

# Feature-Based Multi-Resolution Registration of Immunostained Serial Sections

Oleg Lobachev<sup>a,\*</sup>, Christine Ulrich<sup>b</sup>, Birte S. Steiniger<sup>c</sup>, Verena Wilhelmi<sup>c</sup>, Vitus Stachniss<sup>d</sup>, Michael Guthe<sup>a</sup>

<sup>a</sup>Visual Computing of University Bayreuth, 95440 Bayreuth, Germany

<sup>b</sup>Psychology of Philipps-University Marburg, 35037 Marburg, Germany

<sup>c</sup>Institute of Anatomy and Cell Biology of Philipps-University Marburg 35037 Marburg, Germany

<sup>d</sup>Restorative Dentistry and Endodontics of Philipps-University Marburg, 35037 Marburg, Germany

---

## Abstract

This document serves as a supplementary text to the main paper. It handles some details not mentioned there, shows additional figures, and gives an exact definition of quality measures used. Here, we prefix all the references to the main paper with “M.” The main paper prefixes all the references to this document with “S.”

**Keywords:** registration, multi-resolution, feature detection, splines, histology

---

## 1. Supplementary Materials

### 1.1. Technical Details on Matching (Section M3.2)

A *direct* match of all-against-all key points is neither feasible nor produces good results, see Fig. 1. We rematch the key point pairs basing on a weighted implementation of radius match. Then least-squares optimization follows to yield the positions of B-spline control points for the actual non-rigid distortion. Compare the matches of a rigid stage with the matches of the non-rigid stage: in Fig. 1, a) shows decimated direct matches, b) is the rigid match we find in the first iteration, c) shows matches from the non-rigid registration.

### 1.2. Pipeline (Section M3.3)

We present a side view on serial sections of spleen, double-stained, in Fig. 2 (a–c) show the horizontal the  $z$ -stacks and demonstrate that our method is superior not only to the single-resolution registration (“A”), but also to the repeated application of single-resolution registration to the ROI (“A–A”). Compare Fig. 2 a) to Fig. 2 b): we see some positive effects of a second application of non-multi-resolution registration. However, in Fig. 2 c) we see a more significant effect of our method (“A–B”). The shapes of the blood vessels and other objects in this  $z$ -stack are better, esp. in marked positions. This is even more pronounced in the volume renderings in Fig. M11 in the main text.

In Fig. 3, (a) shows the crop zone and the ROI (whole image shown) used for the fine-grain registration. Fine capillaries are not properly registered using the slice-wide, coarse registration in (b). We used the method “A” here. It served as an input for (c) and (d). Compare the fine-grain registration using (c) single-resolution registration “A” again (yielding “A–A”) with (d) our method “B”, the multi-resolution registration (resulting in “A–B”). Small capillaries should profit from our method as

they are too small to produce features that are large enough to be regarded by a single-resolution method, esp. in a slice-wide application. We compare these two methods in detail in main paper.

### 1.3. Results: Volume Renderings (Section M4.3)

Fig. 4 shows the full volume rendering of a single-stained spleen specimen. Figure M7 (in main paper) shows a region from this image that was only cut out, without any re-registering.

Fig. 5 shows the processed bone marrow specimen as a volume rendering of the alignment with our method. We used two levels of multi-resolution registration.

### 1.4. Discussion (Section M5)

To further highlight our method we show  $z$ -stacks and further volume renderings of another bone marrow ROI in Figures 6–7. Observe the precision of section matching in the  $z$ -stacks (Fig. 6). These images are not directly comparable to each other, because different registrations have distorted them differently. Instead, compare the quality of the separate shapes without comparing the shapes directly. We also show complete volume renderings accompanied by more detailed crops (Fig. 7). The crops in Fig. 6 b), d), f) show visually the benefit from using our method.

### 1.5. Applying Our Method to Standard Images (Section M5.5)

To show that our method is applicable not only to histological images, we generated distorted series from standard test images and registered them with our method.

We took two standard images from signal processing: “Lena” and “mandrill”. To use them for the image registration, we randomly distorted the images. We moved random points of the input in random direction and distance using Shepard’s distortion. Both the offsets and the initial positions were generated using a normal distribution with  $\sigma = 204$  for positions; with  $\mu = 2.5$  and  $\sigma = 4$  for offsets. (Fig. 8) All the difference images here are negated for the sake of presentation—white is no

---

\*E-mail: oleg.lobachev(at)uni-bayreuth.de

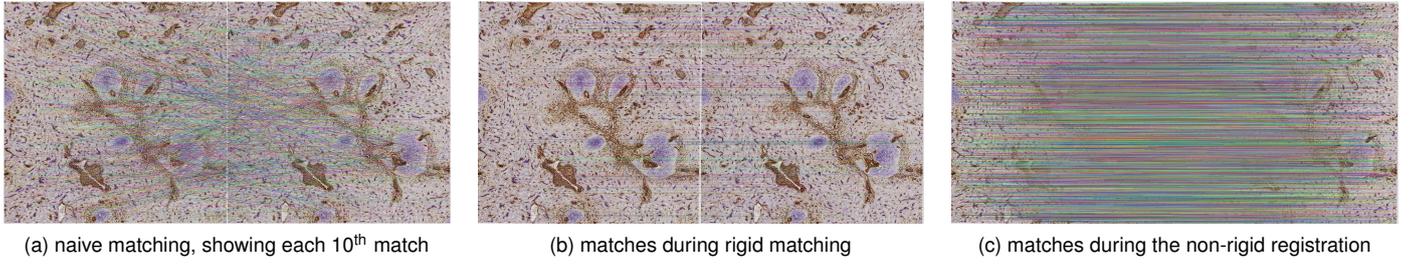


Figure 1: Matches in a ROI from spleen data, double staining. Fig. (c) shows matches at the end of the first non-rigid registration phase, i. e. those with largest search radius and size. We add further (smaller) matches within a smaller search radius in the next phases. The registered data set is in Figure M8 in the main text.



Figure 2: Stack of 24 serial human spleen sections double-stained for CD34 plus SMA (brown) and CD271 (blue). Compare the  $z$ -profiles of this specimen. All the sections underwent a single-resolution registration and were then cropped to ROI. Note the differences in the vascular network at the marked positions.

difference. The image size was  $512 \times 512$  pixels. We upscaled the input images after the distortion  $6\times$ . Then we performed the registration using our method, see Fig. 8, page 6 for the results.

## 2. Extended Quantitative Evaluation

### 2.1. Definitions of the Quality Measures (Section M5.3)

We discuss seven different quantitative measures of registration quality for each of the specimens (Fig. M4–M6) below. The paper has omitted the formula for the quality measures used, we fix this here.

For the notation, consider two images  $\mathbf{A}$  and  $\mathbf{B}$ , two single pixel values at  $(i, j)$  are  $A_{ij}$  and  $B_{ij}$ , the mean values at  $(i, j)$ -centered window are  $\bar{A}_{ij}$  and  $\bar{B}_{ij}$ , i. e. for  $n \times m$ -sized window,

$$\bar{A}_{ij} = \frac{1}{nm} \sum_{k=i-n/2}^{j+n/2} \sum_{l=j-m/2}^{j+m/2} A_{kl}.$$

The cross-correlation in the same  $n \times m$  window is

$$\rho_{ij}(\mathbf{A}, \mathbf{B}) = \frac{1}{nm} \sum_{k=i-n/2}^{i+n/2} \sum_{l=j-m/2}^{j+m/2} (A_{kl} - \bar{A}_{ij})(B_{kl} - \bar{B}_{ij}).$$

With  $\partial A_{ij} / \partial x$  we denote the derivative of  $\mathbf{A}$  at  $(i, j)$  in the direction of  $x$  axis.

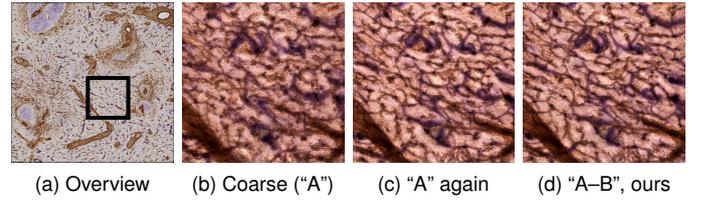


Figure 3: Stack of 24 serial human spleen sections double-stained for CD34 plus SMA (brown) and CD271 (blue). We showcase how well small capillaries are represented, this is pivotal for histological research basing on this images. Capillary sheaths are visible as blue areas around brown capillaries. Compare the crops from the volume rendering, from left to right: (a) legend, a single slice showing the registered ROI and post-registration crop location, magnification factor 30, (b) coarse, section-wide registration only using single-resolution registration of Ulrich et al., (c) the series from b) was registered *again* using the single-resolution method, (d) the series from b) was registered using our method. We used four levels of multi-resolution registration. The magnification factor for (b–d) is 120.8.

To obtain the data for quality evaluation we registered the *whole* series with each method, picked a pair of consecutive images and compared them using the following measures over the whole ROI (we show later a detailed view when applicable):

- structural similarity (SSIM) is

$$\frac{(2\bar{A}_{ij}\bar{B}_{ij} + C_1)(2\rho_{ij}(\mathbf{A}, \mathbf{B}) + C_2)}{(\bar{A}_{ij}^2 + \bar{B}_{ij}^2 + C_1)(\rho_{ij}(\mathbf{A}, \mathbf{A}) + \rho_{ij}(\mathbf{B}, \mathbf{B}) + C_2)},$$

with stabilization constants  $C_1$  and  $C_2$ , using  $11 \times 11$  Gaussian weighting. Final single SSIM value is the mean over the overall SSIM image.

- pixel-wise sum of squared differences (SSD), i. e.  $\sum (A_{ij} - B_{ij})^2$ .
- zero-mean sum of absolute differences (ZSAD) with window  $16 \times 16$ ,

$$\sum (A_{ij} - \bar{A}_{ij}) - (B_{ij} - \bar{B}_{ij}).$$

We compute the absolute differences pixel-wise, but beforehand subtract the mean of the values in the window around the pixel from the actual pixel values. The resulting images are quite dark, for the sake of presentation we adjust the brightness and contrast to  $+50$  in all presented ZSAD images. The numerical values are not corrected.

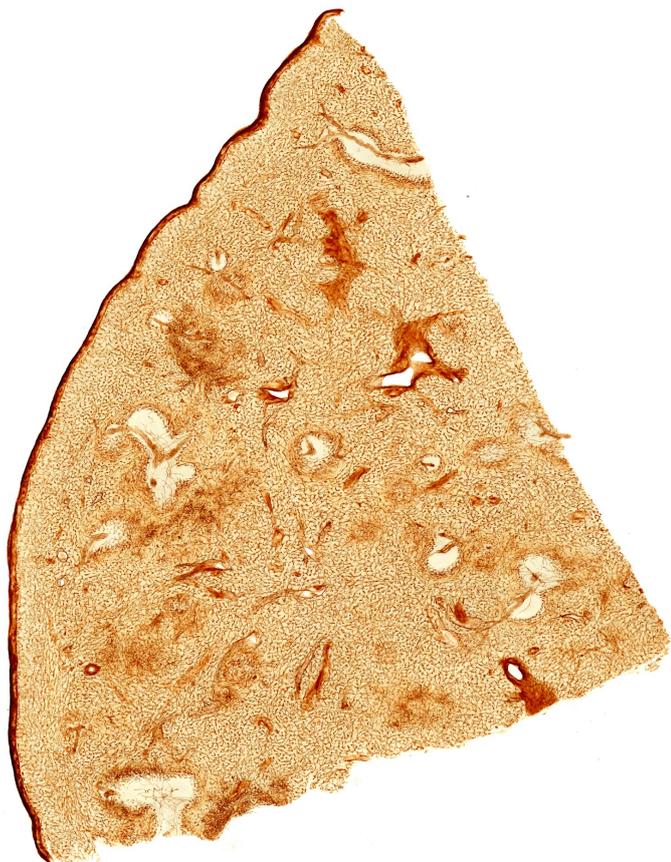


Figure 4: Volume rendering of spleen, single-stained for CD34. We overlay 24 images after 4-levels of our multi-resolution registration. Figure M7 in the main text shows a crop of this 9× overview image.

- mutual information (MI) increase. We compute the mutual information for an image pair. For the actual value we divide the mutual information in two registered consecutive images of a scan series by the mutual information in the original, unregistered image pair to obtain the increase of mutual information compared to baseline.

- gradient cross-correlation (GCC):

$$\rho_{ij}(\partial A_{ij}/\partial x, \partial B_{ij}/\partial x), \quad \rho_{ij}(\partial A_{ij}/\partial y, \partial B_{ij}/\partial y).$$

We compute the gradient separately in the direction of the  $x$  and  $y$  axes using Scharr operator, compute the cross-correlation using window size 16 and then blend both images together. As the result is too dark, we use gamma correction with  $\gamma = 6$  before blending. The numerical values are not corrected.

- Farneback's dense optical flow that is defined in the following way. We model a signal with quadric polynomials, e.g.  $f_1(\mathbf{x}) = \mathbf{x}^T \mathbf{A}_1 \mathbf{x} + \mathbf{b}_1^T \mathbf{x} + c_1$  for a symmetric matrix  $\mathbf{A}_1$ , a vector  $\mathbf{b}_1$  and a scalar  $c_1$ . A globally displaced by  $\mathbf{d}$  signal  $f_2$  is defined in a similar way in terms of  $\mathbf{A}_2, \mathbf{b}_2$  and  $c_2$ . We can express  $\mathbf{d}$  as  $-\mathbf{A}_1^{-1}(\mathbf{b}_2 - \mathbf{b}_1)/2$ . Now, to make the

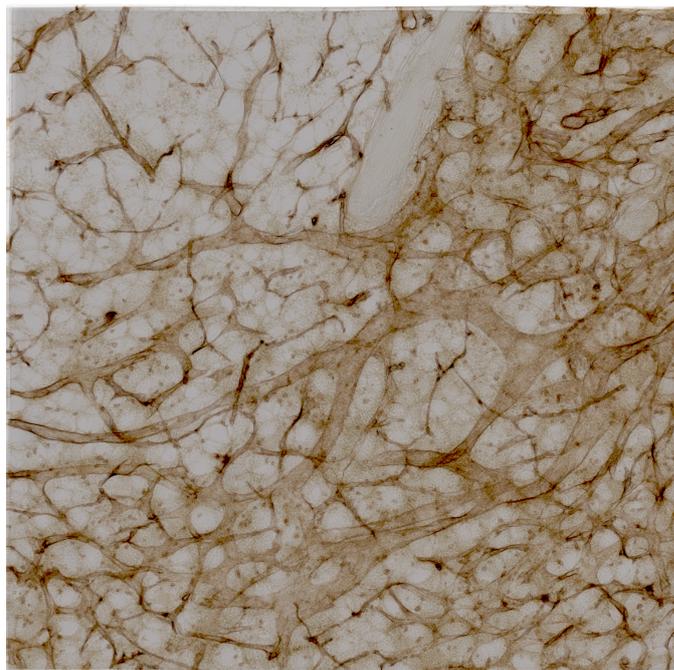


Figure 5: Stack of 30 human bone marrow sections single-stained for CD34 and CD141 (both in brown). In contrast to Section M4 and Figure M10, it was registered with *two* levels of our method, after the single-resolution registration was applied both to the full section and to the ROI. We evaluate this ROI in Figure 16.

computation more practical we use *local* polynomial approximations of images. In the following description of the optical flow we omit the  $\mathbf{x}$  argument, e.g.  $\mathbf{A}$  always stands for  $\mathbf{A}(\mathbf{x})$ , a local approximation. Let

$$\mathbf{A} := (\mathbf{A}_1 + \mathbf{A}_2)/2 \quad \text{and} \quad \Delta \mathbf{b} := -(\mathbf{b}_2 - \mathbf{b}_1)/2.$$

The distance is now a spatially varying displacement field  $\mathbf{d}$  with  $\mathbf{A}\mathbf{d} = \Delta \mathbf{b}$ . Let  $w$  be a weight function—the edges of an image have a smaller weight. Now, in a neighborhood  $I$  following holds:

$$\mathbf{d}(\mathbf{x}) = \left( \sum_I w \mathbf{A}^T \mathbf{A} \right)^{-1} \sum_I w \mathbf{A}^T \Delta \mathbf{b},$$

the sums run over all  $\mathbf{x}$  in  $I$ , as  $\mathbf{x}$  is omitted for brevity argument to  $\mathbf{A}$ ,  $\Delta \mathbf{b}$ , and  $w$ . In practice we compute  $\mathbf{A}^T \mathbf{A}$ ,  $\mathbf{A}^T \Delta \mathbf{b}$ , and  $\Delta \mathbf{b}^T \Delta \mathbf{b}$  (needed for the confidence value) point-wise.

The flow is computed on multiple scales to alleviate the problem with larger displacements. We computed it based on a classical  $2^n$  image pyramid with 3 levels, averaging window size of 15 pixels, three iterations at each level, neighborhood size for polynomial expansion of 5 pixels. The derivatives for polynomial expansion are smoothed with  $\sigma = 1.2$ .

- The Jaccard similarity measure for two sets  $X$  and  $Y$  is defined as

$$J(X, Y) = \frac{\#(X \cap Y)}{\#(X \cup Y)},$$

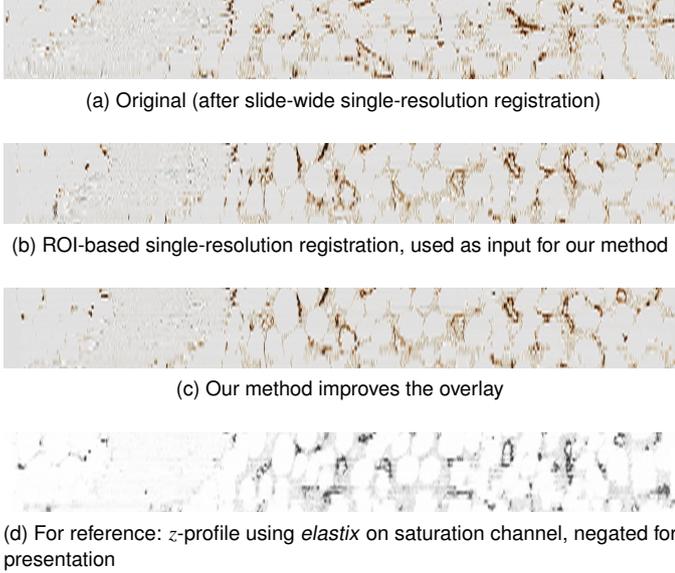


Figure 6: Human bone marrow, z-profiles of 30 sections single-stained for CD34 and CD141 (both in brown), showing the same ROI as in Figure 7

the valid values lie between zero (no similarity) and one (full similarity). As our inputs are images, say,  $\mathbf{A}$  and  $\mathbf{B}$  we threshold the data at a given value, yielding a coarse kind of segmentation. Let the binary threshold of an image  $\mathbf{A}$  at value  $x$  be  $t_x(\mathbf{A})$ . Then we compute the Jaccard measure for the resulting binary images. For multi-channel images we perform the computation channel-wise and take the maximum. Let  $r, g, b$  be the channels, and  $\mathbf{A}^r$  a single channel of an image  $\mathbf{A}$ . Hence, the Jaccard measure we use is

$$J(\mathbf{A}, \mathbf{B}) = \max_{c \in \{r, g, b\}} \frac{\#(t_x(\mathbf{A}^c) \cap t_x(\mathbf{B}^c))}{\#(t_x(\mathbf{A}^c) \cup t_x(\mathbf{B}^c))},$$

we use RGB channels when applicable.

The result is presented in Fig. 9 (SSIM for all specimens), Figs. 10–12 and Table M2 in the main text (optical flow), Table 1 (Jaccard measure for all specimens), Fig. 14 for spleen, single-stained, Fig. 15 for spleen, double-stained, Figs. 16 and 17 for bone marrow present results of SSD, ZSAD, mutual information increase, and GCC for respective specimens.

Figs. 9, 16 and 17 show the results both for bone marrow with two levels of registration with our method and with four levels of registration. This facilitates the comparison of different multi-resolution approaches and underlines the importance of iteratively decreasing feature sizes in a registration.

## 2.2. Full Evaluation (Section M5.3)

Section M5.3 (main document) does not show the complete evaluation because of article size limitations. We show it here in complete detail.

Images in Figs. 14–16 show crops to emphasize details of the registration. Below the images we show mean values for whole images, not for the crops shown. All differences and improvements in per cent are relative values.

The SSIM and Jaccard measure images and values are better when lighter and higher. SSD, ZSAD, GCC, optical flow images and values are better when darker and lower.

### 2.2.1. SSIM for All Specimens (Fig. 9)

Figure 9 shows the structural similarity index for spleen, single-stained (b–f), spleen, double-stained (h–l), and bone marrow (n–r). Figs. 9 a), g), m) show corresponding specimens registered using our method.

The SSIM images were very convincing for spleen, single-stained specimen. The input overlay Fig. 9 b) produced a very dark (i.e. bad) image, our method (f) was much lighter, the lightest in the series. The images for *elastix* (d) and the method of Ulrich et al. (e) were quite similar to rigid-only transformation (c), but a bit *darker*, which corresponded to a minor decrease of the quality value. Overall, our method showed 45% improvement over baseline, while *elastix* yielded 31% of improvement. The relative difference between the SSIM for our method Fig. 9 f) and for *elastix* (d) was over 10%. The difference between SSIM for our method (f) and for the method of Ulrich et al. (e) was over 14%.

Again, SSIM for spleen, double-stained were much darker for the initial overlay (h) than for our method (l). The SSIM value here improved more than by 146%, the best overall result. As for *elastix* (j) the improvement was 90.5%. All three non-rigid methods, (j–l), showed SSIM images much lighter than the rigid-only transformation (i).

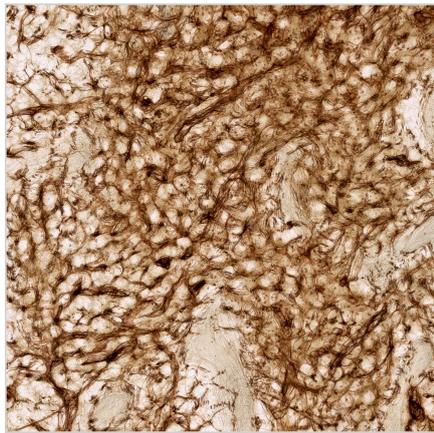
As for bone marrow at two levels of our method, the difference was not that high. The image for our method Fig. 9 r) appeared a bit lighter (i.e. better) than the input overlay (n). Our multi-resolution method improved the overlay of two images by 5.7%, however repeated application of the non-multi-resolution method by Ulrich et al. (q) did not improve the situation at all. The SSIM quality decrease was here  $-6.8\%$ . Same holds for *elastix* (p), it impaired the registration by  $-19\%$ .

The bone marrow specimen at four levels of our method (x) showed more than 50% of SSIM improvement over the coarse-only registration (t) for a  $4k \times 4k$  ROI. The result of *elastix* (v) was visibly darker, indeed it showed a decrease in SSIM value w.r.t. baseline (t). It corresponded to over 44% quality decrease when compared to our method. Ulrich et al. (w) facilitated 37.6% improvement over baseline (t).

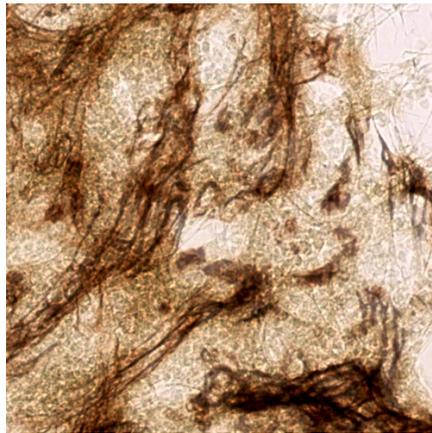
### 2.2.2. Optical Flow (Section M5.3.2, Figs. 10–12)

We measure the magnitude of dense optical flow between to consecutive images based on work of Farneback. This method basically approximates the both signals locally in neighborhood with quadric polynomials, then solves for the translation vector. For the practical reasons the neighborhoods are averaged and the fitting happens iteratively on multiple scales. We use pairs of consecutive registered images as input and evaluate the motion vector field in terms of its magnitude and angle.

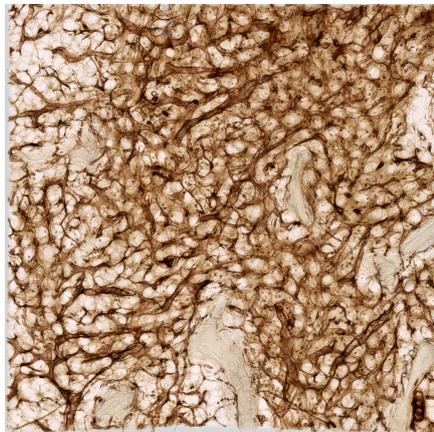
Table M2 lists the means of the magnitude of translation vector between two consecutive registered images. Here we extend it with Figures 10, 11, and 12 that show angle and magnitude between two slices as a color coded image for single-stained spleen, double-stained spleen, and bone marrow specimens.



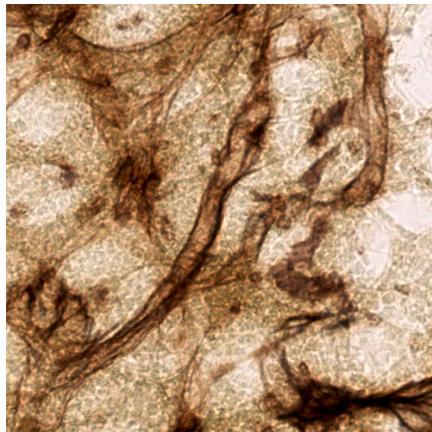
(a) Original (after slide-wide single-resolution registration)



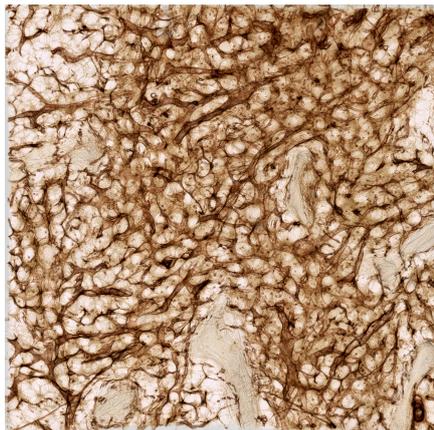
(b) A detail from (a)



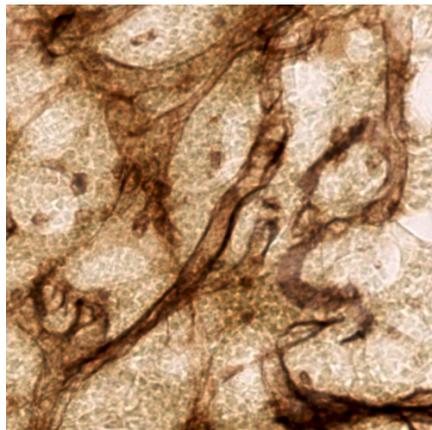
(c) ROI-based single-resolution registration, shows almost no effect in volume rendering, used as input for our method



(d) Single-resolution registration, a detail from (c)



(e) Our method improves the overlay



(f) Our method, a detail from (e)

Figure 7: Stack of 30 human bone marrow sections single-stained for CD34 and CD141 (both in brown), showing another ROI.

The stack shows capillaries (microvessels with small diameter) and sinuses (microvessels with large diameter). Both microvessels form a network with smooth transitions from capillaries to sinuses. Sinuses may form round or flattened structures. It is impossible to represent blood microvessels in such length without registering a series of sections.

A repeated application of single-resolution registration to a) would not show a better result, as it only operates on larger key points. These, however, were already aligned in (c). We need to align smaller key points with our method to yield a difference (e). These results are clearly visible in selected details (b), (d), and (f): notice the blurry blood vessel contour in d) and how it is more pronounced in f).

Figures b), d), f) where shown in a smaller magnification in the main paper as Fig. M3. Fig. 6 also shows the same ROI. Our registration method here uses *four* levels to produce an even better stack than in Fig. 5. Fig. M10 (main document) is similar to (e), but shows another ROI, a larger view of Fig. 5.

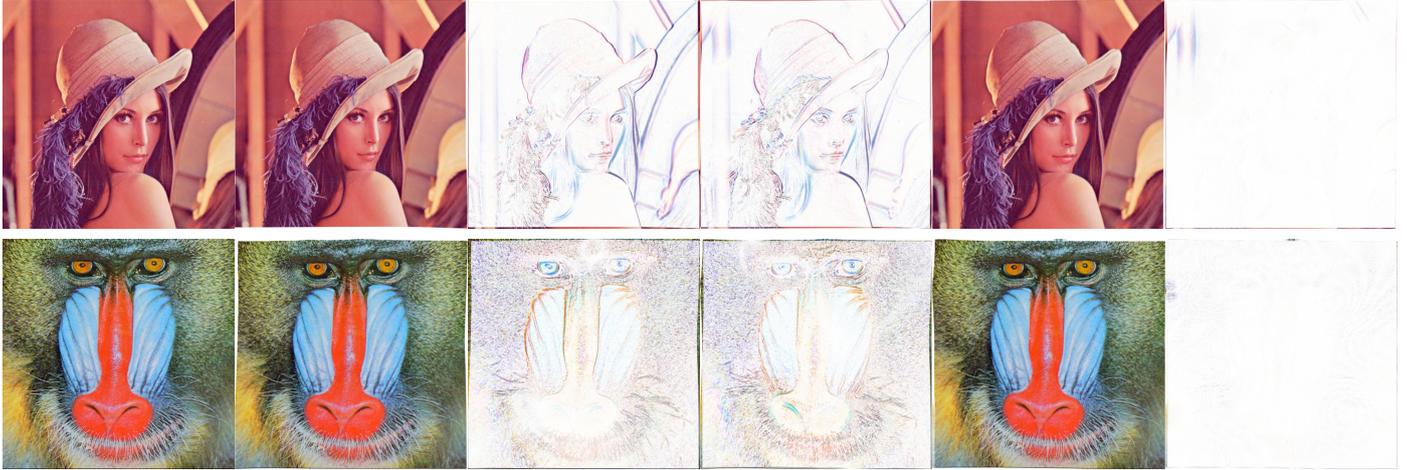


Figure 8: Registration of the standard images. Top line: Lena image, bottom line: mandrill image, both from the USC-SIPI image database. From left to right: original image, a distorted image used for registration, the difference between them, the difference between two differently distorted images used as input for registration, a registered image using our approach, the difference between two registered distorted images.

The color-coded optical flow for single-stained spleen specimen (Fig. 10) appeared quite dark for rigid-only transform (b) and our method (e) which is good, Ulrich et al. (c) and *elastix* (d) appeared somewhat lighter, while input overlay was the brightest (i.e. the worst). This corresponded well with Table M2.

In double-stained spleen analysis in Figure 11, *elastix* (d) and our method (e) were dark, where Ulrich et al. (c) showed some colored spots. These were almost everywhere in input overlay (a) and rigid-only transform (b), though in a different manner. This visual did not quite correspond with the numerical findings of Table M2, but the probable culprit was the difference between single-channel and color data (*elastix* ran on blue channel only), the distribution of motion magnitudes, or even the dimension of the filter kernels in the implementation of dense optical flow we used as detailed in the main document (Sec. M5.3.2).

Bone marrow specimen showed a very bright picture of optical flow between two consecutive images (Fig. 12) in input overlay (a) and *elastix* (d). Images for rigid-only transform (b), method of Ulrich et al. (c), and our method (e) were quite dark. The numerical evaluation in Table M2 stated that Ulrich et al. outperformed our method by a relatively small margin, but it also stated that rigid-only transform was worse than *elastix*, which contradicted with the visuals. Judging from visuals only (i.e. Fig. 12 c) and e)) we would deem these two methods as approximately equal.

At this point we can only re-state the assumption from the main paper that the averaging window of the dense optical flow method used was larger than the finest-grain improvements in capillaries and similar small structures. Visual inspection and results of 3D reconstruction (Figs. 6–7) clearly indicated that the result of Ulrich et al. was less usable than the result of our registration, which contradicts Table M2.

### 2.2.3. Jaccard Measure (Section M5.3.3, Table 1)

Table 1 extends the Table M3 from the main text. Fig. 13 shows the results of thresholding for double-stained spleen spec-

imen (top) and bone marrow specimen (bottom).

We chose the threshold value quite high to obtain distinctive values; the Jaccard measure for our method applied to single-stained spleen specimen reaches 1.0 at threshold 100. The stainings we used were quite distinctive, thus we chose the threshold 220 for the same specimen in Table 1.

In Table 1 our method outperformed all other methods used. Both the method Ulrich et al. and *elastix* were worse than baseline for spleen, single-stained. In fact, *elastix* performed not so well against this measure, it was consequently below the baseline for all specimens.

This table differs from Table M3 in the main text in the last line: we experimented with a larger threshold on bone marrow specimen. Our method showed again its superiority w.r.t. this measure. As for *elastix*, we obtained in this setting an absurdly low value which we did not include in the table.

### 2.2.4. Spleen, Single-Stained (Fig. 14)

Visually, the SSD did not change a lot between rigid-only registration, *elastix*, and our method; ours might be a little clearer. The numerical values, however, showed that our method displayed a better mean value. Our method was 14.5% better.

ZSAD for our method in single-stained spleen was visually much clearer and of less intensity. (For the sake of presentation we even had to increase brightness and contrast in all ZSAD images.) The numerical values showed that the ZSAD mean was the best in our method, with 22.7% improvement.

MI showed a slight increase between rigid-only registration and our method. The value for *elastix* was formally better, but it was not really comparable as *elastix* used only the red channel of the images. Thus we evaluated the information increase for this color channel only. (Section 2.2.5 discusses an experiment with grayscale-based *elastix* and grayscale-based MI on our registration to obtain comparable values.)

The GCC values—basically, the “movement” in the image pair—improved in all cases compared to the original overlay.

Structural similarity

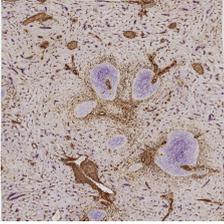
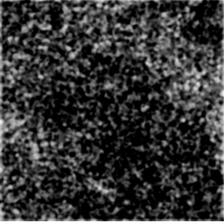
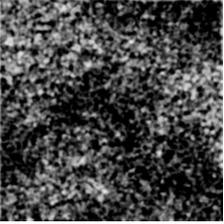
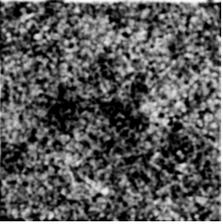
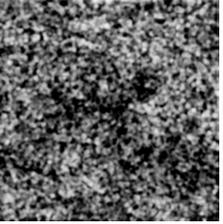
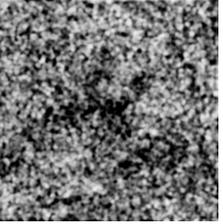
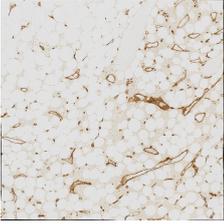
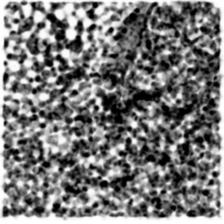
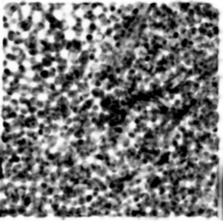
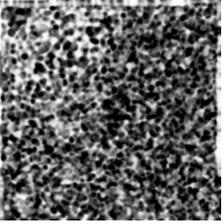
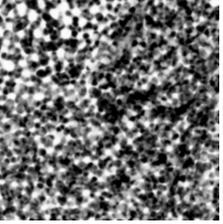
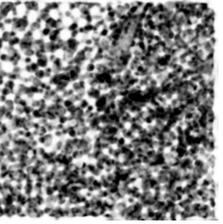
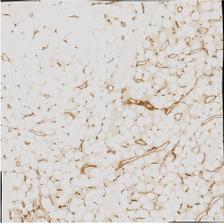
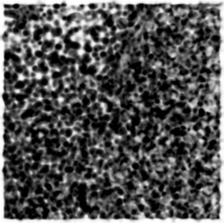
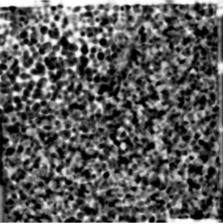
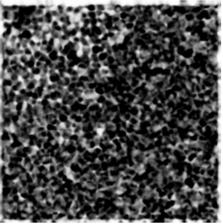
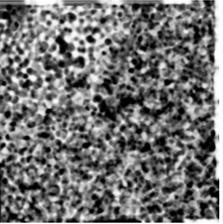
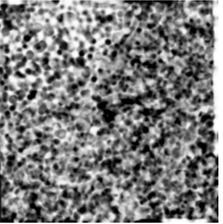
Registered image	SSIM for				
					
(a) Spleen, single-stained	0.556 (b) input	0.746 (c) rigid	0.730 (d) <i>elastix</i>	0.706 (e) Ulrich et al.	<b>0.808</b> (f) our method
					
(g) Spleen, double-stained	0.180 (h) input	0.252 (i) rigid	0.343 (j) <i>elastix</i>	0.377 (k) Ulrich et al.	<b>0.443</b> (l) our method
					
(m) Bone marrow, 3k	0.470 (n) input	0.480 (o) rigid	0.380 <sup>†</sup> (p) <i>elastix</i>	0.438 <sup>†</sup> (q) Ulrich et al.	<b>0.497</b> (r) our method
					
(s) Bone marrow, 4k	0.314 (t) input	0.351 (u) rigid	0.263 <sup>†</sup> (v) <i>elastix</i>	0.432 (w) Ulrich et al.	<b>0.474</b> (x) our method

Figure 9: Structural similarity index (SSIM). This is an extended version of Figure M12 in the main text, showing SSIM for spleen, single-stained (top row, (a–f), magnification factor 3.0), spleen, double-stained (second row, (g–l), magnification 12.08×), bone marrow (3k side, two levels of our method, third row, (m–r), magnification 20.05×), bone marrow (4k side, four levels of our method, bottom row, (s–x), magnification 15.01×) specimens. The left column, (a), (g), (m), (s), shows a registered image, all other columns show SSIM images and SSIM values: the second column (b), (h), (n), (t) shows SSIM for input data, third one, (c), (l), (o), (u), shows SSIM for rigid-only registration, fourth and fifth columns show state of the art: (d), (j), (p), (v) features *elastix*, (e), (k), (q), (w) shows Ulrich et al., sixth and last column, (f), (l), (r), (x), shows our method.

The higher the value and the lighter the image, the better. **Bold** shows the best value, *italics* with † denotes decrease from baseline.

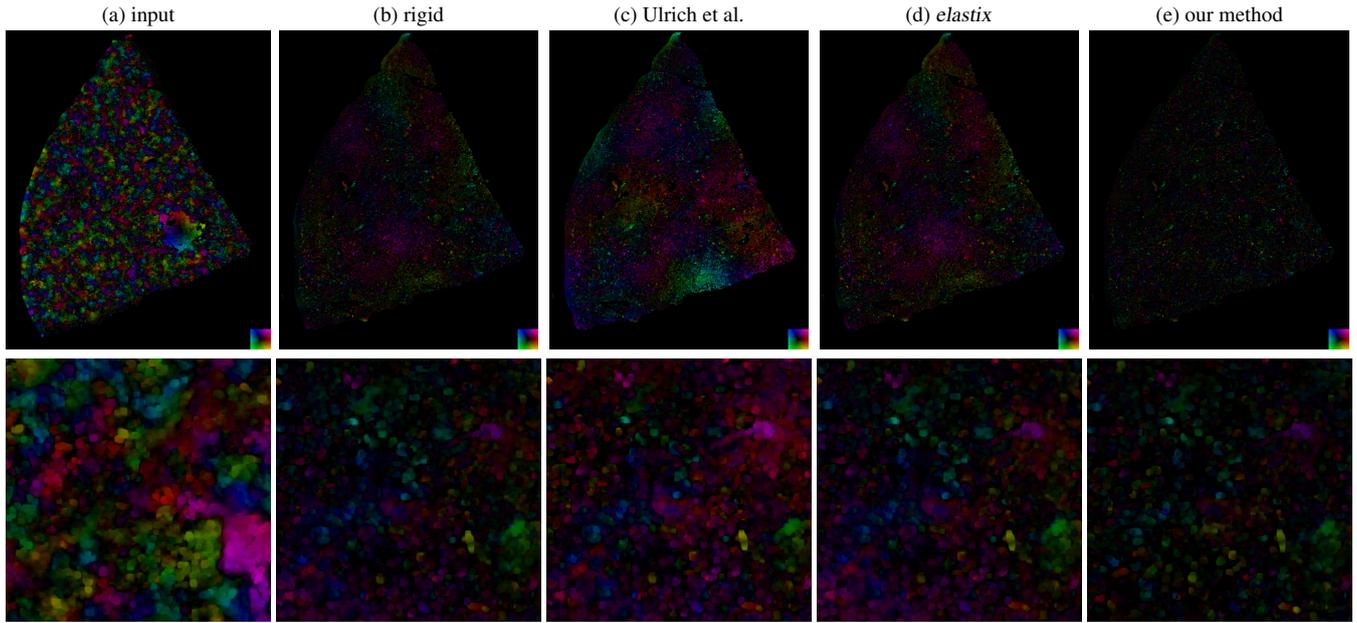


Figure 10: Color-coded optical flow as displacement angle and magnitude for single-stained spleen specimen.

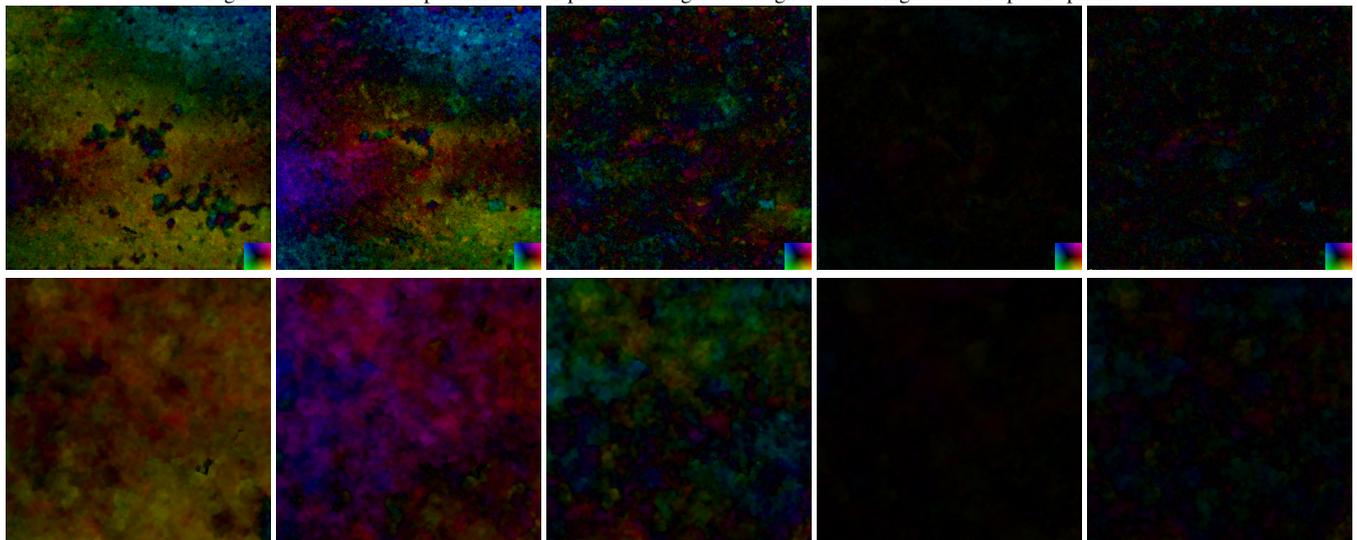


Figure 11: Color-coded optical flow as displacement angle and magnitude for double-stained spleen specimen.

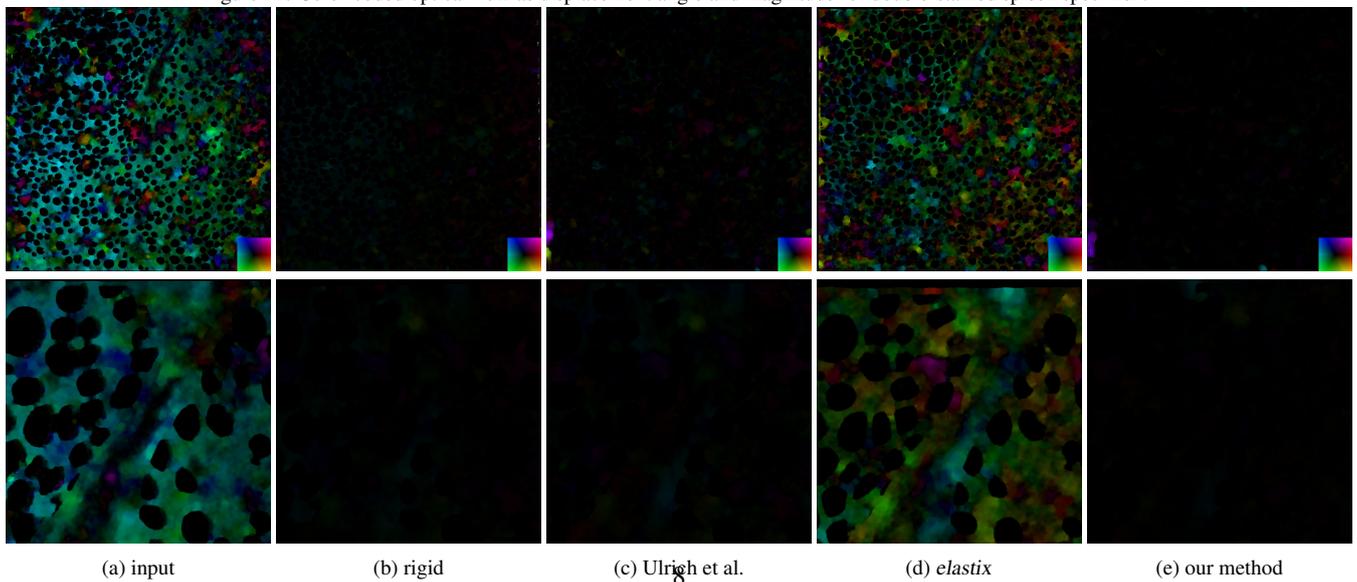


Figure 12: Color-coded optical flow as displacement angle and magnitude for double-stained spleen specimen. From left to right: input, Ulrich et al., *elastix*, our method. Top row shows whole slice, bottom row shows a crop. The smaller the intensity of the image, the better.

Table 1: Jaccard measure for all three specimens. The values were computed channel-wise where applicable, we show here the value for the best channel. This is the extended version of Table M3.

We denote the best overall value in **bold**, the values worse than baseline are marked with *emphasis* and †, values with possible compatibility problems (single-channel vs. best of three RGB channels) are marked with \*. The value marked with ¶ was too low (< 0.06558) due to a too high threshold.

The values are scaled between 0 and 1.0, the higher the better.

Specimen / Registration	Size	Levels	Threshold	Method				
				input	rigid	Ulrich et al.	<i>elastix</i>	ours
Spleen, single-stained	8000	4	220	0.9701	0.9785	0.9762 <sup>†</sup>	0.9774 <sup>*,†</sup>	<b>0.9829</b>
Spleen, double-stained	5032	4	100	0.8622	0.8755	0.8917	0.8158 <sup>*,†</sup>	<b>0.9062</b>
Bone marrow	4032	4	150	0.9583	0.9562 <sup>†</sup>	0.9612	0.8083 <sup>*,†</sup>	<b>0.9631</b>
			220	0.6744	0.6905	0.7169	-¶	<b>0.7224</b>

Our method delivered the best average value overall: 0.64 vs. 0.80, i.e. a 25% improvement.

Our method showed best results for all quality measures up to MI increase, where the data was not quite comparable.

### 2.2.5. Spleen, Double-Stained (Fig. 15)

The SSD images showed basically noise, however, in case of *elastix* and our method one might find some stationary points, especially with our method. This corresponded to best mean (and also median, not shown) improvements.

ZSAD was much more informative for this specimen: the "noise" was greatly reduced, the image for our method looked less noisy and also visibly darker. Numerically this meant that our method delivered the best results. The mean value of ZSAD for *elastix* showed roughly 4% quality decrease, while our method had a 21% better mean ZSAD value than rigid-only registration. The median for our method (not-shown) was 24% better.

We obtained an MI increase of 195% which is the best value across the methods tested. However, the value delivered by *elastix* (183%) was not really comparable to our results. The reason for this was that mutual information increase (i.e. a quotient of mutual information in the registered image pair and in original image pair) was computed for RGB images in case of rigid-only transform and our method in Figure 15, but only the blue channel was used in both image pairs for *elastix*. This was the same image channel *elastix* operated on in Fig. M11. If we used only the blue channel for our method, we would obtain 176.48% increase. However, this would not be fair towards our method as it did not operate on separate color channels. If we compared the intensities only (and run *elastix* on intensity images), *elastix* would outperform our method by 3.6%.

The gradient cross-correlation (GCC) showed some darker areas in case of our method, these might guide us to perceive the whole image as more dark, esp. when compared with the original, unregistered image pair or with the rigid-only method. As for mean values, the improvement of our method translated to 12.3%. (The non-shown maximal values improved up to 61%.)

Our method showed better results than *elastix* for *all* measures on double-stained spleen specimen, however the MI values were not quite comparable.

### 2.2.6. Bone Marrow, Two Levels (Fig. 16)

Here we compare the method of Ulrich et al., a non-multi-resolution feature-based registration method, to our approach w.r.t. bone marrow images at two levels of multi-resolution registration. We also show the unregistered overlay and rigid-only registration as intended baselines.

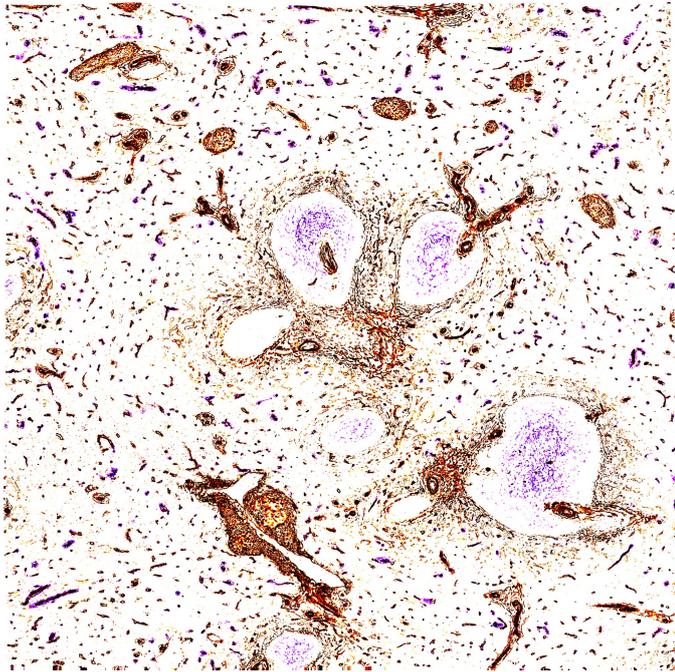
The sum of squared differences (SSD) for bone marrow images basically showed noise everywhere besides the interior of large unstained cells (adipocytes, light gray blobs in the microscopic images) and the non-stained interior of a blood vessel. Still, our method showed the best mean value (by a narrow margin) and was equal to the older method in median and maximum (not shown).

ZSAD quite clearly showed the edges of the blood vessels in a light color with an apparent matching problem, e.g. in rigid-only registration (h). These values seemed to increase for the method of Ulrich et al. (i) and decrease for our method (j) which led to the best mean value (and also maximum, not shown) in our case and to a ZSAD-quality decrease for the method of Ulrich et al. The difference between mean ZSAD for Ulrich et al. and our method was about 10%.

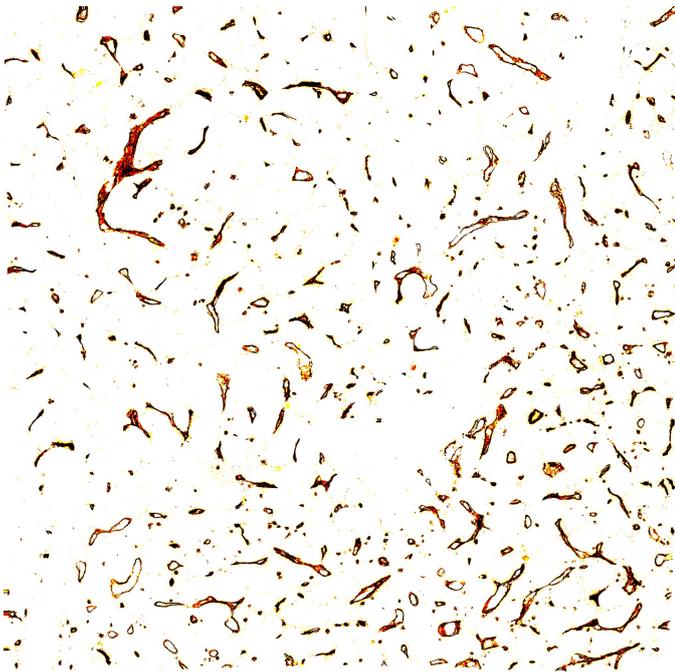
As we compared MI for RGB images, all values were fully comparable. Unfortunately, *all* registrations of the ROI were worse than the initial coarse-grained registration ("original"). Our method posed a 21% improvement over Ulrich et al., which was, however, still not enough to come over the baseline. We have no plausible explanation for this behavior.

The GCC images showed quite nicely how the gradient cross-correlation values in our method fade on a larger surface to darker gray, the same color as the distinctively visible large unstained cells. These images were gamma-corrected with a quite high value of  $\gamma = 6$  for presentation sake, so that dark gray was a rather low value. The numerical data for the full and gamma-uncorrected ROIs confirmed this. Our method was best in both directions for the mean value.

If we would simply sum up how often we obtained the best values for each quality measure, we obtain would a score of our method to the method of Ulrich et al. of 5 to 1. We ignored here MI, as both methods performed non-ideally w.r.t. this measure. Still, the MI value for our method was 21% better than that of Ulrich et al. We discussed the non-ideal performance of optical flow as a quality measure above.



(a) Spleen, double-stained



(b) Bone marrow

Figure 13: The pixels occurring in numerator or denominator of the Jaccard measure shown for: (a): a ROI in the double-stained spleen specimen at threshold 100, (b): a ROI in the bone marrow specimen at threshold 150. In (a) the blue staining shows additional information, the brown staining mirrors the microvascular system. In (b) the pixels shown show quite nicely the stained blood vessel walls. Hence the Jaccard measure compares the performance of the registration methods exactly at the point of our interest.

*Spleen, single-stained*

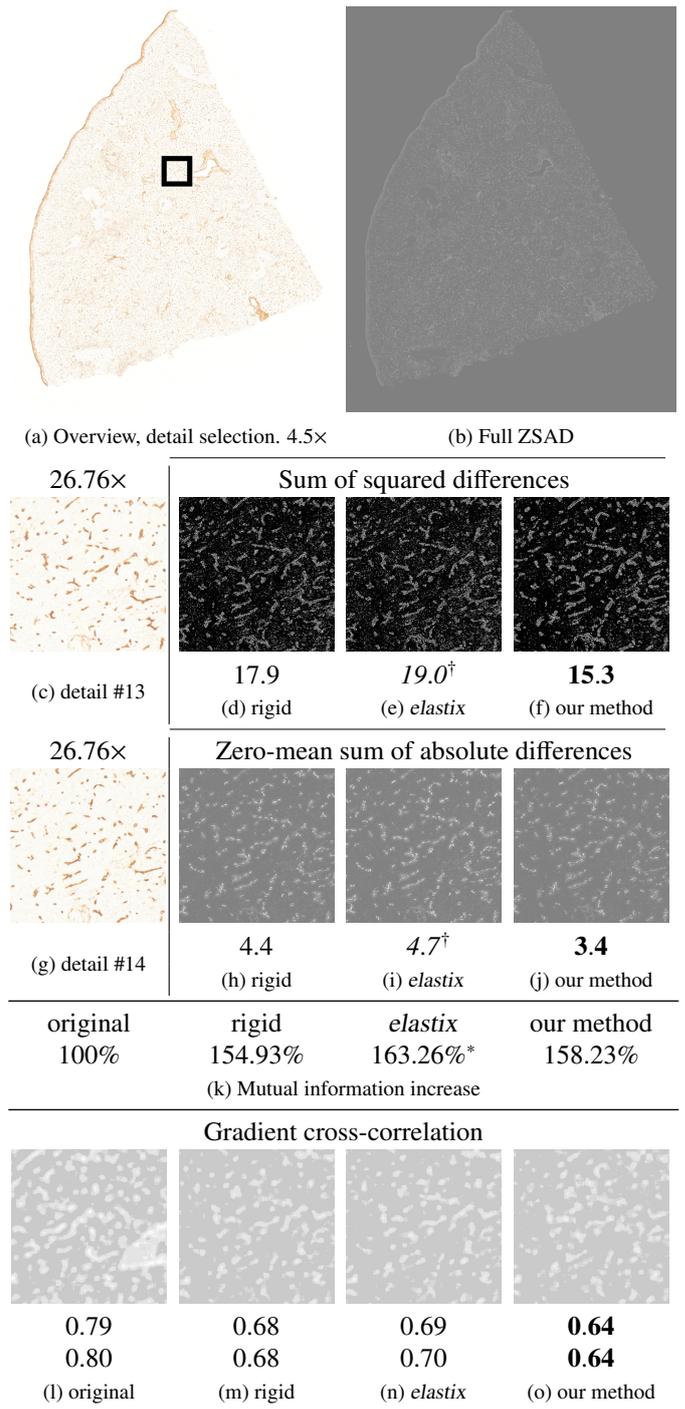


Figure 14: Quality measures for single-stained spleen, full section (same as Fig. 4). (a) Detail selection. (b) Zero-mean sum of absolute differences (ZSAD) measure for the whole image. Brightness and contrast for ZSAD images was corrected with +50. (c) and (g) The detail of the first and second sections. Magnification factor 26.76. (d–f) sum of squared differences (SSD). (h–j) Zero-mean sum of absolute differences (ZSAD). (k) Mutual information (MI) increase. (l–o) Combined cross-correlation of gradient approximation (GCC) with gamma enhanced to +6. The values below the images show the mean values of full image-wide quality measure. In case of GCC the first (second) line states the mean of derivatives in the direction of the  $x$  ( $y$ ) axis.

Aside from (k), the smaller the values and the darker the images the better. **Bold** is best, *italics* with <sup>†</sup> denotes decrease from the baseline, \* marks not comparable values.

*Spleen, double-stained*

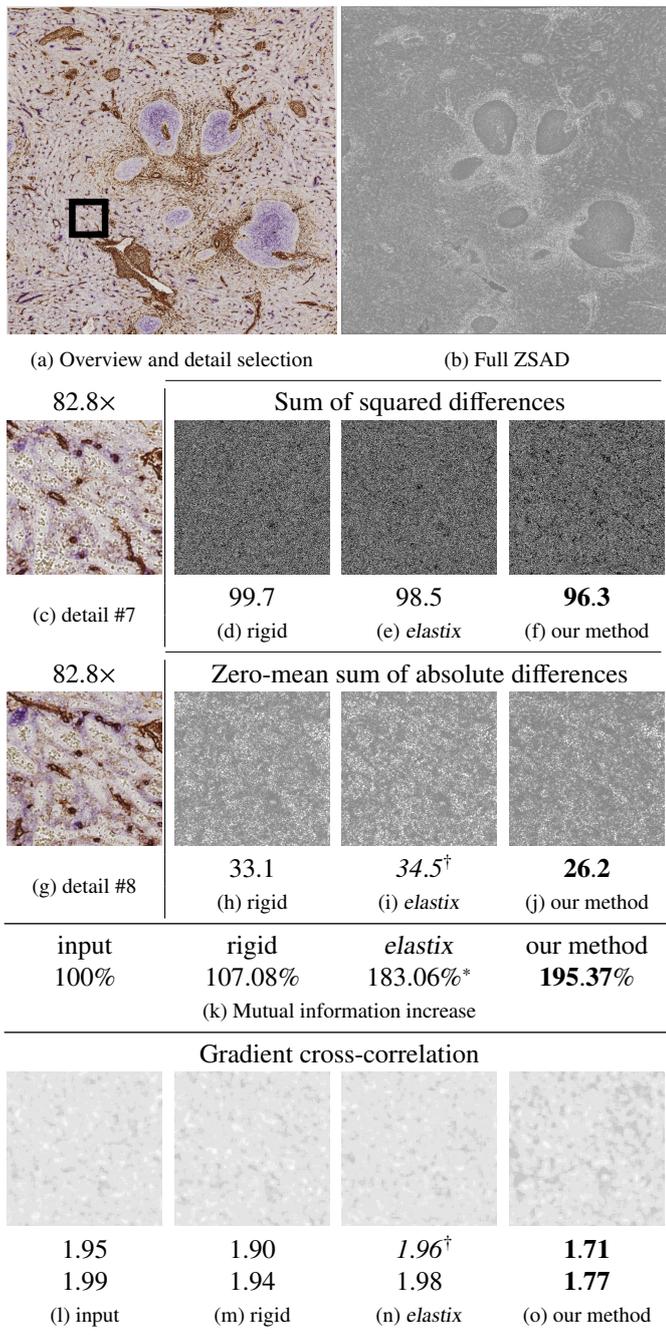


Figure 15: Quality measures for double-stained spleen (same ROI as in Fig. M8.) (a) Detail selection. The ROI is displayed at magnification 17.53. (b) Zero-mean sum of absolute differences (ZSAD) for full ROI. (See (h–j) for crops.) ZSAD images were enhanced with brightness and contrast +50. (c, g) The details of the first and second sections, displayed at 82.8x. (d–f) Sum of squared differences (SSD). (h–j) Zero-mean sum of absolute differences (ZSAD). (k) mutual information (MI) increase with original image pair as a baseline. Note that *elastix* (marked with asterisk) shows data for the blue channel only, but our method produces RGB images. (l–o) Combined cross-correlation of gradient approximation (GCC). Images are +6 gamma-corrected. We show mean values of full image-wide quality measures. For GCC we state both the mean of the derivative in the direction of the  $x$  axis (first line) and  $y$  axis (second line). Aside from (k), the smaller the values and the darker the images the better. **Bold** shows the best value, *italics* with <sup>†</sup> denotes decrease from baseline, asterisk (\*) marks incomparable values.

*Bone marrow, 2 levels, 3k*

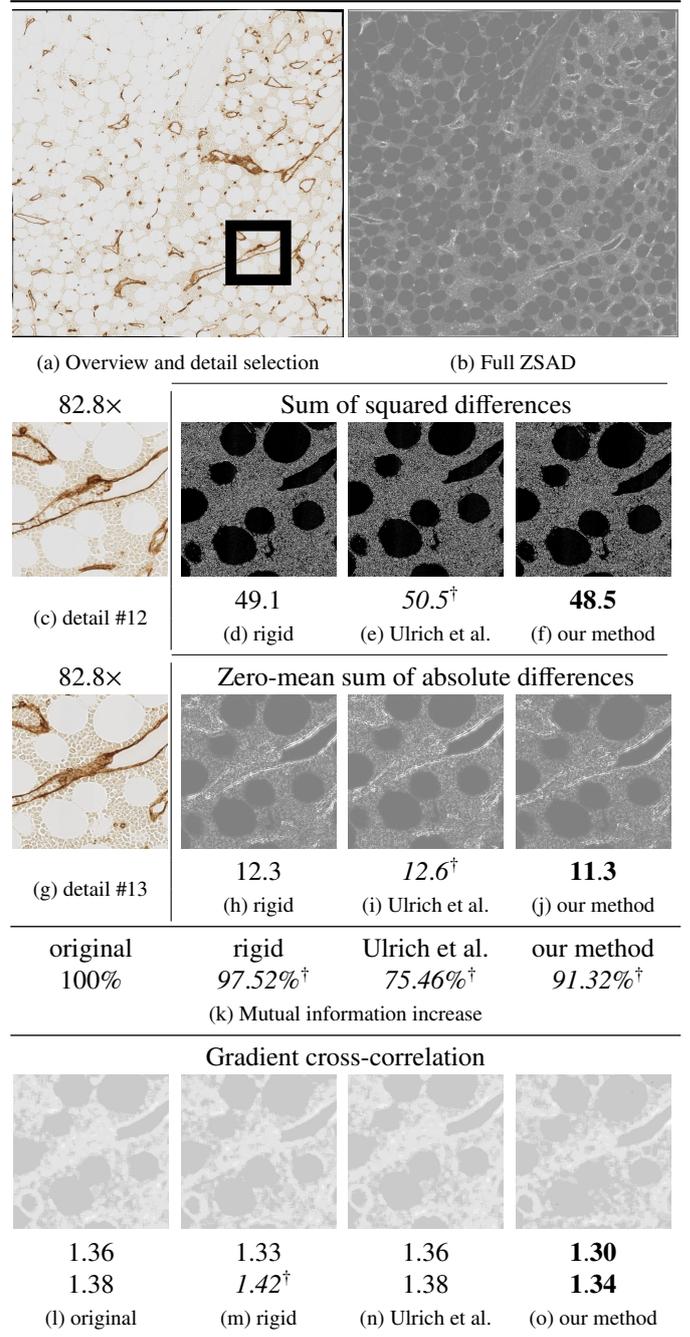


Figure 16: Quality measures for bone marrow specimen, same ROI as in Fig. 5. (a) The detail selection at 29.09x, (b) Complete ZSAD quality measure. (See (h–j) for crops.) (c) The detail of the first section, at 82.8x. (d–f) Sum of squared differences (SSD). (g) The detail of the second section. (h–j) Zero-mean sum of absolute differences (ZSAD). We enhanced the brightness and contrast. (k) Mutual information (MI) increase from baseline to the method from the corresponding column. (l–o) Combined cross-correlation of gradient approximation (GCC). These images are enhanced with brightness and contrast +50. We state the mean values for each image-wide quality measures. For GCC the first (second) line states the mean of derivatives in the direction of the  $x$  (second) axis. Aside from (k), the smaller the values and the darker the images the better. **Bold** shows the best value, *italics* with <sup>†</sup> denotes decrease from baseline.

### 2.2.7. Bone marrow, Four Levels (Fig. 17)

This discussion is based on four levels of our registration method, the input ROI is  $4k \times 4k$ , everything else is as in Section 2.2.6.

Our method showed the best overall mean SSD value. The result of Ulrich et al. was marginally *worse* than the input overlay. In a contrast, our method managed to improve over the same overlay by 4.7%. The SSD images were not very informative, it is basically noise outside of larger cells. Still we saw some minor improvements (dark areas) in SSD for our method, that were not present in other versions.

ZSAD showed the contours of the blood vessels with an apparent matching problem quite clearly in a light color, e. g. in (e) and (f). These values seemed to decrease both for Ulrich et al. (g) and for our method (h), however our method showed less bright spots that would correspond to a mismatch. This led to the best mean value (and also maximum and median, not shown) The difference between mean ZSAD for the method of Ulrich et al. and our method was almost 20%.

As we compared MI for RGB images, all values were this time fully comparable. Here, the single-resolution registration was marginally ( $< 0.7\%$ ) better than our method. Both showed improvements over rigid-only registration and input images.

The reason for the not so well performance of our method w. r. t. MI might lie in the specimen, there were a lot small cells that basically constitute background noise. Visual comparisons of reconstructions based on method of Ulrich et al. and on our method (Figs. 6–7) underlined the superiority of our method in the representation of small capillaries.

The GCC images showed how the gradient cross-correlation values in our method faded on a larger surface to darker gray, the same color as the distinctively visible large unstained cells. We also saw in Fig. (m) some black spots not present in other images. Notably, even dark gray was a quite low value ( $\gamma = 6$  was applied). The numerical data on full and gamma-uncorrected ROIs confirmed this. In the direction of  $x$  ( $y$ ) axis, Ulrich et al. (l) improved 10.6% (6.6%) over input overlay, our method (m) showed 17.2% (16.1%) improvement, and *elastix* (not shown) was 9.2% (7.6%) better than its single-channel input. Our method (m) was best in both directions.

If we simply summed up how often we obtained the best values for each quality measure, our method would clearly outperform the method of Ulrich et al.

### 3. Online Materials

Further supplementary material, including input data, full-scale volume renderings, full-scale images of quantitative evaluations, Windows executable of our application, and *elastix* parameter files, is available under <https://gdv-server.inf.uni-bayreuth.de/gdvcloud/index.php/s/NnSov0065n9Gp01>.

The supplementary video for this paper is uploaded as supplementary material alongside this document, but it is also available under <https://gdv-server.inf.uni-bayreuth.de/gdvcloud/index.php/s/s13s5I1yf4BrDe7>.

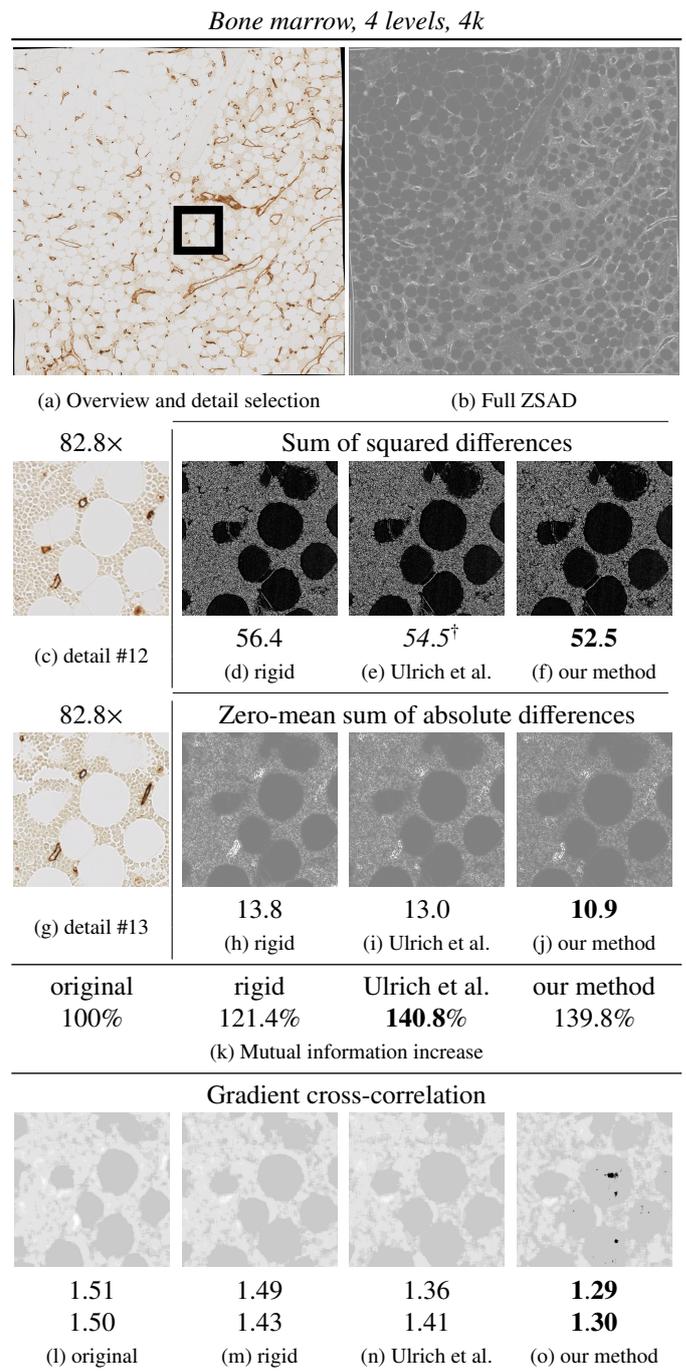


Figure 17: Quality measures for bone marrow specimen, extended version of Fig. M15. (a) The detail selection at 29.09×, (b) Complete ZSAD quality measure. (See (h–j) for crops.) (c) The detail of the first section, at 82.8×. (d–f) sum of squared differences (SSD). (g) The detail of the second section. (h–j) Zero-mean sum of absolute differences (ZSAD). We enhanced the brightness and contrast. (k) Mutual information (MI) increase from baseline to the method from the corresponding column. (l–o) Combined cross-correlation of gradient approximation (GCC). These images are enhanced with brightness and contrast +50.

We state the mean values for each image-wide quality measures. For GCC the first (second) line states the mean of derivatives in the direction of the  $x$  ( $y$ ) axis. Aside from (k), the smaller the values and the darker the images the better. **Bold** shows the best value, *italics* with <sup>†</sup> denotes decrease from baseline.