Detecting symmetry in ascidian embryo

Grégoire Malandain, Patrick Lemaire

DETECTING SYMMETRY IN ASCIDIAN EMBRYO

Grégoire Malandain[⋆] *Patrick Lemaire*[†]

[⋆] Université Côte d'Azur, Inria, CNRS, I3S, France
[†] CRBM, Université de Montpellier, CNRS, France

ABSTRACT

Modern microscopy techniques allow to acquire temporal sequences of 3D stacks. It provides a unprecedented means to image developing organs or embryos at the cellular level, but results in a huge quantity of data. Proposing efficient processing methods is then absolutely crucial. We are addressing here the study of ascidian embryo development. The stereotypy of their development enables population studies at a cellular level but requires embryo cells to be named. Cell naming can be achieved by name transfer from a database, which, in turn, requires embryo-to-embryo registration, that is eased by embryo symmetry detection [1]. In this article, we propose a novel symmetry detection method for that purpose.

Index Terms— Fluorescence microscopy, symmetry detection, embryogenesis

1. INTRODUCTION

Developmental biology aims at understanding the dynamics of formation of tissues or organs within an organism. To study development, many model organisms are used [2]. Most of them are animals, like the C. elegans nematod or the Drosophila. Each of these models has its own specificity, which may make it more suitable for observation and/or analysis. For instance, C. elegans is transparent, which makes it ideal for observation under a light microscope, and it has a remarkable stereotyped development, every adult having exactly the same number of cells, and each cell having its homologue in other individuals: in other words, each cell can be identified by its name [3]. This last property allows not only cell-based comparisons in population studies, but also aggregating different information coming from several individuals.

When coming to vertebrates, mice are usually the model of choice. However, as the closest relatives to vertebrates [4], tunicates are also widely used as models to study embryonic development, ascidians being the largest group within tunicates. The embryos of some ascidian species are transparent, making them suitable for microscopic observation. Last, the development of ascidian embryos (in the first stages) shows a strict left/right bilateral symmetry and is stereotyped down to the cellular level, even across species. As homologous cells can be recognized across individuals within and between species, ascidian embryos are also ideal to aggregate information from a population and compare development between individuals [5]. This development stereotypy has been known since the late 19th century [6, 7], and the currently used cell nomenclature was proposed in 1905 [8].

Live microscopy is a method of choice [9] for embryogenesis studies in ascidians. Recent technical advances, including light-sheet fluorescence microscopy [10], allow to acquire temporal series of 3D images, which can be automatically segmented, thus enabling to follow organismal development with a spatial subcellular resolution and at high temporal frequency. To compare a newly acquired ascidian embryo with a database of previously acquired ones, it is mandatory to name its cells.

This can be done by name transfer, assuming cells can be paired between two different segmented embryos. Embryo registration, as described in [1, section 4], can efficiently achieve this pairing. Registration is conducted with an ICP algorithm [11] where cell centers of mass are the features to be registered. ICP being an iterative optimization method, its result depends on the starting point (here an initial transformation). In [1], multiple initial transformations are explored, to ensure one will be close enough to the global optimum (ie the right transformation). Instead of sampling the set of 3D rotations, we took advantage of the left-right symmetry of ascidian embryos. A symmetry vector candidate of the embryo to be named is aligned with the symmetry vector of a database embryo (this allows to superimpose the symmetry planes) and then 2D rotations around the symmetry vector defined a set of initial transformations to be tested. The registration computational cost is then proportional to the number of symmetry vector candidates to be tested. For such a strategy, it is crucial to efficiently detect the symmetry direction of embryos, which is the purpose of this article.

2. DATA

We have at hand 7 temporal sequences of 3D images issued from [12], named `Pm[1,3,4,5,7,8,9]`, each sequence being composed of around 100 time points. An example of a segmented temporal series can be seen at morphonet.org/dataset/1095. A sequence of such segmented images may be up to 1 Tb of data. For an easier analysis, each sequence can be summarized in a so-called property file (encoded in XML) that contains, for each 3D image, cell information (cell volume, surface, center of mass, neighboring cells together with a measure of their contact surface, etc), plus the temporal cell lineage information that links cells between consecutive 3D images keeping track of cell divisions. These embryo property files can then be augmented by useful information, for example cell names. Such a property file weights only a few tens of Mb, and is thus an efficient means to represent a population of embryos. The purpose of this work is the detection of the embryo symmetry direction directly from its property file. For an easier comparison between embryos, acquisition times are first linearly warped so that a common *developmental time* value corresponds approximately to the same cell count in each embryo (see Fig. 1). This compensates for different acquisition starting timepoint across embryos as well as for different developmental speeds

Cell names are of the form $[a,b]p.q[\_,*]$ where $[a,b]$ indicates a anterior or posterior position, p denotes the generation, q allows to individualize the cell within a generation, and $[\_,*]$ indicates whether

Corresponding author: gregoire.malandain@inria.fr

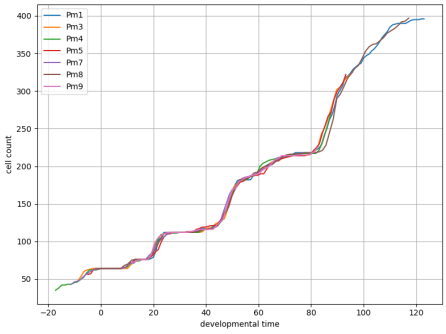

Fig. 1. Time warping allows to align cell counts across embryos.

the cell is in the left or right hemisphere. The (unit) vector from the center of mass of the left hemisphere (the '_' cells) to the one of the right hemisphere (the '$*$' cells) will be the ground truth we will compare to.

3. DIRECTION DISTRIBUTION

We hypothesize that vectors aligned to the symmetry direction may be extracted from the data, so that a mode, corresponding to the symmetry direction, can be extracted from a vector distribution. This sections aims at describing this distribution computation.

Let assume we have a number of samples $N = \{(w_i, \mathbf{n}_i)\}$ where $\mathbf{n}_i$ is an unit vector and w_i its associated weight.

To compute the vector distribution, we discretize the unit sphere into the set of unit vectors $\{\mathbf{u}_k\}$ (typically around 4000 vectors). The kernel-based [13] vector distribution density $d(\mathbf{u}_k)$ at unit vector $\mathbf{u}_k$ is estimated by

$$d(\mathbf{u}_k) = \frac{1}{2\sum_i w_i}\left(\sum_i w_i \frac{\phi(\mathbf{n}_i, \mathbf{u}_k)}{\Phi(\mathbf{n}_i)} + \sum_i w_i \frac{\phi(-\mathbf{n}_i, \mathbf{u}_k)}{\Phi(-\mathbf{n}_i)}\right)$$

where

$$\phi(\mathbf{n}, \mathbf{u}_k) = \exp \frac{-\arccos(|\mathbf{n}.\mathbf{u}_k|)^2}{2\sigma^2} \text{ and } \Phi(\mathbf{n}) = \sum_k \phi(\mathbf{n}, \mathbf{u}_k)$$

The Gaussian is used as kernel (typically $\sigma = 0.15$ radians). For every sample $\mathbf{n}_i$ we add the opposite vector (since we want a direction distribution, not a vector distribution), hence we end up with $2|N|$ samples. Normalization by $\Phi(\mathbf{n})$ ensures that every sample has exactly a w_i contribution to the distribution. Once the distribution has been computed (eg see Fig. 2), we extracted its local maxima (ie modes) that are considered as symmetry vector candidates. Maxima can be sorted (in decreasing order) with respect to their distribution values. The distribution is symmetrical (since we added the opposite samples), this implies that every couple of two successive maxima have the same distribution value and are opposite vectors (corresponding to one only direction). The detection of the symmetry direction is efficient if

- the maximum closest to the ground truth symmetry vector is of low rank (hence only a few candidates have to be tested to co-register the embryos), and
- the angle of this maximum with the ground truth symmetry direction is small (hence the registration is likely to be successful).

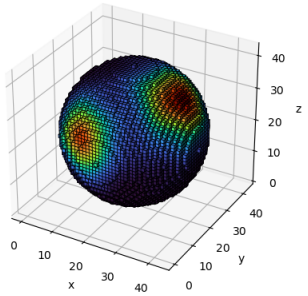

Fig. 2. Direction distribution example with two modes in directions $(1, 0, 1)$ and $(-1, -3, 1)$.

4. SYMMETRY DETECTION

In [1], the symmetry detection was based on images. A membrane detection filter was computed [14], which allowed to extract an image of directional maxima (at the center of the membrane) thanks to an estimation of the membrane normal vector. Directional maxima were subsequently thresholded, resulting in a binary image B, and it was hypothesized that most of the membrane vectors were aligned to the symmetry direction. The direction distribution was computed with the samples $\{(1, \mathbf{n}(i, j, k))|B(i, j, k) \neq 0\}$ where $\mathbf{n}(i, j, k)$ is the membrane normal vector at voxel (i, j, k).

4.1. Symmetry from membranes

In [1], the symmetry was deduced from local information, ie the orientation of membrane at voxels separating cells. This local information is no more at hand, but we can mimic [1] by using more regional information. Consider two cells c_k and c_ℓ. In an image, there were a number of voxels from the membrane separating c_k from c_ℓ that participate to the direction distribution estimation. It seems reasonable to assume that

- this number of voxels is equal to the contact surface $s_{k,\ell} = s(c_k, c_\ell)$ measured between cells c_k from c_ℓ,
- the contact surface normal vector is aligned with the direction joining the cell centers of mass, g_ℓ and g_k.

Let $\mathbf{n}_{k,\ell} = (g_\ell - g_k)/\|g_\ell - g_k\|$ denotes the unit vector from cell c_k to c_ℓ. It comes to compute the distribution with the set of samples $\{(s_{k,\ell}, \mathbf{n}_{k,\ell})\}$ where (c_k, c_ℓ) are all the pairs of adjacent cells.

Fig. 3 (top left) indicates the rank of the distribution maximum closest to the embryo symmetry vector (the best candidate). It comes out that the worst case is a 12th rank, meaning that we would have to retain a lot of symmetry candidates (distribution maxima) to co-register two embryos. Moreover, a few of these best candidates have an angular error of more of 15 degrees (Fig. 3 bottom left): this may impair the ability of ICP to converge to correctly register the two embryos. The reported results under-performed with respect to the princeps method of [1]. This may be due to the poor ability of the vector joining the centers of mass of two adjacent cells to represent the normal direction of the membrane separating the two cells.

4.2. Symmetry from cells

Identifying pairs of homologous symmetrical cells in the left and right hemispheres would be a more efficient way to detect the symmetry direction. We hypothesize here that, thanks to the left/right bilateral symmetry of the embryo, the contact surface pattern is conserved between symmetrical cells: we then build measures upon these contact surfaces. To compare the two cells c_k and c_ℓ, we have then to pair the neighboring cells $\{c_{k_i}\}$ of c_k to the neighboring cells $\{c_{\ell_j}\}$ of c_ℓ (see Fig. 4).

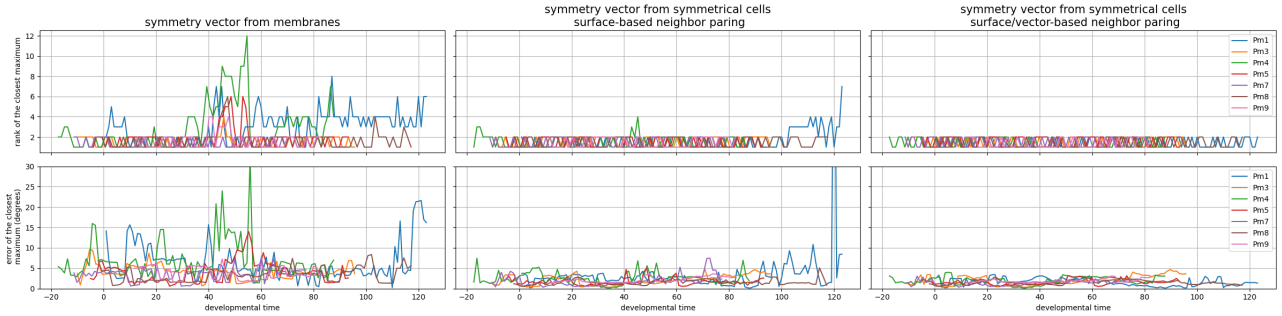

Fig. 3. Top row: rank of the vector distribution maximum closest to the embryo symmetry. Bottom row: angular error of the vector distribution maximum closest to the embryo symmetry. From left to right: symmetry computed from membranes, from symmetrical cells with surface-based neighbor pairing, from symmetrical cells with surface/vector-based neighbor pairing.

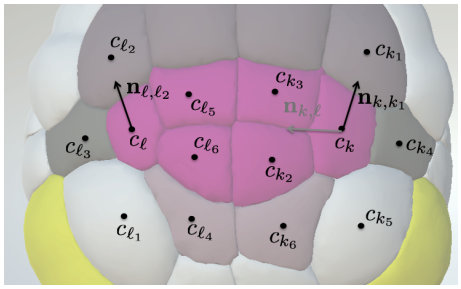

Fig. 4. To compare cells c_k and c_ℓ, their neighbors have to be paired. Here, perfect pairing will be $\{(k_1, \ell_2), (k_4, \ell_3), (k_5, \ell_1), \ldots\}$.

4.2.1. *Surface-based neighbor pairing*

Let consider we have a set of neighbor pairings $P = \{(k_i, \ell_j)\}$ and the normalized L1 distance between the contact surface vectors (with 0 padding to get the same vector length)

$$d_s(c_k, c_\ell; P) = \frac{\sum_{(k_i,\ell_j)\in P} |s_{k,k_i} - s_{\ell,\ell_j}|}{\sum_{i_k} s_{k,i_k} + \sum_{j_\ell} s_{\ell,j_\ell}} \quad (1)$$

$\sum_{i_k} s_{k,i_k} = s_k$ is nothing but the surface of cell c_k. Dividing by the sum of the two cell surfaces allows to get a measure in $[0, 1]$, 0 meaning a perfect equality of the paired contact surfaces. The index $\hat{k}$ of the most likely symmetrical cell for cell c_k can be found by minimizing this normalized distance: $\hat{k} = \arg\min_\ell (\min_P d_s(c_k, c_\ell; P))$. The optimal pairing for $d_s(c_k, c_\ell; P)$ is obtained when the neighbors are sorted with respect to their contact surfaces, so that both largest contact surfaces are paired together, then the second largest contact surfaces are paired together, and so on. Let denote this surface-based optimal pairing by P_s^* and let $d_s^*(c_k, c_\ell) = d_s(c_k, c_\ell; P_s^*)$. Thus the index $\hat{k}$ of the most likely symmetrical cell for cell c_k is given by

$$\hat{k} = \arg\min_\ell d_s^*(c_k, c_\ell) \quad (2)$$

The direction distribution can be built from the unit vectors joining each cell and its symmetrical homologous cell, $\mathbf{n}_{k,\hat{k}}$. We weighted the vector by the cell surface multiplied by $1 - d_s^*(c_k, c_{\hat{k}})$ to account for poor similarity between paired cells. It comes to build the distribution with the set of samples $\{(s_k(1 - d_s^*(c_k, c_{\hat{k}})), \mathbf{n}_{k,\hat{k}})\}$.

The distribution maximum closest to the embryo symmetry vector (the best candidate) is almost always the first or the second (Fig. 3 top middle), meaning that the largest maximum gives the embryo symmetry direction (recall that the distribution is symmetrical, so the two first maxima are opposite vectors), with an error (Fig. 3 bottom middle) almost always less than 10 degrees. The symmetry detection failed for one time point of embryo `Pm1` where the error reached 96 degrees: the embryo has a little bit less than 400 cells at this particular time.

4.2.2. *Surface/vector-based neighbor pairing*

In the previous section, symmetrical cells were only assessed based on adjacency information (contact surfaces), without accounting for neighbors ordering or relative spatial position.

However, if cells c_k and c_ℓ are symmetrical (see Fig. 4) and if the neighbors c_{ℓ_2} and c_{k_1} are homologous, then the vector $\mathbf{n}_{\ell,\ell_2}$ is the mirror vector of $\mathbf{n}_{k,k_1}$ with respect to $\mathbf{n}_{k,\ell}$. We thus add the directional information towards the neighbors in the figure of merit to better characterize symmetrical cells. If we hypothesize that $\mathbf{n}_{k,\ell}$ is the embryo symmetry direction, then the mirror vector of $\mathbf{n}_{\ell,\ell_j}$, denoted by $\mathbf{n}'_{\ell,\ell_j}$ is calculated by

$$\mathbf{n}'_{\ell,\ell_j} = \mathbf{n}_{\ell,\ell_j} - 2(\mathbf{n}_{\ell,\ell_j}.\mathbf{n}_{k,\ell})\mathbf{n}_{k,\ell}$$

Comparing neighbors comes to replace the surface difference, $|s_{k,k_i} - s_{\ell,\ell_j}|$, by $\|s_{k,k_i}\mathbf{n}_{k,k_i} - s_{\ell,\ell_j}\mathbf{n}'_{\ell,\ell_j}\|$ in the figure of merit that assesses symmetrical cell similarity, thus introducing

$$d_v(c_k, c_\ell; P) = \frac{\sum_{(k_i,\ell_j)\in P} \|s_{k,k_i}\mathbf{n}_{k,k_1} - s_{\ell,\ell_j}\mathbf{n}'_{\ell,\ell_j}\|}{\sum_{i_k} s_{k,i_k} + \sum_{j_\ell} s_{\ell,j_\ell}} \quad (3)$$

However, computing the optimal pairing (the one that minimizes $d_v(c_k, c_\ell; P)$) between neighbors of c_k and c_ℓ is no more straightforward. It is an assignment problem where one has to pair neighbors c_{k_i} of c_k with neighbors c_{ℓ_j} of c_ℓ with a minimal sum of costs, the cost of a pairing being $\|s_{k,k_i}\mathbf{n}_{k,k_i} - s_{\ell,\ell_j}\mathbf{n}'_{\ell,\ell_j}\|$. It can be efficiently solved by minimum weight matching methods in bipartite graphs[1].

Let denote this surface-based optimal pairing by P_v^*, the index $\hat{k}$ of the most likely symmetrical cell for cell c_k is given by

$$\hat{k} = \arg\min_\ell d_v(c_k, c_\ell; P_v^*) \quad (4)$$

The direction distribution can be built from the set of samples $\{(s_k(1 - d_v(c_k, c_{\hat{k}}; P_v^*)), \mathbf{n}_{k,\hat{k}})\}$.

Now, the distribution maximum closest to the embryo symmetry vector (the best candidate) is always the first or the second (Fig. 3 top

[1] We use `linear_sum_assignment` from `scipy`.

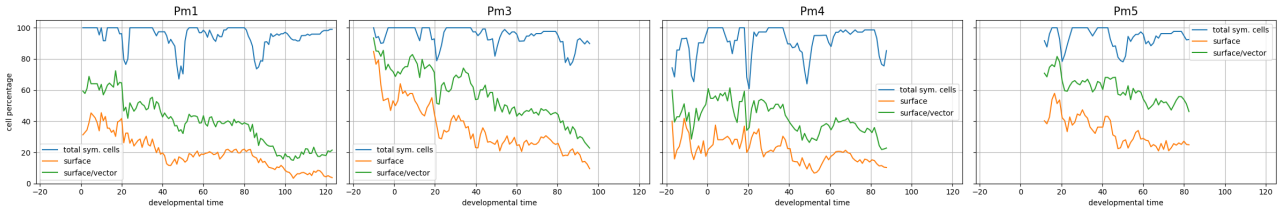

Fig. 5. Identification of symmetrical cells for 4 embryos. Blue: percentage of cells having a symmetric homologous cell; orange: percentage of symmetric cells correctly identified with the surface-based neighbor pairing; green: percentage of symmetric cells correctly identified with the surface/vector-based neighbor pairing.

right), meaning that the largest maximum always gives the embryo symmetry direction, with an error (Fig. 3 bottom right) always less than 5 degrees. Please note that this error could be decreased if we increase the number of vectors used to discretize the sphere (Sec. 3).

4.2.3. Symmetrical cell identification assessment

Fig. 3 demonstrates that the normalized distances (Eqs 1 or 3) seems to be powerful enough to build a direction distribution where the main maximum is close to the embryo symmetry direction. This suggests that Eqs 2 or 4 are able to find the correct pairs of symmetrical cells. Since we have named embryos, this can be easily assessed. Indeed, the symmetrical cell of a cell named $[a,b]p.q\_$ is named $[a,b]p.q*$ and conversely. Fig. 5 depicts the count of correctly identified symmetric homologous cells for 4 embryos. Since symmetric homologous cells did not divide at the same exact time in the right and left hemisphere, some of the cells do not have their symmetric counterpart: the blue curve indicates the percentage of cells having a symmetrical homologous cell at the given time. Low values corresponds to rounds of cell divisions. It can be observed that correct identification decreases with the cell count (as time increases). Surface/vector-based neighbor pairing outperforms the surface-based neighbor pairing for correct symmetric cell identification for a percentage roughly twice larger.

5. CONCLUSION

This article proposed a novel method for ascidian embryo left-right symmetry detection, which can be used for further embryo co-registration. Although we only have at hand textual summaries of segmented images, the (best) proposed method allows to achieve the detection of the plane of left-right symmetry in all the embryo time-points with an error less than 5 degrees. This not only outperformed our previous method [1] in terms of accuracy, but it also provides a much more efficient method since the calculation of the symmetry detection is a matter of seconds (with a `python` code) while it took several tens of minutes with the previous method (code in C).

As a side result, it suggests that contact surfaces may be an efficient signature for cell identification.

Compliance with Ethical Standards: this is a numerical simulation study for which no ethical approval was required.

Acknowledgments: this work was supported by the Cell-whisper (ANR-19-CE13-0020) grant of French National Research Agency.

Code: The described methods have been implemented within the ascidian package (see astec.gitlabpages.inria.fr/ascidian). Figs 3 and 5 have been built upon results issued from this package (see Gallery).

6. REFERENCES

[1] G Michelin, L Guignard, UM Fiuza, P Lemaire, C Godin, and G Malandain, "Cell Pairings for Ascidian Embryo Registration," in *ISBI*, Apr. 2015.

[2] U Irion and C Nüsslein-Volhard, "Developmental genetics with model organisms," *Proc Natl Acad Sci U S A*, vol. 119, no. 30, July 2022.

[3] J E Sulston, E Schierenberg, JG White, and JN Thomson, "The embryonic cell lineage of the nematode caenorhabditis elegans," *Dev Biol*, vol. 100, no. 1, pp. 64–119, November 1983.

[4] A Fodor, J Liu, L Turner, and BJ Swalla, "Transitional chordates and vertebrate origins: Tunicates," *Curr Top Dev Biol*, vol. 141, pp. 149–171, TI 2021.

[5] M Brozovic, C Martin, C Dantec, D Dauga, M Mendez, P Simion, M Percher, B Laporte, and alii, "Aniseed 2015: a digital framework for the comparative developmental biology of ascidians," *Nucleic Acids Res*, vol. 44, no. D1, 2016.

[6] CA Kofold, "On some laws of cleavage in limax. a preliminary notice," *Proceedings of the American Academy of Arts and Sciences*, vol. 29, no. 180-203, 1893.

[7] WE Castle, "The early embryology of ciona intestinalis, flemming (l.)," *Bulletin of the Museum of Comparative Zoology*, vol. XXVII, no. 7, pp. 202–310, 1896.

[8] EG Conklin, *The organization and cell-lineage of the ascidian egg*, Philadelphia :[Academy of Natural Sciences], 1905.

[9] PJ Keller, "Imaging Morphogenesis: Technological Advances and Biological Insights," *Science*, vol. 340, no. 6137, pp. 1234168+, June 2013.

[10] U Krzic, S Gunther, TE Saunders, SJ Streichan, and L Hufnagel, "Multiview light-sheet microscope for rapid in toto imaging," *Nat Methods*, vol. 9, no. 7, pp. 730–3, 2012.

[11] PJ Besl and ND McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, February 1992.

[12] L Guignard, UM Fiuza, B Leggio, J Laussu, E Faure, G Michelin, K Biasuz, L Hufnagel, G Malandain, C Godin, and P Lemaire, "Contact area-dependent cell communication and the morphological invariance of ascidian embryogenesis," *Science*, vol. 369, no. 6500, pp. 158, July 2020.

[13] E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

[14] G Michelin, L Guignard, UM Fiuza, and G Malandain, "Embryo Cell Membranes Reconstruction by Tensor Voting," in *International Symposium on Biomedical Imaging*, Apr. 2014.