



Registration of Multi-modal Volumetric Images by Establishing Cell Correspondence

Manan Lalit^{1,2} , Mette Handberg-Thorsager^{1,2} , Yu-Wen Hsieh^{1,2} ,
Florian Jug^{1,2,3} , and Pavel Tomancak^{1,2} 

¹ Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

{jug,tomancak}@mpi-cbg.de

² Center for Systems Biology Dresden (CSBD), Dresden, Germany

³ Fondazione Human Technopole, Milan, Italy

Abstract. Early development of an animal from an egg involves a rapid increase in cell number and several cell fate specification events accompanied by dynamic morphogenetic changes. In order to correlate the morphological changes with the genetic events, one typically needs to monitor the living system with several imaging modalities offering different spatial and temporal resolution. Live imaging allows monitoring the embryo at a high temporal resolution and observing the morphological changes. On the other hand, confocal images of specimens fixed and stained for the expression of certain genes enable observing the transcription states of an embryo at specific time points during development with high spatial resolution. The two imaging modalities cannot, by definition, be applied to the same specimen and thus, separately obtained images of different specimens need to be registered. Biologically, the most meaningful way to register the images is by identifying cellular correspondences between these two imaging modalities. In this way, one can bring the two sources of information into a single domain and combine dynamic information on morphogenesis with static gene expression data. Here we propose a new computational pipeline for identifying cell-to-cell correspondences between images from multiple modalities and for using these correspondences to register 3D images within and across imaging modalities. We demonstrate this pipeline by combining four-dimensional recording of embryogenesis of Spiralian annelid ragworm *Platynereis dumerilii* with three-dimensional scans of fixed *Platynereis dumerilii* embryos stained for the expression of a variety of important developmental genes. We compare our approach with methods for aligning point clouds and show that we match the accuracy of these state-of-the-art registration pipelines on synthetic data. We show that our approach outperforms these methods on real biological imaging datasets. Importantly, our approach uniquely provides, in addition to the registration, also the non-redundant matching of corresponding, biologically meaningful entities within the registered specimen which is the prerequisite for generating biological insights from the combined datasets. The complete pipeline is available for public use through a Fiji plugin.

Keywords: Image registration · *Platynereis dumerilii* · Shape context

© Springer Nature Switzerland AG 2020

A. Bartoli and A. Fusiello (Eds.): ECCV 2020 Workshops, LNCS 12535, pp. 458–473, 2020.

https://doi.org/10.1007/978-3-030-66415-2_30

1 Introduction

Development of an animal embryo is a highly dynamic process spanning several temporal and spatial scales, and involves a series of dynamic morphogenetic events that are driven by gene regulatory networks encoded by the genome. One of the major challenges in developmental biology is to correlate the morphological changes with the underlying gene activities [18]. Recent advances in fluorescence microscopy, such as light-sheet microscopy [12, 23], allows investigating the spatio-temporal dynamics of cells in entire developing organisms and in a time-resolved manner. The three-dimensional time-lapse data produced by light-sheet microscopes contain information about positions, trajectories, and divisions of most cells in the embryo during development. However, such data sets typically lack information about gene activities in the living system.

The molecular information is provided by complementary approaches, such as confocal imaging of fixed specimens, stained for expression of a certain gene (following the molecular protocols of whole-mount *in-situ* hybridization (ISH)). The three-dimensional images of the fixed and stained embryos contain information about the spatial position of all cells or nuclei and in addition some cells are specifically labelled to indicate the expression of a gene of interest. Images of many such stained specimen showing expression of different genes at a particular stage of development can be readily collected. In order to systematically connect the molecular state of a cell to its fate during embryo morphogenesis, one needs to detect the cells in both live and fixed imaging modalities and identify cell-to-cell correspondences. This can be achieved, in principle, by aligning the images. However, the process of chemical fixation during ISH leads to a global and non-linear deformation of the specimen. Additionally, the round embryos are scanned in random orientations, and each specimen is a distinct individual showing stochastic differences in numbers and positions of the cells. This makes the problem of image registration in this context non-trivial.

We reasoned that since the primary objective is to transfer information between the imaging modalities and since cells (or nuclei) are the units of biological interest, it is more important to establish precise correspondences between equivalent cells across specimens and modalities, and that once this is achieved the registration will be obtained implicitly (Fig. 1A). We aimed to solve two matching and registration problems. Firstly, intramodal registration, where different fixed embryos stained for different gene expression patterns are registered to one reference specimen (Fig. 1B). When successful, the intramodal registration will transfer information about expression of multiple genes derived from distinct staining and imaging experiments to a single reference atlas. Secondly, intermodal registration where individual fixed and stained specimens are registered to an appropriate matching time-point of a time-lapse series of the same animal species imaged live (Fig. 1C). When successful, the intermodal registration will transfer gene expression information from fixed data to the live imaged specimen where it can be propagated along the developmental trajectories of the cells. In both cases, the common denominator are the labelled nuclei and the task is to establish the correspondences between them as precisely as possible.

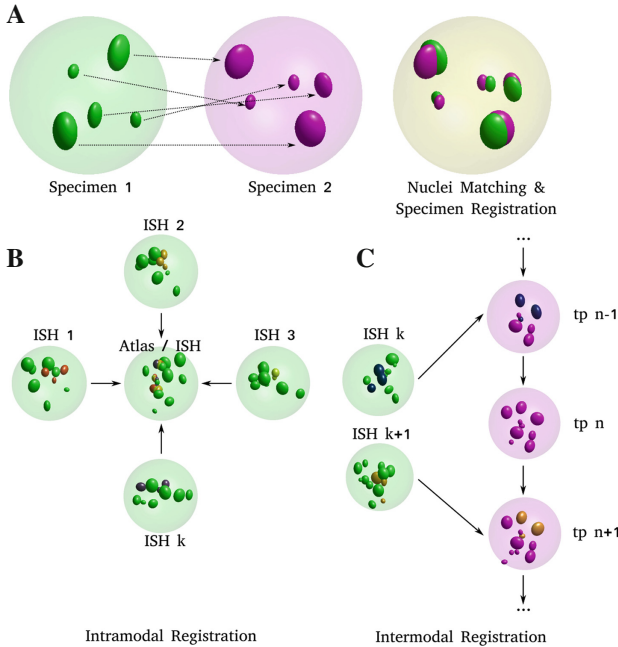


Fig. 1. Establishing cell correspondences enables registration. (A) 2-D schematic illustrating the idea: two distinct specimens (left: source and middle: target) are compared in order to estimate pair-wise cell nuclei correspondences and an optimal transform that registers the source onto the target (right) (B, C) 2-D schematics illustrating the two use cases: (B) images of distinct, independent *in-situ* specimens, acquired through confocal microscopy are registered to each other, which enables formation of an average, virtual atlas. (C) images of *in-situ* specimens, acquired through confocal microscopy are registered to the appropriate frame (tp: time point) in a time-lapse movie acquired through SPIM imaging. Nuclei indicated in darker shades are the ones expressing the gene being investigated. In both cases, the information about gene expression is transferred from the source nucleus to the corresponding target nucleus.

To address these challenges, we developed a new computational pipeline to identify cell-to-cell correspondences between images from the same and multiple imaging modalities and use these correspondences to register the images. We demonstrate the results of the pipeline on fixed ISH images of the embryos of the marine annelid worm *Platynereis dumerillii* at 16 hpf and the corresponding long term time-lapse acquired with light-sheet microscopy. This worm is particularly suitable for demonstrating our approach because its embryonic development is highly stereotypic, meaning that the number, arrangement and dynamic behaviour of cells is highly similar across individuals.

We compare our algorithm with methods for matching point clouds from computer vision such as Coherent Point Drift [19] and a variant of ICP (which we refer to as PCA-ICP) and show that our method outperforms the accuracy

of these state-of-the-art global registration pipelines on real biological data. We also perform a series of controlled experiments on synthetic data in order to demonstrate that our method is robust to initial conditions, and noisy nuclei detections. Importantly, the pipeline is made available to the biology research community through an easy-to-use plugin distributed on the Fiji platform [21], accessible through the project page <https://juglab.github.io/PlatyMatch>.

2 Related Work

2.1 Registration Approaches Applied to Images of *Platynereis Dumerillii* embryonic and larval development

Platynereis dumerillii has been a playground for image registration approaches in the recent years, due to the efforts to infer gene regulatory networks underlying neuronal development by registering ISH expression patterns. Most of this work has emphasized non-linear registration of an *in-situ* specimen to a virtual atlas. For instance, a new computational protocol was developed to obtain a virtual, high resolution gene expression atlas for the brain sub-regions in embryos at 48 hpf and onwards [22]. The reference signal used in this protocol was the larval axonal scaffold and ciliary band cells stained with an acetylated-tubulin antibody. This signal has a very distinctive 3D shape within the larva and so this approach relied on intensity based registration where linear transformations were initially applied on the source image to obtain a coarse, global registration. This was then followed by applying a non-linear, deformable transformation which employed mutual-information as the image similarity metric [26].

Another approach, more related to the path we took, leveraged the DAPI image channel (which localises the cell nuclei) to obtain registration of high-quality whole-body scans to a virtual atlas for embryos at stages 48 and 72 hpf [2] and for a larva at 144 hpf [25]. Also these approaches relied on voxel intensities of the DAPI channel rather than on the matching of segmented nuclei as in our approach. Most similar to our work is the approach of [27] where the early lineages of developing embryos were linked to gene expression ISH data by identifying corresponding nuclei between embryos imaged in two modalities based on their shape, staining intensity, and relative position.

The embryo specimens targeted in our study are spherical and highly symmetrical, lack distinctive features such as a prominent ciliary band and the nuclei are densely packed. Therefore, intensity-based registration approaches using DAPI or neuronal marker channel either fail or perform poorly on such data. Contrary to these approaches and driven by the objective to, first and foremost, transfer gene expression information with cellular resolution between modalities, we adopt a matching-by-detection workflow, where we first detect nuclei in the source and target DAPI image channels and use the detections to estimate an initial transform. We then refine this transform and estimate optimal pair-wise nuclear correspondences. Therefore, after the nuclei detection step, the problem is cast into the realm of point cloud geometric registration methods that has received substantial attention in both biological and computer vision

research communities. We discuss the existing approaches in the following two sections.

2.2 Matching of Cells or Nuclei in Biological Specimens

The work on matching nuclei between biological specimens has focused mainly on *Caenorhabditis elegans* (*C. elegans*) model system that exhibits perfectly stereotypic mode of development, and in fact, every single cell in the animal has its own name. Using this information, a digital atlas was constructed, which labels each nucleus segmentation in a three-dimensional image with an appropriate name. This was initially achieved using a relatively simple RANSAC based matching scheme [16] and was later extended by an active graph matching approach to jointly segment and annotate nuclei of the larva [14]. The *C. elegans* pipelines work well partly due to the highly distinctive overall shape of the larvae and non-homogenous distribution of the nuclei. Another example of matching nuclei between biological specimens uses identification of symmetry plane, to pair cells between multiple, independent time-lapse movies showing ascidian development [17]. These publications however emphasized nuclei detection and matching between images arising from the same modality. We are not aware of any automated strategy that identified nuclear correspondences between images from different modalities, as we attempt to do (see Fig. 1C).

2.3 Approaches to Point Cloud Matching in Computer Vision

In computer vision, a typical workflow for matching point clouds estimates a rigid or affine transform in order to perform an initial global alignment, which is followed by a local refinement of the initial transform through the Iterative Closest Point (ICP) algorithm. Many global alignment methods identify point-to-point matches based on geometric descriptors [9]. Once candidate correspondences are collected, alignment is estimated from a sparse subset of correspondences and then validated on the entire cloud. This iterative process typically employs variants of RANSAC [7].

One example of geometric descriptors is Shape Context, which was introduced by [3] for measuring similarity between two dimensional point clouds and was employed for registering surfaces in biomedical applications [1, 24]. This work was further extended for use with three dimensional point clouds by [8] and employed for the recognition of three dimensional objects.

A prominent example of geometric descriptor matching inspired by the computer vision work and applied biological image analysis is the bead-based registration of multiview light sheet (Selective Plane Illumination Microscopy (SPIM)) data [20]. Here, fluorescent beads embedded around a specimen are used as fiducial markers to achieve registration of 3D scans of the same specimen from multiple imaging angles (referred to as views). This is achieved by building rotation, translation and scale-invariant bead descriptors in local bead neighbourhoods, which enables identification of corresponding beads in multiple views and thus allows image registration and subsequent fusion of the views. The

approach was extended to multiview registration using nuclei segmented within the specimen instead of beads [13], however the approach is not robust enough to enable registration across different specimen and/or imaging modalities.

A second body of approaches estimate the optimal transform between the source and target point clouds in a single step. One such example is Coherent Point Drift (CPD) algorithm [19] where the alignment of two point clouds is considered as a probability density estimation problem : gaussian mixture model centroids (representing the first point cloud) are fitted to the data (the second point cloud) by maximizing the likelihood. CPD has also been used to perform non-rigid registration of features extracted from biomedical images [6, 11]. In this paper, we use CPD as one of the baselines to benchmark the performance of our approach.

It is important to note, that in computer vision, matching of interest points represented by geometric descriptors is not the goal but rather the means to register underlying objects or shapes in the images and volumes. Therefore, using a subset of descriptors to achieve the registration is perfectly acceptable and in fact many of the schemes rely on pruning correspondence candidates in the descriptor space to a highly reliable subset. By contrast, in biology, the nuclei that form the basis of the descriptors are at the same time the entities of interest and the goal is to match most, if not all of them, accurately.

3 Our Method

The core of our method is to match the nuclei in the various imaged specimens by means of building the shape context descriptors in a coordinate frame of reference that is unique to each nucleus. This makes the problem of matching rotationally invariant (Fig. 2 (II)). The descriptors are then matched in the descriptor space by finding the corresponding closest descriptor in the two specimens and these initial correspondences are pruned by RANSAC to achieve an initial guess of the registration. This alignment is next refined by ICP (Fig. 2 (III)). The performance of this part of the pipeline is compared to two baselines, PCA-ICP and CPD (run in affine mode), which are also able to estimate an optimal alignment. At this point, we diverge from the classical approach and evaluate the correspondences through a maximum bipartite matching to achieve the goal of matching every single nucleus from one specimen to a corresponding nucleus in the other (Fig. 2 (IV)). The pipeline relies on an efficient nucleus detection method. We present one possible approach based on scale-space theory (Fig. 2 (I)) but in principle any detection approach can be used to identify feature points to which the Shape Context descriptors would be attached. Also optionally, after the maximum bipartite matching, the estimated correspondence can be used to non-linearly deform the actual images to achieve a visually more convincing overlap of corresponding nuclei (Fig. 2 (V)). The individual steps of the pipeline are described in detail in the following subsections.

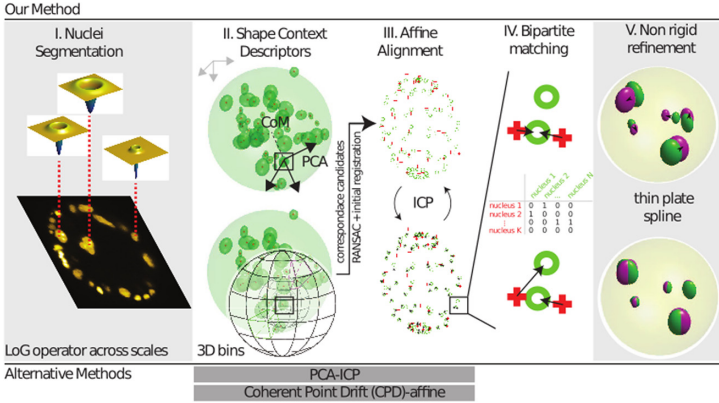


Fig. 2. Overview of the elements of the proposed registration pipeline. Figure illustrating the key elements of our pipeline: (I) A two dimensional slice of a volumetric image of the DAPI channel in a fixed *Platynereis* specimen. The operators which provide the strongest local response are shown for three exemplary cell nuclei. (II) In order to ensure that the shape context geometric descriptor is rotationally covariant, we modify the original coordinate system (shown in gray, top left) to obtain a unique coordinate system (shown in black) for each nucleus detection. The Z-axis is defined by the vector joining the center of mass of the point cloud to the point of interest, the X-axis is defined along the projection of the first principal component of the complete point cloud evaluated orthogonal to the Z-axis. The Y-axis is evaluated as a cross product of the first two vectors. Next, the neighbourhood around each nucleus detection is binned in order to compute the shape context signature for each detection. The resulting shape context descriptors from the two clouds are compared to establish correspondence candidates. These are pruned by RANSAC filtering and subsequently used to estimate a global affine transform which coarsely registers the source point cloud to the target point cloud. (III) Next, Iterative Closest Point (ICP) algorithm is used to obtain a tighter fit between the two clouds of nuclei detections. The procedure involves the iterative identification of the nearest neighbours (indicated by black arrows), followed by the estimation of the transform parameters. (IV) At this stage, a Maximum Bipartite Matching is performed between the transformed source cloud of cell nuclei detections and the static target cloud of cell nuclei detections, by employing the Hungarian Algorithm for optimization. (V) Since the two specimens are distinct individuals, non-linear differences would persist despite the preceding linear registration. We improve the quality of the registration at this stage by employing a non-linear transform (such as thin plate spline and free-form deformation) that uses the correspondences evaluated from the previous step as ground truth control points to estimate the parameters of the transform.

3.1 Detecting Nuclei

Following the scale-space theory [15], we assume that the fluorescent cell nuclei visible in the DAPI image channel inherently possess a range of scales or sizes, and that each distinct cell nucleus achieves an extremal response at a scale σ proportional to the size of that cell nucleus (Fig. 2 (I)). We compute the trace

of the scale-normalized Hessian matrix H of the gaussian-smoothed image $L(x, y, z, \sigma)$ which is equivalent to the convolution (\otimes) response resulting from the scale-normalized Laplacian of Gaussian kernel and the image $I(x, y, z)$ i.e.

$$\begin{aligned} \text{trace}(H_{\text{norm}}L(x, y, z, \sigma)) &= \sigma^2(L_{xx} + L_{yy} + L_{zz}), \text{ where} \\ L_{kk} &= \frac{\partial^2 G_\sigma}{\partial k^2} \otimes I(x, y, z), k \in \{x, y, z\} \text{ and} \\ G_\sigma(x, y, z) &= \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} e^{-\frac{x^2+y^2+z^2}{2\sigma^2}}. \end{aligned} \quad (1)$$

The cell nuclei centroid locations (and additional scale information) are then estimated as the local minima of the 4D (x, y, z, σ) space. At this stage, some of the detections might overlap especially in dense regions. To address this, first we employ the assumption that the estimated spherical radius \hat{r} of a cell nucleus is related to its estimated scale $\hat{\sigma}$ through the following relation $\hat{r} = \sqrt{3}\hat{\sigma}$. Next we state a relation drawn from algebra that if d is the distance between two spheres with radii r_1 and r_2 (and corresponding volumes V_1 and V_2 , respectively), and provided that $d < r_1 + r_2$, the volume of intersection V_i of these two spheres is calculated as in [4], by:

$$V_i = \frac{\pi}{12d} (r_1 + r_2 - d)^2 (d^2 + 2d(r_1 + r_2) - 3(r_1 - r_2))^2. \quad (2)$$

Spheres for which $V_i < t \times \min(V_1, V_2)$ are suppressed greedily, by employing a non-maximum suppression step. In our experiments, we use the threshold $t = 0.05$. An optional manual curation of the nuclei detections is made possible through our Fiji plugin.

3.2 Finding Corresponding Nuclei Between Two Point Clouds

Estimating a Global Affine Transform. In this section, we will provide the details of our implementation of the 3D shape context geometric descriptor, which is a signature obtained uniquely for all feature points in the source and target point clouds. This descriptor takes as input a point cloud P (which represents the nuclei detections described in the previous section) and a basis point p , and captures the regional shape of the scene at p using the distribution of points in a support region surrounding p . The support region is discretized into bins, and a histogram is formed by counting the number of point neighbours falling within each bin. As in [3], in order to be more sensitive to nearby points, we use a log-polar coordinate system (Fig. 2 (II), bottom). In our experiments, we build a 3D histogram with 5 equally spaced log-radius bins and 6 and 12 equally spaced elevation (θ) and azimuth (ϕ) bins respectively.

For each basis point p , we define a unique right-handed coordinate system: the Z-axis is defined by the vector joining the center of mass of the point cloud to the point of interest, the X-axis is defined along the projection of the first principal component of all point locations in P , evaluated orthogonal to the Z-axis. The Y-axis is evaluated as a cross product of the first two vectors (Fig. 2

(II), top). Since the sign of the first principal component vector is a ‘numerical accident’ and thus not repeatable, we use both possibilities and evaluate two shape context descriptors for each feature point in the source cloud. Building such a unique coordinate system for each feature point ensures that the shape context descriptor is rotationally invariant. Additionally since the chemical fixation introduces shrinking of the embryo volume (the intermodal registration use case, see Fig. 1C) and since the embryo volume may considerably differ across a population (intramodal use case, see Fig. 1B), an additional normalization of the shape context descriptor is performed to achieve scale invariance. This is done by normalizing all the radial distances between p and its neighbours by the mean distance between all point pairs arising in the point cloud. Similar to [3], we use the χ^2 metric to identify the cost of matching two points p_i and q_j arising from two different point clouds i.e.

$$C_{ij} := C(p_i, q_j) = \frac{1}{2} \sum_{k=1}^K \frac{(h_i(k) - h_j(k))^2}{h_i(k) + h_j(k)}, \quad (3)$$

where $h_i(k)$ and $h_j(k)$ denote the K -bin normalized histogram at p_i and q_j respectively. By comparing shape contexts resulting from the two clouds of cell nuclei detections, we obtain an initial temporary set of correspondences. These are filtered to obtain a set of inlier point correspondences using RANSAC [7]. In our experiments, we specified an affine transform model, which requires a sampling of 4 pairs of corresponding points. We executed RANSAC for 20000 trials, used the Moore-Penrose Inverse operation to estimate the affine transform between the two sets of corresponding locations, and allowed an inlier cutoff of 15 pixels L_2 distance between the transformed and the target nucleus locations.

Obtaining a Tighter Fit with ICP. The previous step provides us a good initial alignment. Next, we employ ICP which alternates between establishing correspondences via closest-point lookups (see Fig. 2 (III)) and recomputing the optimal transform based on the current set of correspondences. Typically, one employs Horn’s approach [10] to estimate strictly-rigid transform parameters. We see equivalently accurate results with iteratively estimating an affine transform, which we compute by employing the Moore-Penrose Inverse operation between the current set of correspondences.

Estimating the Complete Set of Correspondences. We build a $M \times N$ -sized cost matrix C where the entry C_{ij} is the euclidean distance between the i^{th} transformed source cell nucleus detection and the j^{th} target cell nucleus detection. Next, we employ the Hungarian Algorithm to perform a maximum bipartite matching and estimate correspondences \hat{X} (see Fig. 2 (IV)):

$$\hat{X} = \arg \min_X \sum_{i=1}^M \sum_{j=1}^N C_{ij} X_{ij}, \text{ where } X_{ij} \in \{0, 1\} \text{ s.t. } \sum_{k=1}^{k=M} X_{ik} \leq 1, \sum_{k=1}^{k=N} X_{kj} \leq 1.$$

3.3 Estimating a Non-linear Transform

Since the two specimens being registered are distinct individuals, non-linear differences would persist despite the preceding, linear (affine) registration. We improve the quality of the image registration at this stage by implementing an optional non-linear transform (for example the thin-plate spline transform or the free-form deformation). The correspondences evaluated from the previous step are used as ground truth control points to estimate the parameters of the transform function.

4 Materials

To test our method, we are using two sets of real biological specimen. Firstly, representing the fixed biological specimen containing information about gene expression, we collected whole-mount specimens of *Platynereis dumerilii* stained with ISH probes for several different, developmentally regulated transcription factors at the specific developmental stage of 16 hpf. These specimens were scanned in 3D by laser scanning confocal microscopy resulting in three-dimensional images containing the DAPI (nucleus) channel used in our registration as a common reference and the gene expression channel. Secondly, representing the live imaging modality, we obtained access to a recording capturing the embryological development of the *Platynereis dumerilii* at cellular resolution *in toto* [23] using a SimView light sheet microscope. The embryos were injected with a fluorescent nuclear tracer prior to imaging and thus the time-lapse movie visualizes all the nuclei in the embryo throughout development. This movie includes the 16 hpf stage of *Platynereis* development providing an appropriate inter-modal target to register the fixed specimen to on the basis of the common nuclear signal.

5 Results

We evaluate our proposed strategy on real and simulated data and compare against two competitive baselines. The first baseline, which we refer to as *PCA-ICP* is an extension of ICP and includes a robust initialization prior to performing ICP. The center of mass of the source point cloud is translated to the location of the center of mass of the target point cloud. Next, the translated source point cloud is rotated about its new center of mass such that its three principal component vectors align with the three principal component vectors of the target point cloud. In order to ensure that the orthogonal system forming the three principal components is not mirrored along any axis, we consider all 8 possibilities for the obtained principal component vectors of the source point cloud. We initialize ICP from these 8 setups and iteratively estimated a *similar* transform (scale, rotation and translation). Finally, the configuration which provides the least L_2 euclidean distance between the two sets of correspondences obtained through nearest neighbour lookups upon the termination of ICP, is kept and the rest of the configurations are discarded. The second baseline is *Coherent Point*

Drift (CPD) [19]. In our experiments, we executed CPD in the Affine mode with normalization set to 1, maximum iterations equal to 100 and tolerance equal to $1e^{-10}$. We use two metrics in order to quantify the performance of all considered methods: (i) *Matching Accuracy* which we define as the ratio of the true positive matches and the total number of inlier matches, and (ii) *Average Registration Error* which we define as the average L_2 euclidean distance between a set of ground truth landmarks arising from the two point clouds, evaluated after the completion of the registration pipeline. A higher Matching Accuracy and a lower Average Registration Error are desirable readouts to demonstrate better performance.

5.1 Experiments on Real Data

For the intramodal registration use case (see Figs. 3A & 3C), nuclei detections arising from 11 images of *in-situ* specimens were registered to nuclei detections arising from the image of a typical, target *in-situ* specimen. Since for real data the true correspondences are not known, we asked expert biologists to manually identify 12 corresponding landmark nuclei. This set represents ground truth landmarks against which we evaluated the results of our registration based on the average L_2 euclidean distance of proposed landmark correspondences (Source landmarks are labeled 1, ... 12 and Target Landmarks are similarly labeled 1', ... 12' in Fig. 3).

For the intermodal registration use case (see Figs. 3B & 3D), nuclei detections arising from 7 confocal images of *in-situ* specimens are registered to the corresponding frame from the time lapse movie which contains an equivalent number of nuclei. They were similarly evaluated on the average L_2 euclidean distance in the positions of landmarks identified in the movies by the expert annotators. We noticed that instead of directly registering an *in-situ* specimen to its corresponding time point in the developmental trajectory of the live embryo, our pipeline gives better results if we map the *in-situ* specimen to a reference atlas, then apply a pre-computed affine transform to the atlas to transform it to the domain of the live embryo and lastly refine this coarse registration using ICP.

The results show that after applying our proposed pipeline, the average registration error of corresponding landmarks is around 25 and 35 pixels for our intramodal and intermodal registration use cases respectively (Fig. 3E). The accuracy is significantly better compared to the baseline methods. The exemplary intramodal image shows good overlap of the nuclear intensities (Fig. 3C). The displacement of the corresponding landmarks (denoted by the yellow unprimed numbers) is better in the left part of the specimen compared to the right part. This suggests that significant non-linear deformation occurred during the staining process and our current pipeline relying on affine models is unable to undo this deformation. For the intermodal registration, the pipeline clearly compensated for the mismatch in scale between the fixed and live specimen (Fig. 3D). The remaining error is, similar to the intramodal case, likely due to non-linear distortions. In terms of matching accuracy after performing maximum bipartite matching, our method outperforms the baselines. Since the

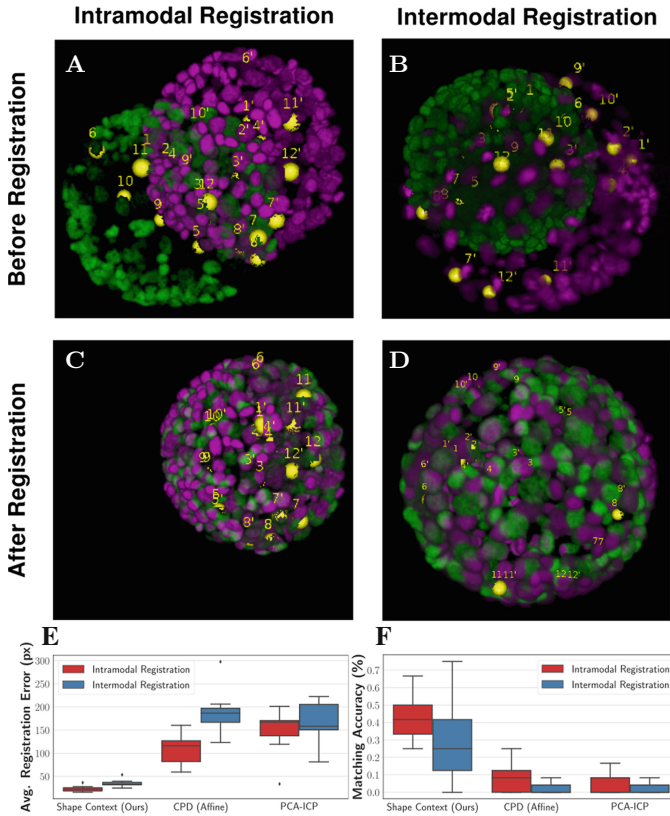


Fig. 3. Experiments on real data. (A) DAPI channels indicating cell nuclei for two distinct *in-situ* specimens before registration (source: green, target: magenta). (B) DAPI channels indicating cell nuclei for an *in-situ* specimen (source: green) before it was registered to the corresponding frame containing equivalent number of cell nuclei, in the time-lapse movie (target: magenta). Landmarks for source image are indicated as yellow spheres and labeled from 1, . . . 12. Similarly, landmarks for the target image are labeled from 1', . . . 12'. (C) Specimen shown in (A) after intramodal registration using our proposed pipeline. (D) Specimen shown in (B) after intermodal registration using our proposed pipeline. (E) Plot indicating the average Euclidean distance between landmarks after applying different registration pipelines. (F) Plot indicating the percent of correct correspondences between landmarks, evaluated through Maximum Bipartite Matching, after applying different registration pipelines. (Color figure online)

matching accuracy is estimated on only 12 corresponding landmarks, which represents only 3.6 % of the total matched nuclei, it is likely subject to sampling error. This is reflected by the broad spread of accuracy for both inter- and intramodal use cases (Fig. 3F).

Since obtaining a larger set of ground truth correspondences is not practical we turn next to evaluating the approach on synthetic data.

5.2 Experiments on Simulated Data

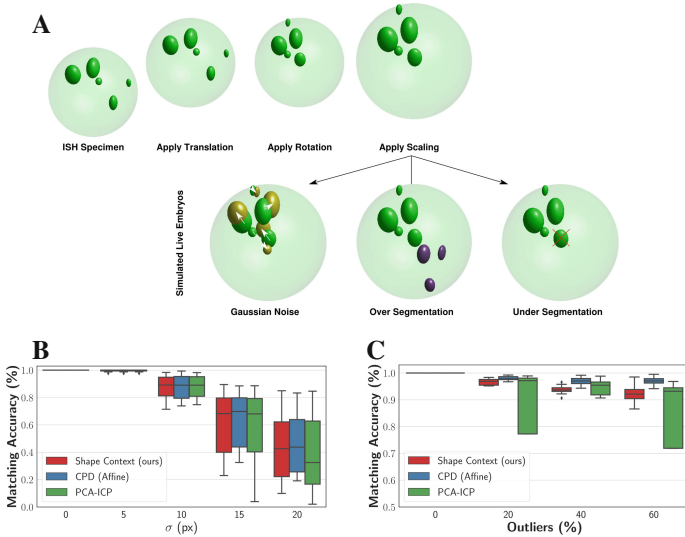


Fig. 4. Experiments on simulated data. Synthetic ‘live’ embryos are simulated by manipulating cell nuclei detections from real *in-situ* specimens globally and locally. (A) First all cell nuclei detections of an *in-situ* specimen were translationally offset, next the translated point cloud was randomly rotated by an angle $\in \{-\pi/6, \pi/6\}$ about a random axis passing through the center of mass of the translated point cloud, and finally, the translated and rotated point cloud was scaled by a random factor. (B) The above globally transformed nuclei are provided independent gaussian noise. Plot indicating the percent of correct correspondences between all pairs evaluated through Maximum Bipartite Matching, after applying different registration pipelines. (C) The above globally transformed nuclei are corrupted with excess outliers. Plot indicating the percent of correct correspondences between all pairs evaluated through Maximum Bipartite Matching, after applying different registration pipelines.

Starting from the nuclei detection on real fixed embryos, we generated simulated ground truth data by random translation, rotation and scaling operations, followed by (i) adding gaussian noise to the location of individual segments (i.e. nuclei) and (ii) randomly adding nuclei (Fig. 4A). The simulated embryos are meant to resemble the live-imaged embryos which in real scenarios are also rotated, translated and scaled compared to the fixed specimens and may have extraneous or missing nuclei due to biological variability or segmentation errors.

Robustness to Gaussian Noise. The synthetic ‘live embryos’ were generated by manipulating nuclei detections from multiple, independent *in-situ* specimens. First, the nuclei detections of each *in-situ* specimen are provided a random

translation offset, next the translated point cloud is rotated by a random angle between -30° and $+30^\circ$ about an arbitrary axis passing through the center of mass of the point cloud, and finally, the translated and rotated point cloud is scaled by a random factor (See Fig. 4A). After these global transformations, we add gaussian noise to each individual detection in order to vary their positions independently along the X, Y and Z axes. We evaluate five levels of Gaussian noise with standard deviations [0:5:20]. The results of evaluation of matching accuracy with respect to different levels of Gaussian noise show that all methods provide equivalent performance (Fig. 4B). The matching accuracy starts to break down when the magnitude of gaussian noise is greater than 10 pixels.

Robustness to Outliers. In order to test robustness against over or under-segmentation of nuclei, we add outliers to both the source fixed *in-situ* volumes and the corresponding simulated ‘live embryo’. New outlier points are generated by sampling existing points and adding a new point at a standard deviation of 20 pixels from their locations. The results show that the CPD Affine method performs the best in the presence of outliers, while our approach is more stable compared to the PCA-ICP (Fig. 4C).

6 Discussion

Our method showed promising results on real biological data in terms of average registration error and provided equivalent performance when compared to state of the art methods on simulated data. The pipeline offers several entry points for further improvement towards achieving more precise one-to-one matching of cells within and across imaging modalities for separate biological specimens. One area open for future investigations is certainly obtaining more accurate initial segmentations. Another performance boost may come from the definition of the 3D geometric descriptor. Our implementation of shape context as a 3D geometric descriptor draws from [3]. We use a log-polar coordinate system and build 3D histograms by evenly dividing the azimuth and elevation axis. This creates bins with unequal sizes, especially near the poles, and makes the matching of feature points non robust to noisy detections. This drawback could be addressed through two approaches: (i) employing the optimal transport distance [5] between two 3D histograms would provide a more natural way of comparing two histograms as opposed to the current χ^2 squared distance formulation, (ii) opting for a more uniform binning scheme (see for example, [28]) would eliminate the issue of noisy detections jumping arbitrarily between bins near the poles. Finally, the method will benefit from non-linear refinement as the specimens are often deformed in an unpredictable manner during the staining and imaging protocols.

By establishing nuclei correspondences between images of *in-situ* specimens and the time lapse movies, biologists will be able to transfer the gene expression information from the fixed specimens to the dynamic cell lineage tree generated by performing cell tracing on the time-lapse movie. This will enable biologists to study the molecular underpinning of dynamic morphogenetic processes occurring during embryo development.

References

1. Acosta, O., et al.: 3D shape context surface registration for cortical mapping (2010). In: IEEE International Symposium on Biomedical Imaging (ISBI) 2010
2. Asadulina, A., Panzera, A., Veraszto, C., Lieblig, C., Jekely, G.: Whole-body gene expression pattern registration in *Platynereis* larvae. *EvoDevo* **3**(27), (2012)
3. Belongie, S., Puzicha, J., Malik, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 509–522 (2002)
4. Bondi, A.: Van der Waals volumes and radii. *J. Phys. Chem.* **68**, 441–451 (1964)
5. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transportation distances. In: *Advances in Neural Information Processing Systems*, vol. 26, pp. 2292–2300 (2013)
6. Farnia, P., Ahmadian, A., Khoshnevisan, A., Jaberzadeh, N., Kazerooni, A.: An efficient point based registration of intra-operative ultrasound images with MR images for computation of brain shift; a phantom study. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2011)
7. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)
8. Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: Pajdla, T., Matas, J. (eds.) *ECCV 2004*. LNCS, vol. 3023, pp. 224–237. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24672-5_18
9. Guo, Y., Bennamoun, M., Sohler, F.: A comprehensive performance evaluation of 3D local feature descriptors. *Int. J. Comput. Vis.* **116**, 66–89 (2016). <https://doi.org/10.1007/s11263-015-0824-y>
10. Horn, B.: Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am.* **4**, 629–642 (1987)
11. Hu, Y., Rijkhorst, E.-J., Manber, R., Hawkes, D., Barratt, D.: Deformable Vessel-Based Registration Using Landmark-Guided Coherent Point Drift. In: Liao, H., Edwards, P.J.E., Pan, X., Fan, Y., Yang, G.-Z. (eds.) *MIAR 2010*. LNCS, vol. 6326, pp. 60–69. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15699-1_7
12. Huisken, J., Swoger, J., Bene, F.D., Wittbrodt, J., Stelzer, E.H.K.: Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* **305**, 1007–1009 (2004)
13. Hörl, D., et al.: BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples. *Nat. Methods* **16**, 870–874 (2019)
14. Kainmueller, D., Jug, F., Rother, C., Myers, G.: Active graph matching for automatic joint segmentation and annotation of *C. elegans*. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) *MICCAI 2014*. LNCS, vol. 8673, pp. 81–88. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10404-1_11
15. Lindeberg, T.: Feature detection with automatic scale selection. *Int. J. Comput. Vis.* **30**(2), 70–116 (1998). <https://doi.org/10.1023/A:1008045108935>
16. Long, F., Peng, H., Liu, X., Kim, S.K., Myers, E.: A 3D digital atlas of *C. elegans* and its application to single-cell analyses. *Nat. Methods* **6**(4), 667–672 (2009)
17. Michelin, G., Guignard, L., Fiuza, U.M., Lemaire, P., Godin, C., Malandain, G.: Cell pairings for ascidian embryo registration. In: *2015 IEEE International Symposium on Biomedical Imaging (ISBI)* (2015)

18. Munro, E., Robin, F., Lemaire, P.: Cellular morphogenesis in ascidians: how to shape a simple tadpole. *Curr. Opin. Genet. Dev.* **16**(4), 399–405 (2006)
19. Myronenko, A., Song, X.: Point set registration: coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(12), 2262–2275 (2010)
20. Preibisch, S., Saalfeld, S., Schindelin, J., Tomancak, P.: Software for bead-based registration of selective plane illumination microscopy data. *Nat. Methods* **7**, 418–419 (2010)
21. Schindelin, J., et al.: Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**(7), 676–682 (2012)
22. Tomer, R., Denes, A.S., Tessmar-Raible, K., Arendt, D.: Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell* **142**(5), 800–809 (2010)
23. Tomer, R., Khairy, K., Amat, F., Keller, P.J.: Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nat. Methods* **9**, 755–763 (2012)
24. Urschler, M., Bischof, H.: Registering 3D lung surfaces using the shape context approach. In: *Proceedings of the International Conference on Medical Image Understanding and Analysis*, pp. 512–215 (2004)
25. Vergara, H.M., et al.: Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of *Platynereis dumerilii*. *Proc. Natl. Acad. Sci.* **114**(23), 5878–5885 (2017)
26. Viola, P.A., Wells, W.M.: Alignment by maximization of mutual information. *Int. J. Comput. Vis.* **24**, 137–154 (1995). <https://doi.org/10.1023/A:1007958904918>
27. Vopalensky, P., Tosches, M., Achim, K., Thorsager, M.H., Arendt, D.: From spiral cleavage to bilateral symmetry: the developmental cell lineage of the annelid brain. *BMC Biol.* **17**, 81 (2019). <https://doi.org/10.1186/s12915-019-0705-x>
28. Zhong, Y.: Intrinsic shape signatures: a shape descriptor for 3D object recognition. In: *International Conference on Computer Vision ICCV Workshop* (2009)