection (see *Fig. 5*) one can identify that at resolution level 4 good results can be obtained for all methods. Little improvement can be observed for the finer level 3.

Although the number of detected features grows exponentially at finer levels, only close to linear increase of initially matched features was observed (see Fig. 7). It therefore suggests that the SIFT descriptor is not well discriminative at the fine resolution. This behavior was expected. At coarse levels the features correspond to macroscopic structures which can be easily traced across slices. At fine resolution cells become visible, the feature points extracted at this level become more similar and therefore do not pass the NNDR test. The need for finer resolutions thus shows to be insignificant for our global matching approach.

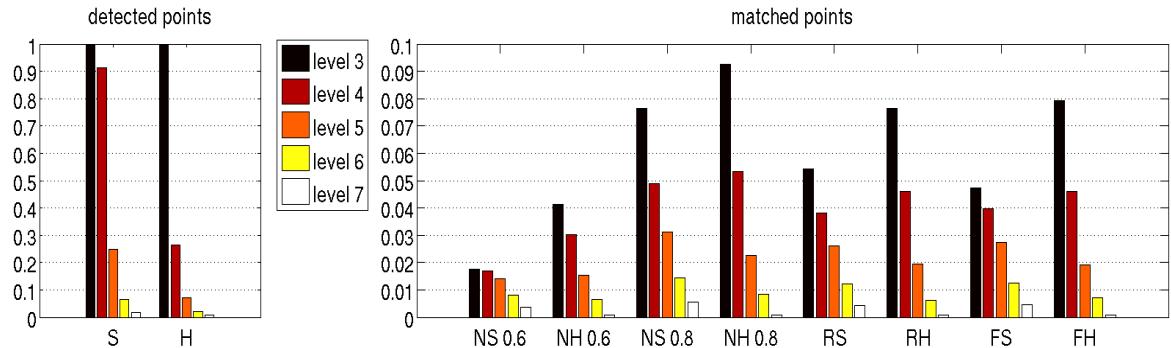


Figure 7: Average amount of detected and filtered feature points across resolution levels 3 to 7. Relative numbers to the count of Harris feature points detected at resolution level 3 are presented. Note that the count of detected SIFT points at level 3 was 4.5 times larger than that of the Harris points, exceeding thus the plotted graph. The NNDR filtering method is shown for the matching ratio values 0.6 and 0.8. The output of the latter setting is the input for further outlier filtering by RANSAC or MBF.

Overall, the MBF and the RANSAC performance is similar. Nevertheless, the RANSAC algorithm demonstrated several deficits. It performs slow due to the repetitive splitting of the feature point set and the error estimation of each transformation model for the outlier identification. Moreover, the random sampling of the RANSAC method causes that the quality of the results between different runs varies. Careful selection of the intrinsic parameters is necessary to mitigate this effect. We therefore prefer the MBF technique for its faster execution and throughout search of the transformation space.

The superiority of Harris detector is noticeable for all methods applied, through all levels. Even with no outlier filtering it significantly outperforms the SIFT detector. Note that this has also been observed during the experimental parameter setting (see Fig. 4) at finer resolution levels. Checking the number of points extracted (see Fig. 7) we observed that the SIFT detector finds the most features at all levels. Nevertheless, already after the initial NNDR matching the counts of the feature points become similar. It is also visible that further outlier filtering removes more SIFT than Harris points, notable especially at higher resolutions. Both, the large amount of filtered SIFT points and the superior registration results obtained when using Harris detector demonstrate that the SIFT detector is too sensitive for our data. Consequently, we conclude that the Harris operator is more suitable for our image data.

The Figures 8 and 9 demonstrate the detection and matching pipeline for SIFT and Harris feature points. The feature extraction as well as the matching were performed at level 4. First, a close-up view of the detected points is shown (see Fig. 8, top). Around 7000 SIFT and 2000 Harris points were detected on each slice. The putative matches obtained using the

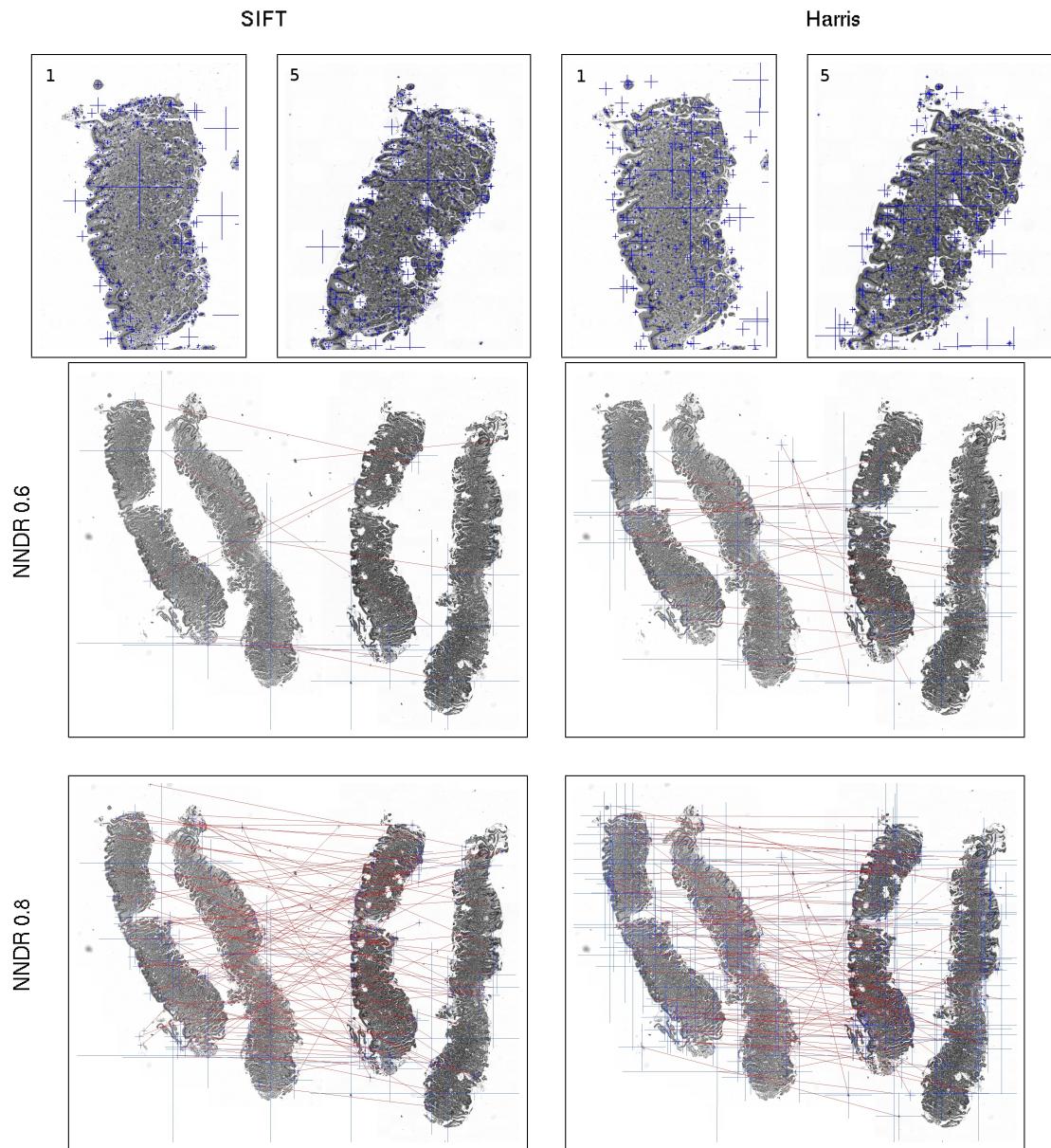


Figure 8: Detection and matching results for SIFT (left half) and Harris (right half) feature points. The process was run at resolution level 4. A close-up view of the detected points and matching obtained by the NNDR method with parameter 0.6 (middle), and 0.8 (bottom) are shown. The sizes of the blue crosses in the top images corresponds to 2×2 pixels regions at the detected scale of the feature point. In following images the blue crosses correspond to the area over which the descriptor was computed (16×16 pixels square).

NNDR parameter 0.6 and 0.8 follow. 9 and 92 SIFT correspondences were obtained in contrast to 19 and 80 Harris matches, respectively for the NNDR parameter 0.6 and 0.8. Results using further outlier filtering by RANSAC and MBF are shown in Fig. 9. In both figures the SIFT points setting is shown on the left, Harris points setting on the right. Superior matching

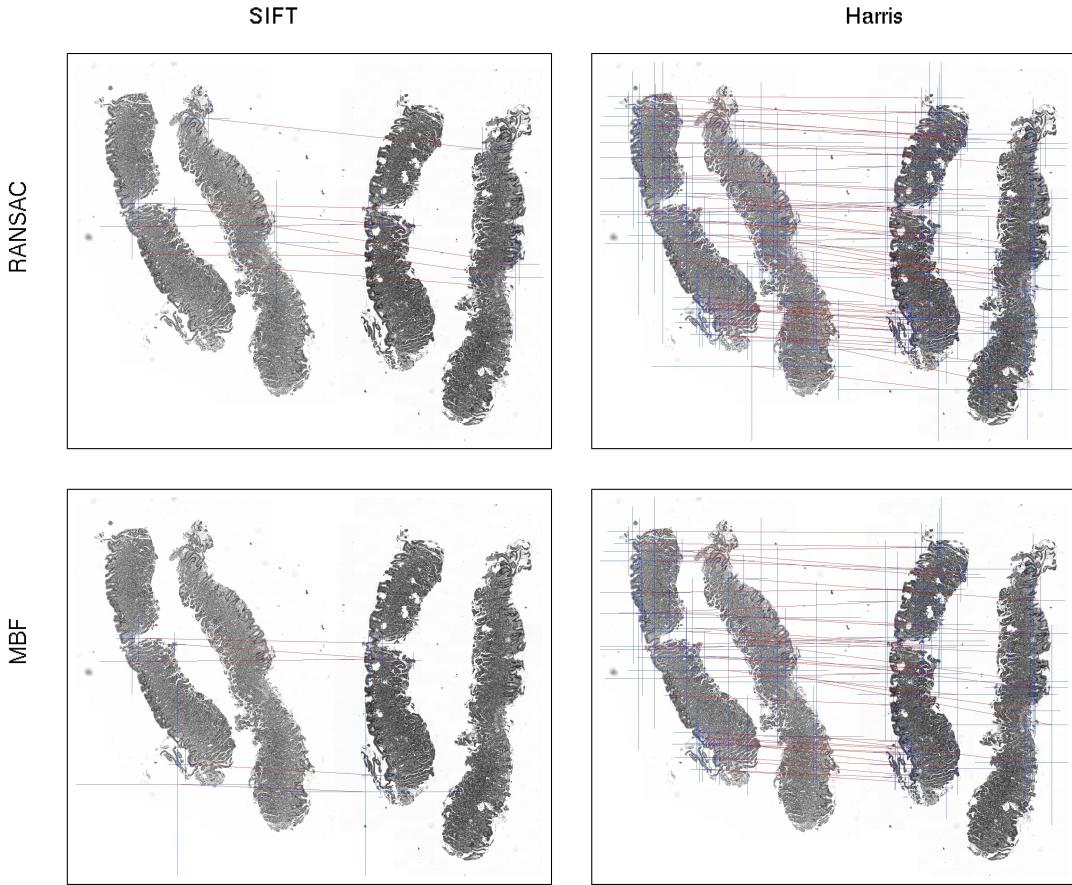


Figure 9: Matching results for SIFT (left) and Harris (right) feature points, shown in Fig. 8, as obtained by methods RANSAC (top) and MBF (bottom). The sizes of the blue crosses correspond to the area over which the descriptor was computed (16×16 pixels square). The process was run at resolution level 4.

performance for the Harris feature points over the SIFT ones is notable. This is due to more SIFT mismatches being present after the initial matching. One can also see that the Harris feature points were better distributed over the different scales. The SIFT features, on the other hand, showed to be too sensitive at the finest scale. The final results obtained after the further outlier filtering are similar.

In Fig. 10 and Fig. 11, two examples of registration results are presented. In the former one, results of the NS, NH and FH settings are compared, registering two consecutive slices. We can observe worse results when using the SIFT feature points (compare top and middle images), but similarly good results when Harris points are used, even without further outlier filtering by MBF (compare middle and bottom images). Note that the additional outlier filtering proved to be more important for cases when nonconsecutive slices are registered. In the latter figure, the registration on nonconsecutive slices is compared. The slice images are more different and therefore less correspondences can be found in contrast to the previous case. First, the result of the NH method when using NNDR parameter 0.6 is shown (Fig. 11, top). Few outliers can be seen, however, some tissue parts are not matched. Since all initial

matches are used in this setting, the additional outlier filtering would hardly recuperate the missing sections. We have to increase the number of considered feature points. We do this by selecting higher value of the NNDR parameter (*i.e.* 0.8). This, however, introduces outliers in the previously well matched regions (*Fig. 11, middle*). Finally, applying outlier filtering by MBF we obtain superior results (*Fig. 11, bottom*).

5 Conclusions

We examined point-based matching methods for the registration of histological images. It was not clear whether general purpose point descriptors can provide enough discriminative power in order to be applied in this specific task. Our results show that salient feature detectors, like SIFT and Harris, combined with the SIFT descriptor are well adaptable for this purpose. In this context we identified Harris-based corner features more representative for the tissue samples than the blob like SIFT features.

We have also shown that already at middle resolution levels fair alignments can be obtained. Going for higher resolutions the cells become visible and one would face the problem of less discriminative descriptor assigned to the new fine level features. This would mostly result in unnecessary computational overhead.

We also evaluated two techniques which can be applied for filtering the outliers, RANSAC and MBF. They greatly improve the registration results when good parametrization is chosen. A convenient method for the automatic parameter setting was also suggested. Although for the MBF method minor precision was obtained in the automatic setting, the suggested K-NN approach offers more intuitive parameter setting than the direct selection of the filtering bandwidth in the transformation space. Finally, we suggest to use the MBF method which is deterministic, applies global optimization and is executed faster, all in contrast to the presented RANSAC method.

Acknowledgments.

This work was supported by the German Federal State of Berlin in the framework of the Zukunftsfonds Berlin and the Technology Foundation Innovation center Berlin (TSB) within the project Virtual Specimen Scout (<http://www.specimen-scout.de>). It was co-financed by the European Union within the European Regional Development Fund (ERDF).

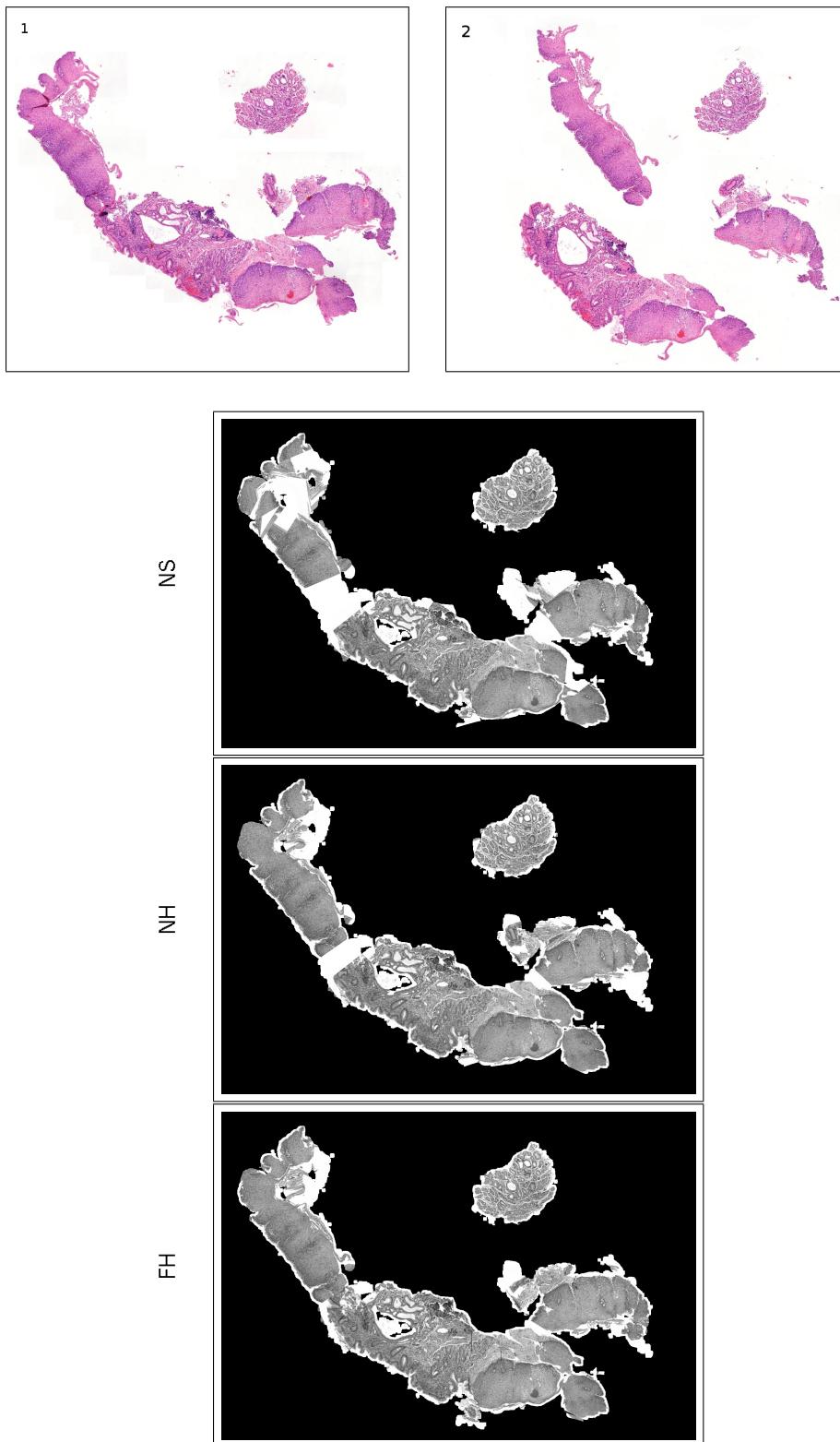


Figure 10: Two subsequent slices registered by methods NS, NH and FH (from top to bottom). Better matching is obtained when Harris points are used. Little difference is observed between NNDR matching and MBF outlier filtering in this setup. The process was run at image level 4. The input images of the selected slices are shown in the first row, the deformed moving slices follow below. The tissue mask of the first slice is overlaid to better visualize the final registration result.

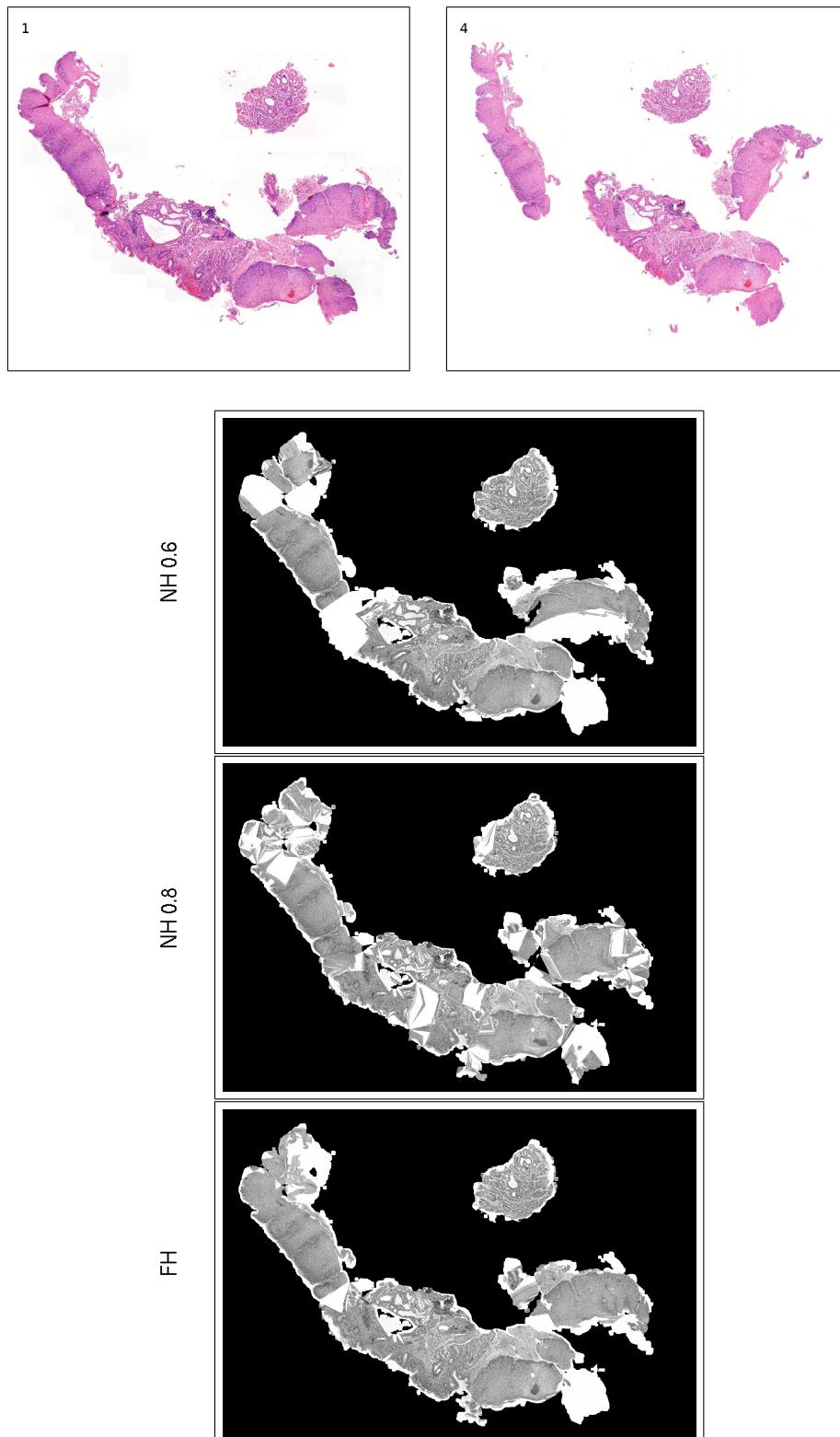


Figure 11: Registration of slices separated by two intermediate ones as obtained by methods NH (with parameter 0.6), NH (with parameter 0.8) and FH (from top to bottom). In this setting the outlier filtering by MBF noticeably enhances the result. The process was run at image level 4. The input images of the selected slices are shown in the first row, the deformed moving slices follow below. The tissue mask of the first slice is overlaid to better visualize the final registration result.

References

- I. Arganda-Carreras, C. Sorzano, R. Marabini, J. Carazo, C. Ortiz-de Solorzano, and J. Kybic. Consistent and elastic registration of histological sections using vector-spline regularization. In *Computer Vision Approaches to Medical Image Analysis*, volume 4241 of *LNCS*, pages 85–95. Springer Verlag, 2006.
- F.S. Cohen, Z. Yang, Z. Huang, and J. Nianov. Automatic matching of homologous histological sections. *Biomedical Engineering, IEEE Trans.*, 45(5):642 –649, 1998.
- L. Cooper, O. Sertel, J. Kong, G. Lozanski, K. Huang, and M. Gurcan. Feature-based registration of histopathology images with different stains: An application for computerized follicular lymphoma prognosis. *Computer Methods and Programs in Biomedicine*, 96(3):182 –192, 2009.
- A. du Bois d'Aische, M. De Craene, X. Geets, V. Gregoire, B. Macq, and S. K. Warfield. Efficient multi-modal dense field non-rigid registration: alignment of histological and section images. *Medical Image Analysis*, 9(6):538 – 546, 2005.
- M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.
- P. Leškovský and O. Hellwich. Point based registration of high-resolution histological slices. In *Proc. Medical Image Understanding and Analysis 2011*, pages 213–217, 2011.
- D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60:91–110, 2004.
- A. Pitiot, E. Bardinet, P. M. Thompson, and G. Malandain. Piecewise affine registration of biological images for volume reconstruction. *Medical Image Analysis*, 10(3):465 – 483, 2006.
- C.O.S. Sorzano, P. Thevenaz, and M. Unser. Elastic registration of biological images using vector-spline regularization. *Biomedical Engineering, IEEE Trans.*, 52(4):652 –663, 2005.
- Adam Stanski. *Konstruktives Probabilistisches Lernen*. PhD thesis, Technische Universität Berlin, DE, 2012. URL www.cepel.de. (date of last access: 3 Aug. 2012).
- R. Subbarao and P. Meer. Nonlinear mean shift over Riemannian manifolds. *Int. Journal of Computer Vision*, 84:1–20, 2009.
- T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3:177–280, 2008.
- J. Yang, J. P. Williams, Y. Sun, R. S. Blum, and C. Xu. A robust hybrid method for nonrigid image registration. *Pattern Recognition*, 44(4):764 – 776, 2011.
- T.S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxes, and R. Whitaker. Engineering and algorithm design for an image processing api: A technical report on itk - the insight toolkit. In *Proc. Medicine Meets Virtual Reality*, pages 586–592. IOS Press Amsterdam, 2002.