Supporting Information: Automated stitching of microtubule centerlines across serial electron tomograms

See main manuscript for references and figures.

Algorithms for fine alignment

This section describes in detail the algorithms for fine alignment. The method uses the expectation maximization algorithm. To apply expectation maximization, we derive 1) the update equations for the posterior $P(m|x)$ in the E-step and 2) the update equations for the distribution and transformation parameters in the M-step, which is done by minimizing the expectation of the complete negative log-likelihood function $L$.

We give the detailed derivation of the update equations for each of the three methods and summarize the algorithms in Figures S2, S3 and S4 in the style of Myronenko and Song [28].

Notation In addition to the notation defined in the main text we will use the following notation: For two sets of endpoints $x_1,...,x_N$ and $y_1,...,y_M$ the matrix $X$ contains all endpoint positions in $X$ as rows $X = (x_1,...,x_N)^\top$ and likewise $Y = (y_1,...,y_M)^\top$.

$P$ denotes an $M \times N$ matrix that contains the values of the posterior $P(m|x)$ for all possible pairs $x_n,y_m$. Furthermore, we write $N_P = \sum_{m,n=1}^{M,N} P(m|x)$.

The operator $\text{diag}(v)$ creates a diagonal matrix from a vector $v \in \mathbb{R}^D$ by $\text{diag}(v) = \begin{pmatrix} v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_D \end{pmatrix}$.

We denote a vector containing only ones as $1$.

Further prerequisites To account for outliers, Myronenko and Song introduce an additional term $P(x,M+1) = 1/N$ to the joint distribution. This outlier term can be weighted with $0 \leq w \leq 1$. The marginal probability of $x$ then takes the form:

$$P(x) = \sum_{m=1}^{M+1} p(m)P(x|m) = \frac{1}{M} \sum_{m=1}^{M} P(x|m) + \frac{w}{N}. \quad (12)$$

We use the same outlier term in all three settings.

For convenience, we will often write sums as matrices using the following equalities:

$$\sum_{m,n=1}^{M,N} P(m|x_n)x_n = X^\top P^\top 1 \quad (13)$$

$$\sum_{m,n=1}^{M,N} P(m|x_n)y_m = Y^\top P 1 \quad (14)$$

$$\sum_{m,n=1}^{M,N} P(m|x_n)x_n^\top x_n = \text{tr}(X^\top \text{diag}(P^\top 1)X) \quad (15)$$

$$\sum_{m,n=1}^{M,N} P(m|x_n)x_n x_n^\top = X^\top \text{diag}(P^\top 1)X \quad (16)$$
\[ \sum_{m,n=1}^{M,N} P(m|x_n)y_m^\top y_m = \text{tr}(Y^\top \text{diag}(P_1)Y) \quad (17) \]

\[ \sum_{m,n=1}^{M,N} P(m|x_n)y_m^\top Y = \text{tr}(Y^\top \text{diag}(P_1)Y) \quad (18) \]

\[ \sum_{m,n=1}^{M,N} P(m|x_n)x_n^\top R y_m = \text{tr}(X^\top P^\top Y R^\top) \quad (19) \]

\[ \sum_{m,n=1}^{M,N} P(m|x_n)x_n^\top R Y = X^\top P^\top Y R^\top \quad (20) \]

**Algorithm for linear alignment from orientation**

**E-step** The posterior \( P(\hat{x}|\bar{x}) \) is computed by

\[ P(\hat{x}|\bar{x}) = \frac{P(\hat{x}, \bar{x})}{P(\bar{x})} \quad (21) \]

To compute \( P(\hat{x}, \bar{x}) \), we add an outlier term \( P(\bar{x}|M + 1) = 1/N \) and weight the priors with \( 0 \leq w \leq 1 \) so that \( P(m) = (1-w)/M \) and \( P(M+1) = w/N \). \( P(\hat{x}|\bar{x}) \) can be computed from the priors and \( P(\bar{x}|\hat{x}) \) which is given by Equation 6 for \( m \leq M \). The joint distribution becomes:

\[ P(\hat{x}, \bar{x}) = P(\hat{x})P(\bar{x}|\hat{x}) = \frac{(1-w)}{M} \frac{\kappa}{2\pi(e^\kappa - e^{-\kappa})} \exp(\kappa(R\bar{y}_m)^\top \bar{x}), \text{for all } m \leq M \quad (22) \]

\[ P(\bar{x}, M+1) = \sqrt{w}/N, \text{for } m = M+1 \quad (23) \]

\( P(\bar{x}) \) can be computed from \( P(\hat{x}, \bar{x}) \) by marginalizing out \( \hat{x} \):

\[ P(\bar{x}) = P(\bar{x}, M+1) + \sum_{m=1}^{M} P(\hat{x}, \bar{x}) = \sqrt{w}/N + (1-w) \sum_{k=1}^{M} \frac{\kappa}{2M\pi(e^\kappa - e^{-\kappa})} \exp(\kappa(R\bar{y}_k)^\top \bar{x}) \quad (24) \]

The posterior becomes

\[ P(\hat{x}|\bar{x}) = \frac{P(\hat{x}, \bar{x})}{P(\bar{x})} = \frac{\exp(\kappa\bar{x}_n^\top R\bar{y}_m)}{\sum_{k=1}^{M} \exp(\kappa\bar{x}_n^\top R\bar{y}_k) + w/M \frac{\sqrt{w}(e^{\kappa} - e^{-\kappa})}{e^{\kappa} - e^{-\kappa}} \quad (25) \]

**M-step** In the M-step, we must write out the expectation of the negative log-likelihood function \( Q \) (Equation 3) and minimize it with respect to the transformation parameters \( R \) and the distribution parameter \( \kappa \). \( Q \) becomes:

\[ Q(R, \kappa) = -\kappa \sum_{m,n=1}^{M,N} p^{\text{mod}}(m|x_n)x_n^\top R\bar{y}_m - N_P \log(\frac{\kappa}{e^\kappa - e^{-\kappa}}) + \text{const.} = -\kappa \text{tr}((\bar{X}^\top P^\top \bar{Y})^\top R) - N_P \log(\frac{\kappa}{e^\kappa - e^{-\kappa}}) + \text{const.} \quad (26) \]

We split the term \( \text{tr}((\bar{X}^\top P^\top \bar{Y})^\top R) \) into terms that are dependent on and independent of \( \bar{R} \):

\[ \text{tr}((\bar{X}^\top P^\top \bar{Y})^\top R) = \text{tr}((\bar{X}_{2d}^\top P^\top \bar{Y}_{2d})^\top \bar{R}) + \bar{X}_{2d}^\top P^\top \bar{Y}_{2d} \quad (27) \]
where \( \bar{X}_{2d} \) is a matrix containing the first two columns of \( \bar{X} \) and \( \bar{X}_z \) contains the last column of \( \bar{X} \).

The optimal rotation can be obtained with a Singular Value Decomposition from the term \( \bar{X}_{2d}^T P \bar{Y}_{2d} \) (equivalently to Myronenko and Song [28], see Umeyama [42] and Myronenko and Song [43] for details on the method). \( \kappa \) cannot be obtained in closed form, but it can be computed with Newton’s method.

We iterate

\[
\kappa = \kappa^{old} - \frac{\partial Q}{\partial \kappa} \frac{\partial^2 Q}{\partial \kappa^2}
\]

until convergence, where

\[
\frac{\partial Q}{\partial \kappa} = -\text{tr}((\bar{X}_{2d}^T P \bar{Y}_{2d})^T \bar{R}) - \text{tr}(\bar{X}_z^T P \bar{Y}_z) - NP(\frac{1}{\kappa} - \coth \kappa),
\]

and

\[
\frac{\partial^2 Q}{\partial \kappa^2} = -NP(-\frac{1}{\kappa^2} + \frac{1}{\sinh^2(\kappa)}).
\]

Because we only compute a rotation, the endpoints are not aligned. To align them, we first note that even though we did not explicitly take the positions into account, the likelihood that two orientations match also applies to the corresponding endpoint positions. To compute an alignment of the positions, we compute the squared distances of the positions, which we weight by the posterior computed from the obtained optimal parameters \( \Theta^* \), \( P(\bar{m}|\bar{x}, \Theta^*) \):

\[
S = \sum_{m,n=1}^{M,N} p(\bar{y}_m|\bar{x}_n, \Theta^*)|x_n - (\bar{R}y_m + t)|^2.
\]

To minimize the distances, we first eliminate the brackets and take the derivative with respect to \( t \):

\[
S = \sum_{m,n=1}^{M,N} p(\bar{y}_m|\bar{x}_n, \Theta^*)(-2x_n + 2\bar{R}y_m + 2t).
\]

Solving for \( t \) yields:

\[
t = \frac{1}{NP} \sum_{m,n=1}^{M,N} p(\bar{y}_m|\bar{x}_n, \Theta^*)x_n - \frac{1}{NP} \sum_{m,n=1}^{M,N} p(\bar{y}_m|\bar{x}_n, \Theta^*)\bar{R}y_m = \mu_x - \bar{R}\mu_y.
\]

We summarize the algorithm in Figure S2.

**Algorithm for linear alignment from position and orientation**

**E-step** The joint distribution of \( x, \bar{x} \) and \( m, \bar{m} \) is (Equation 2 and 6):

\[
P(x, \bar{x}, m, \bar{m}) = \mathcal{N}(x; m, \sigma^2) \mathcal{F}(\bar{x}; \bar{m}, \kappa) \frac{1}{\sqrt{2\pi}} \delta_{\text{ind}(m)} \delta_{\text{ind}(\bar{m})}
\]

To compute the posterior \( P(m, \bar{m}|x, \bar{x}) \), we must again apply Bayes’ rule. We add an outlier term as in the previous algorithm and weight the priors as before with \( 0 \leq w \leq 1 \). The posterior then is:

\[
P(m, \bar{m}|x, \bar{x}) = \frac{P(x, \bar{x}, m, \bar{m})}{P(x, \bar{x})} = \frac{(1-w)\mathcal{N}(x; m, \sigma^2) \mathcal{F}(\bar{x}; \bar{m}, \kappa) \frac{1}{\sqrt{2\pi}} \delta_{\text{ind}(m)} \delta_{\text{ind}(\bar{m})}}{w \mathcal{N}(x; y_k, \sigma^2) \mathcal{F}(\bar{x}; \bar{y}_k, \kappa) \frac{1}{\sqrt{2\pi}} \delta_{k,k}} = \frac{\sum_{k=1}^{M} \exp(-\frac{1}{2\sigma^2}||x_n - s\bar{R}y_m - t||^2) \exp(\kappa \bar{x}_{n}^T \bar{R} \bar{y}_m)}{\sum_{k=1}^{M} \exp(-\frac{1}{2\sigma^2}||x_n - s\bar{R}y_k - t||^2) \exp(\kappa \bar{x}_{n}^T \bar{R} \bar{y}_k)} + \frac{wM}{(1-w)N} \exp(-\kappa)}
\]
**M-step** The expectation of the complete negative log-likelihood function is given by

\[
Q(s, \tilde{R}, t, \sigma^2, \kappa) = \frac{1}{2\sigma^2} \sum_{m,n=1}^{M,N} p^{old}(y, \tilde{y}|x, \tilde{x})||x_n - s\tilde{R}y_m - t||^2 + N_P \log \sigma
\]

\[
-\kappa \sum_{m,n=1}^{M,N} p^{old}(y, \tilde{y}|x, \tilde{x})\tilde{x}_n^T \tilde{R} \tilde{y}_m - N_P \log \left( \frac{\kappa}{e^{\kappa} - e^{-\kappa}} \right) + \text{const.}
\]  

(32)

To solve for each of the parameters during the M-Step, we first compute \( t \) by taking the partial derivative of \( Q \) with respect to \( t \), as we did in the last section, to obtain (equivalently to Myronenko and Song [28])

\[
t = \frac{1}{N_P} \sum_{m,n=1}^{M,N} p^{old}(y_m, \tilde{y}_m|x_n, \tilde{x}_n) x_n - \frac{1}{N_P} \sum_{m,n=1}^{M,N} p^{old}(y_m, \tilde{y}_m|x_n, \tilde{x}_n) \tilde{R} \tilde{y}_m = \mu_x - s\tilde{R} \mu_y
\]

(33)

We then plug \( t \) back into Equation 32. The log-likelihood becomes

\[
Q(s, \tilde{R}, \sigma^2, \kappa) = -\frac{1}{2\sigma^2} \sum_{m,n=1}^{M,N} p^{old}(y, \tilde{y}|x, \tilde{x})||\tilde{x}_n - s\tilde{R} \tilde{y}_m||^2 + \frac{N_P}{2} \log \sigma
\]

\[
-\kappa \sum_{m,n=1}^{M,N} p^{old}(y, \tilde{y}|x, \tilde{x})\tilde{x}_n^T \tilde{R} \tilde{y}_m - N_P \log \left( \frac{\kappa}{e^{\kappa} - e^{-\kappa}} \right) + \text{const.}
\]

(34)

where \( \tilde{x}_n = x_n - \mu_x \) and \( \tilde{y}_m = y_m - \mu_y \). We then write out \( ||\tilde{x}_n - s\tilde{R} \tilde{y}_m||^2 \) as a product \((\tilde{x}_n - s\tilde{R} \tilde{y}_m)^T (\tilde{x}_n - s\tilde{R} \tilde{y}_m)\), eliminate the brackets and write \( Q \) in matrix form:

\[
Q(s, \tilde{R}, \sigma^2, \kappa) = +\frac{1}{2\sigma^2}(\text{tr}(\tilde{X}^T \text{diag}(P^T) \tilde{X}) - 2s \cdot \text{tr}((\tilde{X}^T P^T \tilde{Y})^T \tilde{R})) + s^2 \text{tr}(\tilde{Y}^T \text{diag}(P) \tilde{Y})-
\]

\[
-\kappa \text{tr}((\tilde{X}^T P^T \tilde{Y})^T \tilde{R}) + N_P \log \sigma^2 - N_P \log \left( \frac{\kappa}{e^{\kappa} - e^{-\kappa}} \right) + \text{const.}
\]

(35)

We could not find a closed form solution for any of the parameters. Instead we search the optimum for the parameters numerically, except for \( t \), which we already eliminated. In general, using numerical optimization in this case is problematic due to the constraints \( \tilde{R}^T \tilde{R} = \mathbf{I} \) and \( \det(\tilde{R}) = 1 \). In our case, however, the rotation matrix is only two-dimensional and can therefore be parameterized by a single angle \( \rho \) that can be treated as one additional parameter. To find a numerical solution for this problem, usually first and second derivatives of the objective function are required. These are not hard to find from Equation 35, but it is tedious work. For the convenience of the reader, we give details on the derivative with respect to \( \rho \) below (Section Q derivatives) and list all first and second order derivatives in Figure S15. Figure S3 summarizes the algorithm.

The numerical optimization in the M-step is an unfortunate consequence of the scaling parameter \( s \). Without the scaling, the objective function \( Q \) could probably be minimized by formulating the transformation as dual quaternions and solving the transformation parameters as described by Walker [44]. While this still would not provide a final optimization in closed form, it might be possible to calculate closed form update equations for a generalized expectation maximization. Note that we cannot use other approaches that use scaled vectors such as Wells [26], because our orientations are unit vectors.

**Algorithm for elastic alignment**

For the elastic transformation model, \( y \) is transformed via a displacement vector \( \nu(y) \) (Myronenko and Song [28])

\[
T(Y, \nu) = Y + \nu(Y).
\]

(36)
Here, the task is to find a function $\nu$ that gives enough flexibility to capture the deformation of the data but still maintains the overall neighborhood structure of the points.

Myronenko and Song derive the displacement vectors by solving the regularized expectation of the negative log-likelihood function

$$f(\nu, \sigma^2) = Q_d(\nu, \sigma^2) + \frac{\lambda}{2} \Phi(\nu),$$

where $\Phi$ penalizes high frequencies of $\nu$. See [28] for a detailed derivation of the update equations for the M-step. The resulting vectors $\nu(y)$ are the product of an $M \times N$ smoothing matrix $G$ and, for the 2D case, an $M \times 2$ matrix $W$ by $T(y_m, \Theta) = y_m + (G(m, \cdot)W)^\top$, where $G(m, \cdot)$ denotes the $m$th row of $G$. The transformation parameters $\Theta$ here are the entries in the matrix $W$, whereas $G$ is initialized once and stays fixed.

**M-step** Coupling the transformation of positions and orientations is not obvious, because the transformation is formulated as position displacement vectors $\nu$ (Equation 36) and the rotational part of the transformation is not explicitly accessible. In practice, we assume that the elastic transformation contains only a minimal rotation that can be ignored, because we have applied a linear alignment before. With this assumption we can rewrite Equation 37 with the orientation as

$$f(\nu, \sigma^2, \kappa) = Q_d(\nu, \sigma^2) + \frac{\lambda}{2} \Phi(\nu) + Q_d(\kappa),$$

with

$$Q_d(\kappa) = -\kappa \sum_{m,n=1}^{M,N} P^{old}(y_m, \hat{y}_m|x_n, \hat{x}_n) \frac{\kappa}{2} ||x_n - \nu(y_m)||^2 + N_P \log \left( \frac{\kappa}{e^{\kappa} - 1} \right) + \text{const.}, \quad \text{and}$$

$$Q_d(\sigma^2) = \frac{1}{2\sigma^2} \sum_{m,n=1}^{M,N} P^{old}(y_m, \hat{y}_m|x_n, \hat{x}_n) ||x_n - \nu(y_m)||^2 + N_P \log \sigma^2 + \text{const.}$$

Because $Q_d$ is independent of the displacement vectors $\mu$, the optimization can be performed exactly as described in Myronenko and Song [28]. The only difference here is that the prior $P(m, \hat{m}|x, \hat{x})$ depends on $\kappa$, which must also be updated at each M-step. $\kappa$ can be updated in each iteration again with Newton’s method. With this strategy, the orientation only influences the result via the posterior $P^{old}(m, \hat{m}|x, \hat{x})$.

**E-step** The posterior again is computed by Bayes’ rule

$$P(m, \hat{m}|x, \hat{x}) = \frac{P(x, \hat{x}, m, \hat{m})}{P(x, \hat{x})} =$$

$$\frac{\exp\left(-\frac{1}{\sigma^2}||x - (y_m + (G(m, \cdot)W)^\top)||^2\right) \exp(\kappa \hat{\nu}^\top \hat{y}_m)}{\sum_{k=1}^M \exp\left(-\frac{1}{\sigma^2}||x - (y_k + (G(m, \cdot)W)^\top)||^2\right) \exp(\kappa \hat{\nu}^\top \hat{y}_k)} + \frac{\sigma^2}{(1-w)N} \frac{4\pi^2\sigma^2 (\exp(\kappa) - \exp(-\kappa))}{\kappa}$$

The algorithm is summarized in Figure S4.

**Implementation details**

To optimize $Q$ for the linear alignment from position and orientation (Algorithm S3), we use IPOpt [48], a free optimization library. As the starting point for the optimization, we use the parameters from the previous step, except for $\sigma^2$. In our experiments, $\sigma^2$ is not optimized correctly if the initial value is placed on the wrong side of the peak of the derivative of $Q$ (see Figure S14). To avoid this, we initialize $\sigma^2$ as

$$\sigma^2 = \frac{1}{2N_P} \text{tr}(\hat{X}^\top \text{diag}(P^\top 1)\hat{X}) - 2s \text{tr}(\hat{X}^\top P^\top \hat{Y}^\top \hat{R}) + s^2 \text{tr}(\hat{Y}^\top \text{diag}(P^\top 1)\hat{Y}),$$
which would be the minimum of $\sigma^2$ if the other parameters were fixed. In our experiments this was sufficient to guarantee that the initial $\sigma^2$ is placed on the right side.

Furthermore, the smaller $\sigma$ becomes the sharper the peak that is the minimum of the objective function becomes (see Figure S14). This sharp peak might make it hard to estimate a reasonable step size when optimization is done numerically. We believe that this is the reason why IPOpt does not always find an optimum in later expectation maximization iterations in our implementation. However, for our test data, we found that about 4000 iterations in IPOpt were sufficient to find a new parameter set that decreased the log-likelihood.

The running time of the algorithm for elastic alignment is $O(M^3)$. It can be reduced to linear by computing $P$ with a fast Gauss transform and $G$ by a low rank approximation as described by Myronenko and Song [28]. We did not implement this technique, which is why our implementation has cubic running time.

Q derivatives

Splitting the negative log-likelihood function (Equation 35) again into terms dependent and independent of $\tilde{R}$ (by Equation 4), $Q$ can be written as:

$$Q(s, \tilde{R}, \sigma^2, \kappa) = \frac{1}{2\sigma^2} \left( \text{tr}(\hat{X}^\top \text{diag}(P^\top P) \hat{X}) - 2s \cdot \text{tr}((\hat{X}^\top P^\top \hat{Y})^\top \tilde{R}) + s^2 \text{tr}(\hat{Y}^\top \text{diag}(P^\top P) \hat{Y}) + \frac{\kappa}{2} \text{tr}(D^\top \tilde{R}) + \text{tr}(E) \right)$$

$$- NP \log \sigma^2 - NP \log \left( \frac{\kappa}{2\pi(e^{\kappa}-e^{-\kappa})} \right) + \text{const.}$$

(42)

To compute the partial derivative with respect to $\tilde{R}$, we note that $\tilde{R}$ is parametrized via one rotation angle $\rho$ as $\tilde{R} = \begin{pmatrix} \cos \rho & -\sin \rho \\ \sin \rho & \cos \rho \end{pmatrix}$ in the 2d case. Writing the trace as a sum and taking partial derivatives with respect to this angle we find that

$$\frac{\partial \text{tr}(K^\top \tilde{R})}{\partial \rho} = \text{tr}(K^\top \frac{\partial \tilde{R}}{\partial \rho})$$

(44)

and

$$\frac{\partial \tilde{R}}{\partial \rho} = \tilde{R}' = \begin{pmatrix} -\sin \rho & -\cos \rho \\ \cos \rho & -\sin \rho \end{pmatrix}$$

(45)

The second derivative can be found equivalently.

$$\frac{\partial^2 \tilde{R}}{\partial \rho^2} = \tilde{R}'' = \begin{pmatrix} -\cos \rho & \sin \rho \\ -\sin \rho & -\cos \rho \end{pmatrix}$$

(46)

We give the partial derivatives of Equation 43 with respect to $s, \rho, \sigma^2$ and $\kappa$ in Figure S15.