Puck alignment and interpolation

Léo Guignard

October 15, 2021

When doing spatial single-cell transcriptomics, beads are recorded from pucks. Beads are placed on a 2D matrix where each bead is spaced by a given distance x_{res} (resp. y_{res}) along the x (resp. y) dimension (in our case $x_{res} = y_{res} = XX\mu m$). These distances define the xy resolution (or lateral resolution) of the slice or puck. Then, consecutive pucks are spaced by a given distance z_{res} defining the z resolution (or axial resolution) of the dataset (in our case $z_{res} = 30\mu m$).

Because the pucks are physically moved between their slicing and their acquisition, they are not acquired within the same frame (meaning that they are not aligned). In order to reconstruct a 3D representation of the single cell transcriptomics of the acquired embryo and to do full 3D spatial analysis, it is necessary to align consecutive pucks to recover the spatial integrity of the specimen. Moreover, because the axial resolution is significantly greater than the lateral one, in some cases, it is necessary to interpolate the data between the pucks.

In the following section we will describe how this alignment was performed together with how the beads were interpolated between pucks.

1 Notation

We define our dataset as a set of pucks $\mathcal{P} = \{P_i\}$. The function c_P maps a puck to its height coordinate z_i: $c_P : P_i \in \mathcal{P} \to z_i \in \mathbb{R}$. Each puck P_i is itself a set of beads, $P_i = \{b_{ij}\}$ and similarly to the pucks, the function c_b maps a bead b_{ij} to its xy coordinate within the puck: $c_b : b_{ij} \in P_i \to (x, y) \in \mathbb{R}^2$. From c_P and c_b we define the function $c : b_{ij} \in P_i \to (x, y, z) \in \mathbb{R}$ which maps a bead to its 3D spatial coordinate. Note that c_P can define a total order on the pucks. Let then $\mathcal{P}$ be ordered such that $\forall i, j,\ i < j \iff c_P(P_i) < c_P(P_j)$.

Moreover, using the previously described analysis, we can associate each bead to the tissue it most likely belongs to from the previous analyses, the function $T : b_{ij} \in P_i \to t \in \mathcal{T}$ where $t \in \mathcal{T}$ is a unique identifier for a tissue and $\mathcal{T}$ is the set of tissues. Similarly, to each bead is associated a value for each given gene analysed in the dataset. This value is correlated to the level of expression of said gene for said bead. Given a gene g, we define the function $E_g : b_{ij} \in P_i \to e \in \mathbb{R}$ which maps the value e of expression of the gene g to the bead b_{ij}.

2 Pre-processing

2.1 Removing the remaining outliers

Before aligning the pucks we removed the beads that were likely to be noise. We did so by detecting the beads that were further away from their neighbors from the same tissue than the majority of the beads from the same tissue type. To do so we first computed the distance between any given bead and its closest bead from the same tissue type. We then analysed the distribution of these distances by fitting a gaussian mixture model fixing the number of components to 3. The 1^{st} and 2^{nd} components are the one with the two means and are distributions of distances between real beads. The 3^{rd} component with the higher mean is the distribution of distances of noisy beads. We then discarded all the beads that had a distance to their closest neighbor from the same tissue which had a probability to belong to the first or second component were lower than th_{gmm} (in our case $XX\%$).

3 Aligning the pucks

As previously mentioned, due to the nature of the acquisition process, consecutive pucks do not live within the same frame. They are therefore not spatially comparable.

To align the pucks and register them onto the same frame, we first chose our first puck in $\mathcal{P}$ (P_0) as the reference puck. We then registered each following slide to its preceding one (ie P_1 is registered onto P_0, P_{i+1} is registered onto P_i etc.). Ultimately we can compose all the transformations together to register any puck onto the first puck. To compute the transformation necessary to register two consecutive pucks, we first performed a coarse grain alignment using the center of mass of a subset of the different tissue types. We then refined the alignment by pairing beads across the pucks and by aligning them.

3.1 Coarse grain alignment

We first chose a subset of tissues $\mathcal{L} \subset \mathcal{T}$ that are spatially localised (for example heart tube precursor beads or somite precursor beads). We then discarded (only for the coarse grain alignment) tissues that were spread in space (for example blood beads). An exhaustive list of the discarded tissue types can be found below.

Then we computed the alignment transformation ($r^*_{i\leftarrow j}$) as the rigid transformation (translation plus rotation) that minimizes the sum of the squared distances between corresponding tissue types center of mass:

$$r^*_{i\leftarrow j} \quad = \quad \underset{r\in\mathcal{R}}{\arg\max}\left\{\sum_{t\in\mathcal{L}} \|COM_i(t) - r[COM_j(t)]\|_2\right\} \tag{1}$$

where $COM_i(t)$ is the position of the center of mass of the tissue t in the puck i, $\|\cdot\|_2$ is the L2 norm and $\mathcal{R}$ is the set of all rigid transformations. $r[COM_j(t)]$ is therefore the position of the center of mass of the tissue t in the puck j after application of the rigid transformation r.

We then applied the composition of the necessary transformations to register all pucks onto the first puck. For example, to register the puck j onto the puck 0, we applied the transformation $r^*_{0\leftarrow j} = r^*_{0\leftarrow 1} \circ r^*_{1\leftarrow 2} \circ \cdots \circ r^*_{j-1\leftarrow j}$.

3.2 Alignment refinement

To refine the alignment, we then paired beads from consecutive pucks. Only beads from the same tissue types could be paired. The pairing was the one that minimizes the sum of the distances between paired beads (using the solution of the linear sum assignment optimization). The distances were computed after applying the coarse grain transformation. From this pairing, as previously, we computed the rigid transformation $R^*_{i\leftarrow j}$ that minimizes the sum of the squares of the distances between the paired beads (see eq.(1)).

4 Interpolation

As shown earlier, the distance between two consecutive beads within the same puck is significantly smaller than the one between two pucks (one order of magnitude in our case, $30\mu m$ versus $6\mu m$). This property makes it that some analysis of the 3D volume of the embryo is difficult or even not possible. For example, if one wants to study a plane that is tilted by a small angle to the original puck planes, the beads become too scarce. To overcome this issue, we designed an algorithm to interpolate beads in between pucks. The goal is to create beads where there would likely be and to assign them gene expression values that are likely to be correct.

To interpolate beads between pucks, we first paired beads from the same tissues (similarly to the pairing that was done in 3.2). We then assumed a continuity of expression and position to interpolate the beads. The continuity of expression and position hypotheses is the basis of this interpolation. This means that the interpolation is likely to be incorrect when looking at tissues within which the cells are randomly positioned such as blood cells for example.

4.1 Bead pairings and paths

Beads from consecutive pucks were paired based on their proximity. As explained earlier, we used a linear assignment algorithm to pair beads. The linear sum alignment formulation requires a positive cost to be computed between any beads that can be paired together. Here, we used as our cost function the Euclidean distance. The bead pairing between consecutive pucks creates paths of beads. A path of beads is a set of beads that are paired sequentially by the puck to puck pairing. For example $b_{i,j}$ is paired to $b_{i+1,k}$ which is paired to $b_{i+2,l}$. These beads are paired creating the path $Pa = (b_{i,j}, b_{i+1,k}, b_{i+2,l})$. We can create the set of all possible paths $\{Pa_{ij}\}$ where i, j refers to the first bead $(b_{i,j})$ in the path.

4.2 Interpolation

To interpolate beads between slices we used the paths that we created. A path has z coordinates along which the interpolation is done. Then, given a variable $x_{i,j}$ (for example the position in x or the expression of a gene g) along the path, we build the spline interpolation of x along z. We repeated the operation for the x and y positions and for all genes of interest.