

Registration of Segmented Histological Images using Thin Plate Splines and Belief Propagation

Jan Kybic

Czech Technical University in Prague, Czech Republic

ABSTRACT

We register images based on their multiclass segmentations, for cases when correspondence of local features cannot be established. A discrete mutual information is used as a similarity criterion. It is evaluated at a sparse set of location on the interfaces between classes. A thin-plate spline regularization is approximated by pairwise interactions. The problem is cast into a discrete setting and solved efficiently by belief propagation. Further speedup and robustness is provided by a multiresolution framework. Preliminary experiments suggest that our method can provide similar registration quality to standard methods at a fraction of the computational cost.

1. INTRODUCTION

We present a method to register images with very different appearance and local structure. Stained consecutive histological slices (Fig. 1) are a prime example. The slices need to be registered to fuse information from the different stains.¹ The misalignment is due to the cutting and acquisition process. The different stainings results not only in different colour but also in different elements of the tissue being visible. In addition, the small scale details are different between the slices because of the slice thickness. Finally, the original images can be very large (10^8 pixels).

We propose to register the images based on their segmentations, assuming that a membership of pixels in a small number of classes is the only common information shared between the images and useful for registration. The algorithm is designed to be fast by considering only boundaries between classes, approximating regularization by automatically identified pairwise interactions, and applying a multiresolution belief propagation to the discretized non-linear optimization problem.

Many different methods have been applied to align histology slices, mostly based on minimizing a pixel-based similarity criteria,¹⁻³ but also feature-based methods⁴ or local search methods.^{5,6} Binary segmented images have been registered through descriptors^{7,8} or level sets.⁹ The number of equally classified pixels is maximized in.¹⁰

2. METHODS

The fixed and moving images are represented by two pixel-level 2D segmentations with L_F and L_G classes, respectively. Since classes are typically spatially compact, we expect that sufficient information relevant for finding the correspondences between the two images can be obtained in the vicinity of a sparse set of keypoints $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ located around the edges of the segmentation. where $N \ll M$ and M is the number of pixels in the image. The task is to find a displacement function $\mathbf{g} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\mathbf{g}(\mathbf{x}) = [g_x(\mathbf{x}) \ g_y(\mathbf{x})]^T$ that minimizes the following criterion:

$$\sum_{i=1}^N D_i(\mathbf{y}_i) + \lambda_0 \mathcal{R}(\mathbf{g}) \quad (1)$$

where $\mathbf{y}_i = \mathbf{g}(\mathbf{x}_i)$ are keypoint displacements and D_i measures the local image dissimilarity of the neighborhood of \mathbf{x}_i in the fixed image and $\mathbf{x}_i + \mathbf{y}_i$ in the moving image. Note that while D_i encodes the image information and should be represented accurately, the regularization \mathcal{R} is far less important. Its ‘only’ role is to keep \mathbf{g} smooth. We will therefore take the liberty of approximating it rather crudely later on in order to obtain a fast algorithm.

E-mail: kybic@fel.cvut.cz

2.1 Dissimilarity criterion

For each keypoint \mathbf{x}_i , we compare labels l_F from the fixed image segmentation inside a rectangular window $\Omega(\mathbf{x}_i)$ of size $h \times h$ pixels centered around \mathbf{x}_i , with labels l_G of the moving image segmentation centered around $\mathbf{x}_i + \mathbf{y}_i$. To cope with a general stochastic relationship between labels in the two images, we use a discrete mutual information

$$D_i(\mathbf{y}_i) = - \sum_{m=1}^{L_F} \sum_{n=1}^{L_G} p_{m,n} \log \underbrace{\frac{\bar{p}_{m,n}}{\bar{p}_m \bar{p}_n}}_{\psi(m,n)} \quad (2)$$

$$= - \sum_{\mathbf{x} \in \Omega(\mathbf{x}_i)} \psi(l_F(\mathbf{x}), l_G(\mathbf{x} + \mathbf{y}_i - \mathbf{x}_i)) \quad (3)$$

where \bar{p}_m, \bar{p}_n are marginal probabilities. The class co-occurrence probabilities $\bar{p}_{m,n}$ are calculated over the whole image, while $p_{m,n}$ are calculated inside the window $\Omega(\mathbf{x}_i)$. This implicitly assumes that the relationship between class labels in the two images is not spatially dependent. On the other hand, and unlike standard mutual information on intensities, it permits to reliably assess the local similarity even for very small window size h .

2.2 Thin-plate splines

Choosing the smoothness criterion $\mathcal{R}(g) = \mathcal{R}(g_x) + \mathcal{R}(g_y)$ to penalize the second derivatives,

$$\mathcal{R}(g) = \int \left(\frac{\partial^2 g}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 g}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 g}{\partial y^2} \right)^2 d\mathbf{x} \quad (4)$$

is a natural choice because of its lack of parameters, and rotational, translational, and scale invariance.¹¹ It leads to the well known thin-plate splines (TPS),¹² which we review here shortly. Each of the functions g_x, g_y (note that they are independent and thus for simplicity we will treat only the scalar case here) is uniquely determined by the N values $\mathbf{v} = [v_1 \dots v_N] = [g(\mathbf{x}_1) \dots g(\mathbf{x}_N)]$ and can be written as

$$g(\mathbf{x}) = [\mathbf{x} \ 1] \mathbf{a} + \sum_{i=1}^N w_i \varphi(\|\mathbf{x} - \mathbf{x}_i\|) \quad (5)$$

with $\varphi(r) = r^2 \log r$. The parameters $[\mathbf{w} \ \mathbf{a}]^T$ are found by solving a linear system of equations

$$\mathbf{L} \begin{bmatrix} \mathbf{w} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{v} \\ \mathbf{0} \end{bmatrix} \quad \text{with} \quad \mathbf{L} = \begin{bmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \quad (6)$$

$K_{ij} = \varphi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, and $\mathbf{P}_i = [1 \ x_i \ y_i]$. There is a closed-form expression¹³ for the smoothness

$$\mathcal{R}(g) = \xi \mathbf{w}^t \mathbf{K} \mathbf{w} = \xi \mathbf{v}^t \mathbf{Q} \mathbf{v} \quad (7)$$

where the smoothness matrix \mathbf{Q} is the top left $N \times N$ part of \mathbf{L}^{-1} from (6) and ξ is a known constant. This comes from the fact that the \mathbf{w} are the Lagrange multipliers with respect to the measurements \mathbf{v} . The variational problem (1) is therefore simplified to the following multidimensional minimization with respect to $\mathbf{y} = [\mathbf{v}_x \ \mathbf{v}_y] = [g_x(\mathbf{x}_1) \dots g_x(\mathbf{x}_N) \ g_y(\mathbf{x}_1) \dots g_y(\mathbf{x}_N)]$

$$\sum_{i=1}^N D_i(\mathbf{y}_i) + \lambda \mathbf{y}^T \bar{\mathbf{Q}} \mathbf{y} \quad \text{with} \quad \bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \quad (8)$$

where we have incorporated ξ into $\lambda = \lambda_0 \xi$.

2.3 Large number of keypoints

The bottleneck of evaluating (8) is the matrix inverse L^{-1} , needed to calculate Q , with computational complexity $O(N^3)$. This is acceptable for N up to about 100~1000. When more keypoints are needed, Q needs to be approximated. Simple techniques such as using a low rank approximation of K (suggested e.g. for Gaussian kernels¹⁴) do not work as they neglect small eigenvalues which have a large influence on the inverse matrix. It is possible to heuristically or randomly subsample the basis functions¹³ but there is no clearly justified strategy for which ones to keep. Instead, we have extended a technique by Wood¹⁵ which chooses the approximation space automatically. It starts by calculating a truncated spectral decomposition $\tilde{K} = U\Lambda U^T$ of rank k of the thin-plate spline matrix $K \approx \tilde{K}$. It turns out that K is reasonably well approximated even if we choose $k \ll N$ and that an approximate truncated SVD can be calculated quickly by randomized techniques¹⁶ with complexity $O(kN^2)$. We apply SVD again to the much smaller matrix $P^T U$, to find a basis Z of its null space of dimension $k - 3$. If we choose $w = UZ\tilde{w}$ from this null space, then the orthogonality condition $P^T w = 0$ from (6) is automatically satisfied for any \tilde{w} . Minimizing the squared approximation error

$$\sum_i \|v_i - g(\mathbf{x}_i)\|^2 = \left\| \mathbf{v} - [K \ P] \begin{bmatrix} \mathbf{w} \\ \mathbf{a} \end{bmatrix} \right\|^2 = \left\| \mathbf{v} - [U\Lambda Z \ P] \begin{bmatrix} \tilde{\mathbf{w}} \\ \mathbf{a} \end{bmatrix} \right\|^2 \quad (9)$$

with respect to \tilde{w} leads to a system of k linear equations

$$\tilde{L} \begin{bmatrix} \tilde{\mathbf{w}} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} Z^T \Lambda U^T \mathbf{v} \\ P^T \mathbf{v} \end{bmatrix}, \quad \tilde{L} = \begin{bmatrix} Z^T \Lambda^2 Z & Z^T \Lambda U^T P \\ P^T U \Lambda Z & P^T P \end{bmatrix} \quad (10)$$

which gives us the “best” reduced parameter vector

$$\tilde{\mathbf{w}} = M\mathbf{v}, \quad \text{with} \quad M = L_a Z^T \Lambda U^T + L_b P^T \quad (11)$$

where L_a, L_b are the top left and right blocks of \tilde{L}^{-1} . Note that the size of the matrix \tilde{L} is only $k \times k$.

Given that $\mathbf{v}^T Q \mathbf{v} = \mathbf{w}^T K \mathbf{w} = \tilde{\mathbf{w}}^T Z^T \Lambda Z \tilde{\mathbf{w}} = \mathbf{v}^T M^T Z^T \Lambda Z M \mathbf{v}$ for $\mathbf{w} = UZ\tilde{\mathbf{w}}$, we can see that the approximation of the smoothness matrix

$$Q \approx \tilde{Q} = M^T Z^T \Lambda Z M \quad (12)$$

can be calculated in time $O(kN^2)$.

2.4 Pairwise interaction approximation

With \tilde{Q} or its approximation known, the minimization (8) can be done with standard iterative multidimensional minimization techniques but it remains time consuming due to the non-linearity of D_i and the non-sparsity of Q . We therefore approximate \tilde{Q} by pairwise interactions

$$(1/\xi)\mathcal{R}(\mathbf{g}) = \mathbf{y}^T \tilde{Q} \mathbf{y} \approx \frac{1}{2} \sum_{1 \leq i, j \leq N} \omega_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 \quad (13)$$

which makes the result dependent only on pairwise displacement differences. It is clear that (13) could be made exact by adding suitable unary and constant terms. However, we assume these terms to be zero, making $\mathcal{R}(\mathbf{g})$ depend only on pairwise distances between points. This is motivated by the fact that \mathcal{R} is translation invariant.

Second, we sort the negative elements above the main diagonal of Q and keep only the αN largest ones, setting $\omega_{ij} = \omega_{ji} = -Q_{ij}$. Choosing $\alpha = 2$ makes each \mathbf{y}_i interact with four other points on the average, which was found to give adequate approximation of $\mathcal{R}(\mathbf{g})$ in most cases. The criterion to be minimized becomes

$$J(\mathbf{y}) = \sum_{i=1}^N D_i(\mathbf{y}_i) + \frac{1}{2} \lambda \sum_{(i,j) \in E} \omega_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 \quad (14)$$

where E is a set of αN undirected edges connecting interacting keypoints, i.e. $(i, j) \equiv (j, i)$. This reduces the complexity of evaluating the regularization part of (14) from $O(N^2)$ for (8) to $O(N)$. The advantage of our approach is that the topology and weights of the interaction graph is derived automatically from the data.

2.5 Belief propagation

We discretize (14) by allowing only integer positions \mathbf{y}_i and limiting the maximum displacement, $\|\mathbf{y}_i\|_\infty \leq d$. Then (14) can be efficiently minimized by belief propagation. Each node (keypoint) i sends in time t a message

$$\mu_{i \rightarrow j}^t(\mathbf{y}_j) = \min_{\mathbf{y}_i} \left(\omega_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 + D_i(\mathbf{y}_i) + \sum_{s \neq j} \mu_{s \rightarrow i}^{t-1}(\mathbf{y}_j) \right)$$

to all neighbors j , with $(i, j) \in E$. The $(2d + 1)^2$ values of the dissimilarity criterion D_i need to be precalculated. Each message is a matrix of $(2d + 1)^2$ real numbers, one for each possible displacement \mathbf{y}_j . It can be calculated in time $O(d^2)$,¹⁷ if $D_i(\mathbf{y}_i)$ is precalculated for the $(2d + 1)^2$ possible displacements \mathbf{y}_i .

We stop iterating when the decrease of J is smaller than a threshold. In our experience, very small number of iterations is needed, usually less than 10, so the time complexity of this step is $O(Nd^2)$. Then, the optimum solution is

$$\mathbf{y}_j^* = \arg \max_{\mathbf{y}_j} \left(D_j(\mathbf{y}_j) + \sum_i \mu_{i \rightarrow j}^t(\mathbf{y}_j) \right) \quad (15)$$

2.6 Multiresolution

We employ multiresolution in order to reduce the computational complexity $O(Nd^2)$ of precalculating the dissimilarity criteria D_i and the belief propagation itself when larger displacements d must be handled. We build a dyadic image hierarchy so that pixel (x, y) at level κ corresponds to pixels $\{2x, 2x + 1\} \times \{2y, 2y + 1\}$ at level $\kappa + 1$. Instead of subsampling, for each pixel we keep a number c_m of corresponding finest level pixels having class m . The probability in (3) is then calculated as $p_{m,n} = \sum_{m,n} c_m^F c_n^M / (\sum c_m^F \cdot \sum c_n^M)$. The parameters h and d are kept constant, so at coarse levels we take into account larger spatial features and allow longer maximum displacements than finer levels.

The registration at the coarsest level produces a displacement estimate \mathbf{y}^* , which is used as an initial estimate of \mathbf{y} for the next finer level, $\mathbf{y}_0^{\kappa-1} = 2\mathbf{y}^{*,\kappa}$, where the multiplication by 2 corresponds to the coordinate transformation. This is repeated recursively until the finest level, $\kappa = 0$.

The belief propagation is modified as follows: The messages $\mu_{s \rightarrow i}^{t-1} : \Upsilon_i \rightarrow \mathbb{R}$ where $\Upsilon_i = \{\mathbf{y}; \|\mathbf{y}_i - \mathbf{y}_i^0\|_\infty \leq d\}$ are extended to $\Upsilon_j \cup \Upsilon_i$ by setting all undefined values to $\max_{\mathbf{y} \in \Upsilon_i} \mu_{s \rightarrow i}^{t-1}(\mathbf{y})$. The result $\mu_{i \rightarrow j}^t$ is then restricted to Υ_j .

3. RESULTS

We show the results of registering adjacent histological slices of prostate¹ and rat kidney (Figure 1). The images of a typical size 1400×2050 were very quickly semi-automatically segmented by Ilastik* into four classes. As you can see, the segmentations are rather inconsistent. The segmentations were simplified by a mode filter of size 10 pixels. The keypoints were determined as points where two different classes touch; then they were recursively pruned so that their minimum distance was at least 50 pixels, yielding around 250 keypoints. Our method was run using 6 multiresolution levels, maximum displacement $d = 3$ pixels, window size $h = 10$ pixels, and regularization weight $\lambda = 50$.

As a baseline, we use a standard B-spline non-linear registration with cross-correlation similarity criterion, based on ITK.[†]

Using our unoptimized single-threaded Python+Cython[‡] implementation on a rather weak laptop computer, each level of the multiresolution pyramid took only a couple of seconds, with the whole registration taking about a minute. The most time-consuming part is the final image warping using thin-plate splines. In contrast, the multithreaded C++ ITK B-spline method took over an hour on the same, two-core machine.

From visual inspection, the results of both methods are acceptable, with slightly better alignment provided by the new method. The B-spline results can be improved by using a finer control grid at the expense of further increase of computation time.

*ilastik.org

†itk.org

‡python.org, cython.org

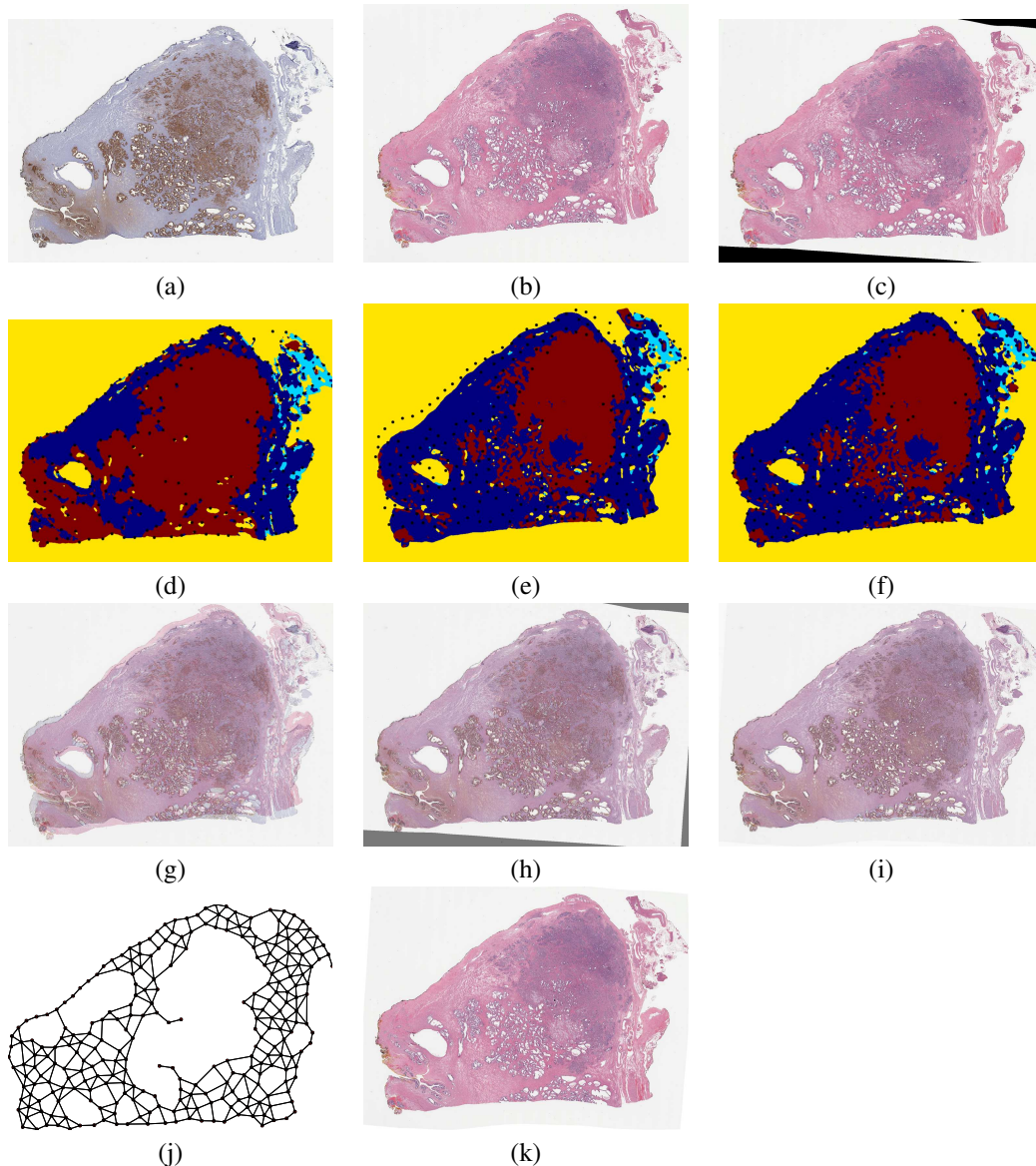


Figure 1. Histology slice of human prostate. Fixed image (a), moving image (b), and a warped moving image after registration (c). Segmentation of the fixed image (d) and moving image (e) with the initial keypoint positions. Keypoints in the moving image after registration (f). Overlay of the input images before registration (g), using the proposed method (h), and using ITK B-spline registration (i). The interaction graph (j) and the warped moving image after the ITK registration (k). (Images best viewed in color, available in the electronic version.)

4. CONCLUSIONS

We have presented a method for registering images based on their segmentations, useful for registering dissimilar images, when traditional pixel-based methods are too slow, or when feature-based methods fail to establish correspondences. The method does not assume any particular topology or relationship between the classes. It uses a number of novel techniques such as class-based mutual information similarity criterion, approximation of TPS regularization by data-derived pairwise interactions, and efficient multiresolution keypoint displacement optimization by loopy belief propagation. We also provide explicit formulas for low-rank thin-plate regression splines. The method can be used with other similarity criteria and regularizers, and can be coupled with fully automatic segmentation methods.

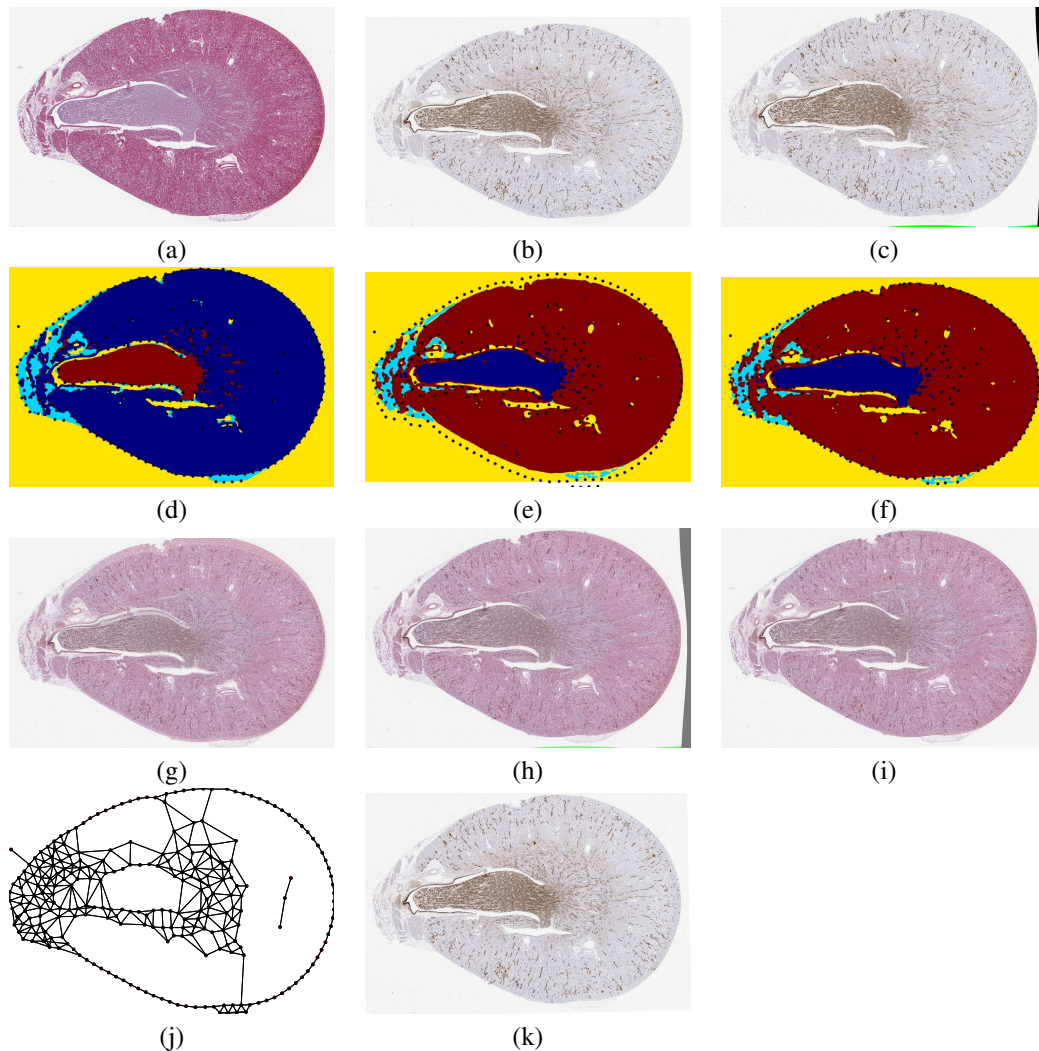


Figure 2. Histology slices of rat kidney. Fixed image (a), moving image (b), and a warped moving image after registration (c). Segmentation of the fixed image (d) and moving image (e) with the initial keypoint positions. Keypoints in the moving image after registration (f). Overlay of the input images before registration (g), using the proposed method (h), and using ITK B-spline registration (i). The interaction graph (j) and the warped moving image after the ITK registration (k). (Images best viewed in color, available in the electronic version.)

On our images, the method performs as good or better than standard B-spline registration in less than 1/10 of the time. By careful implementation, parallelisation, and by using hierarchical methods¹⁸ for TPS interpolation, we expect to gain an additional 10 ~ 100 fold speedup.

5. ACKNOWLEDGEMENTS

This work was supported by the Czech Science Foundation under project P202/11/0111. The data was provided by Flagship Bioscience and University of Minnesota. The ITK registration was programmed by Michal Buřta.

REFERENCES

- [1] Metzger, G. J. et al., "Development of multigene expression signature maps at the protein level from digitized immunohistochemistry slides," *PLoS ONE* 7, e33520 (03 2012).

- [2] Wirtz, S., Fischer, B., Modersitzki, J., and Schmitt, O., “Superfast elastic registration of histologic images of a whole rat brain for 3D reconstruction,” in [*SPIE Conference*], Fitzpatrick, J. M. and Sonka, M., eds., **5370**, 328–334 (May 2004).
- [3] Arganda-Carreras, I. et al., “Consistent and elastic registration of histological sections,” in [*CVAMIA: Computer Vision Approaches to Medical Image Analysis*], 85–95, Springer (May 2006).
- [4] Saalfeld, S., Fetter, R., Cardona, A., and Tomancak, P., “Elastic volume reconstruction from series of ultra-thin microscopy sections,” *Nature Methods* (June 2012).
- [5] Chakravarty, M. et al., “The creation of a brain atlas for image guided neurosurgery using serial histological data,” *NeuroImage* **30**(2), 359 – 376 (2006).
- [6] Hufnagel, H. et al., “Non-linear 2D and 3D registration using block-matching and B-splines,” in [*Bildverarbeitung für die Medizin 2005*], *Informatik aktuell*, 325–329, Springer Berlin Heidelberg (2005).
- [7] Domokos, C., Nemeth, J., and Kato, Z., “Nonlinear shape registration without correspondences,” *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(5), 943–958 (2012).
- [8] Flusser, J. and Suk, T., “A moment-based approach to registration of images with affine geometric distortion,” *IEEE Trans. Geosci. Remote Sensing* **32**(2), 382–387 (1994).
- [9] Droske, M. and Ring, W., “A Mumford-Shah level-set approach for geometric image registration,” *SIAM Appl. Math.* **2007** (2005).
- [10] Bollenbeck, F. and Seiffert, U., “Joint registration and segmentation of histological volume data by diffusion-based label adaption,” in [*ICPR*], 2440–2443 (2010).
- [11] Kybic, J., Blu, T., and Unser, M., “Generalized sampling: A variational approach. Part I — Theory,” *IEEE Transactions on Signal Processing* **50**, 1965–1976 (August 2002).
- [12] Bookstein, F., [*Morphometric Tools for Landmark Data: Geometry and Biology*], Cambridge University Press (1997).
- [13] Donato, G. and Belongie, S., “Approximate thin plate spline mappings,” in [*ECCV*], *ECCV '02*, 21–31, Springer-Verlag, London, UK, UK (2002).
- [14] Myronenko, A. and Song, X., “Point-Set Registration: Coherent Point Drift,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(12), 2262–2275 (2010).
- [15] Wood, S. N., “Thin-plate regression splines,” *Journal of the Royal Statistical Society (B)* **65**(1), 95–114 (2003).
- [16] Halko, N., Martinsson, P. G., and Tropp, J. A., “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM Rev.* **53**, 217–288 (May 2011).
- [17] Felzenszwalb, P. F. and Huttenlocher, D. P., “Efficient belief propagation for early vision,” *International Journal of Computer Vision* (Oct. 2006).
- [18] Zandifar, A., Lim, S.-N., Duraiswami, R., Gumerov, N. A., and Davis, L. S., “Multi-level fast multipole method for thin plate spline evaluation,” in [*ICIP*], 1683–1686 (2004).