A Method for Two-Dimensional Registration and Construction of the Two-Dimensional Atlas of Gene Expression Patterns in Situ

Konstantin Kozlov, Ekaterina Myasnikova, Andrei Pisarev, Maria Samsonova, and John Reinitz

ABSTRACT: We apply the fast redundant dyadic wavelet transform to the spatial registration of two-dimensional gene expression patterns of 736 Drosophila melanogaster embryos. This method is superior to the Fourier transform or windowed Fourier transform because of its ability to reduce noise and is of high resolution. In registration of the dataset we use two cost functions based on computing the Euclidean or Mahalanobis distance. The algorithm shows a high level of accuracy. For early temporal classes the cost function based on Mahalanobis distance gives better results.

INTRODUCTION: Recent progress in the molecular genetics of eukaryotes has revealed the detailed structure of the coding regions of genes as well as cis acting elements such as promoters, enhancers, the binding sites, which these elements contain, as well as information about the interactions of these sites with trans-acting components. These genetic elements, together with their RNA and protein products, comprise the genomic regulatory network that coordinates patterns of gene expression, which control cell type specification, tissue differentiation and ultimately the ontogeny of the zygote. Understanding the structure, logic, inte-
K. Kozlov et al. / A Method for Two-Dimensional Registration and Construction

grated dynamical behavior and evolution of such networks is central to molecular and developmental biology.

We are addressing this problem in the context of a particular biological system, namely the segment determination system of the fruit fly *Drosophila melanogaster*. Our aim is to decipher the molecular mechanisms, which control this process by means of an integrated approach using both theory and experiment [12, 16, 17, 15, 19]. The theoretical component uses mathematical modeling and simulations, while experimental part employs automated image processing and data analysis methods to quantitatively characterize the expression patterns of segmentation genes both in space and time.

Segment determination takes place early in development during the syncytial blastoderm stage. At this stage the embryo is a hollow shell of nuclei not separated by cell membranes. The blastoderm is roughly ellipsoidal with the anterior-posterior (A-P) axis of the embryo lying on the long axis of the ellipsoid. The ellipsoid has residual asymmetry, in the sense that the anterior end is more sharply curved than the posterior and the ventral midline more curved than the dorsal (Fig. 1).

Fig. 1. Embryo stained for the protein products of *bicoid*, *caudal* and *eve*. The anterior end of the embryo is to the left and the dorsal side is up. The nuclei are seen as little colored dots. This is embryo be10 from FlyEx.

The genetic network, which controls segmentation in the fruit fly *Drosophila*, is one of the few genetic networks fully characterized at a genetic and functional level. The initial determination of the segments is a consequence of the expression of 16 genes [14], which are mainly transcription factors [2, 5]. Several of these genes are expressed from the maternal genome to provide asymmetric initial conditions. The others are zygotic and are expressed in patterns that become more spatially refined over time. This refinement is a consequence of members of the network regulating one another’s synthesis in a precise manner. Of particular importance are members of the "gap" and "pair-rule" classes of segmentation genes. Gap genes are expressed with unimodal or bimodal concentration profiles which become gradually steeper, while pair-rule genes initially express protein in a single very broad domain that restricts to seven narrow domains over a relatively short time interval. In general, each gap and pair-rule gene expresses protein in a different set of locations, but these have a characteristic overlap with one another. Thus one can view the expression pattern as a collection of "domains", each of which is a region of expression containing one concentration maximum.

In our experiments gene expression is monitored by confocal scanning microscopy of fixed embryos stained with fluorescence tagged antibodies. The confocal microscope at our disposal permits scanning for the expression of up to three genes at once. Thus to construct a map of the expression patterns of all segmentation genes at the given time we have to combine data from many embryos, each stained for different combinations of three genes.
We have reported previously how this problem can be solved by applying image registration methods to data extracted from a central strip that is 10% of the dorsal-ventral (D-V) axis in width and runs along the anterior-posterior axis of the embryo. These registration methods belong to the point mapping class [3] and are based on extracting “ground control points” (GCPs), which are a small number of characteristic features in each image and using coordinate transformations to make the images coincide as closely as possible. We extract GCPs using a spline approximation [11] or a wavelet decomposition [8] of gene expression levels. This work results in the construction of a one-dimensional map of segmentation gene expression [12]. Here we extend the registration method to two dimensions. In addition we develop a method for construction of a two-dimensional integrated patterns of segmentation gene expression. This method provides the preservation of nuclear structure of a two-dimensional gene expression pattern. It is based on creating an averaged model that reproduces the spatial distribution of nuclei over the embryo image. The average concentrations of each protein in each averaged nucleus are computed from a set of embryos of the same age.

METHODS

Dataset

In our experiments gene expression was measured by confocal scanning microscope. Embryos were stained with fluorescence tagged antibodies [7], put on a glass slide in 50% glycerol and covered with a cover glass. The cover glass presses down on the embryo, tending to flatten it. This has the disadvantage of distorting the curved surface of the blastoderm into a plane, but it has the advantage that over a third of the embryo can be scanned for gene expression using only three optical sections, which are averaged. For each embryo a $1024 \times 1024$ pixel image with 8 bits of fluorescence data in each of 3 channels was obtained. Each embryo was scanned for the expression of three genes at a time. Three embryo images are combined and the resultant image is segmented to construct a binary nuclear mask. In this operation, each pixel on the image is classified as belonging to a nucleus or not by generating an image called a "mask". In the mask, a pixel is equal to one if and only if that pixel is located on a nucleus. The mask was used to reduce the embryo image to a table containing quantitative data on gene expression in each embryo nucleus [6, 24]. Each nucleus is characterized by a unique identification number, the $x$ and $y$ coordinates of its centroid and the average fluorescence levels of three proteins. The embryo is oriented so that the $x$-axis corresponds to the anterior-posterior (A-P) axis of the embryo and the $y$-axis to the dorsal-ventral (D-V) axis. $x$ and $y$ coordinates are expressed as percent of the maximum size of the embryo in the $x$ and $y$ directions.

We used 736 wild type embryos of Oregon-R flies, scanned for the expression of 14 segmentation genes. These are the maternal coordinate genes $\text{bicoid (bcd)}$ and $\text{caudal (cad)}$, the gap genes $\text{Kruppel (Kr)}$, $\text{hunchback (hb)}$, $\text{giant (gt)}$, $\text{knirps (kni)}$ and $\text{tailless (tl)}$; and the pair-rule genes $\text{even-skipped (eve)}$, $\text{runt (run)}$, $\text{hairy (h)}$, $\text{fushi-tarazu (ftz)}$, $\text{odd-skipped (odd)}$, $\text{paired (prd)}$, and $\text{sloppy-paired (slp)}$.

All the embryos belong to cleavage cycle 14A [4]. This cycle is about an hour long and is characterized by a rapid transition of the pair-rule gene expression patterns, culminating in the formation of 7 stripes.

We have subdivided all the embryos in cycle 14A into 8 temporal classes [11] and have established that these classes correlate well with timed morphological changes from in vivo observations [10]. Each image is allocated to one of the temporal classes on the basis of thorough and extensive visual inspection of pair-rule gene expression patterns, in particular that of $\text{eve}$, which is highly dynamic.

The evolution of the expression patterns of $\text{eve}$ during cleavage cycle 14A shows the following tendency. Time classes 1, 2 and 3 do not have seven well-defined $\text{eve}$ stripes and the number and location of stripes changes rapidly. The remaining groups (classes 4 to 8) do have seven well-defined stripes. Initially the amplitude of stripes decreases posteriorly. By the end of cycle 14A, all $\text{eve}$ stripes have reached
maximum and equal intensity and maximum sharpness. In this paper we analyze embryos belonging to time classes 2 through 8.

**Approach to the two-dimensional registration problem**

Although the expression of segmentation genes is largely a function of position along the anterior-posterior (A-P) axis of the embryo body, there is residual curvature in the D-V direction (Fig. 1 and Fig. 2). The biological reason for such curvature is most likely compensation for the greater arc length from the anterior to posterior pole along the ventral midline compared to that along the dorsal midline. Although it has been proposed that this curvature is a consequence of a Turing mechanism [20], we believe it to be a consequence of residual D-V asymmetry in maternal gradients, since in the anterior half of the embryo the curvature of stripes closely follows contours of bicoid concentration [21]. Although the extent of the curvature is exaggerated by mechanical deformation in the embryos considered here, it is clearly seen in non-deformed whole mounts (J. Reinitz, unpublished observations) and in tissue sections hybridized to RNA *in situ* [22]. In previous work [12], we avoided the problem of curvature by considering data along a transect in the A-P direction. The transect had a width of 10% of the total D-V extent of the embryo (50% to 60%) in order to ensure a sufficient number of samples.

![Fig. 2. Two-dimensional expression pattern of the *eve* gene. The domains are approximately vertical in between the lines that correspond to the 50 and 60% of embryo height](image)

The first main result of this paper is that features within this strip are in fact sufficient to induce registration over the entire visible portion of the embryo. This is equivalent to the biologically reasonable statement that information about A-P gene expression is mainly contained in one-dimensional data. This property defines our approach to the two-dimensional registration problem.

We extracted GCPs for registration from the pattern of *eve* expression. This expression pattern is rich in features and so we scanned for Eve protein in all embryos of the dataset. The GCPs were extracted from segmented data excised from the central 10% of *y* values on the midline of an embryo in the A-P direction (*x* coordinate). The *y* values of these data are then ignored and the patterns, demonstrating the variation of *eve* gene expression along the *x*-axis, are presented as diagrams.

The stripes and interstripes on the curve have the form of peaks and valleys respectively (Fig. 3). It is natural to use extrema of *eve* gene expression domains as GCPs for registration. After the GCPs are extracted we find the coordinate transformation that minimizes the distance between corresponding GCPs in different images. At the next step this transformation will be applied to the whole two-dimensional image.

Figures 3a and 3d show an example of *eve* gene expression patterns in the central 10% strip of embryos belonging to temporal classes 3 and 8 respectively. Due to the noise on the peaks it is rather difficult to find the locations of the extrema directly from the raw data. Therefore to find the characteristic features of this pattern we have formulated an algorithm based on the Fast Redundant Dyadic Wavelet Transform (FRDWT) with automatic detection of the correct decomposition level [18, 23]. In the follow-
ing section, we review the FRDWT and discuss its application to feature extraction in *Drosophila* embryos.

**Definition and application areas of the wavelet transform**

An image registration problem requires representing a signal as a set of characteristic features. This set must be of as small size as possible but unique for each signal. Consequently, a transform used for this purpose must have the following properties:

- high resolution
- noise reduction
- stability or shift invariance.

The classical approaches based on the standard Fourier transform possess certain drawbacks, inasmuch as the standard Fourier transform provides poor localization in space, while the windowed Fourier transform uses one and the same window for all the frequencies. The wavelet transform is free of these defects, as we explain below.

**Definition of the FRDWT**

The FRDWT is defined by the equation

\[
(W_{\psi}f)(a,k) = a^{-1/2} \int_{-\infty}^{\infty} \overline{\psi} \left( \frac{x-k}{a} \right) f(x) dx,
\]

where \(\overline{\psi}\) denotes the complex conjugate of \(\psi\).

The FRDWT uses a set of windows proportional to \(a = 2^j\) and transforms a function \(f(t)\) into two sequences,

- low pass: \(s_j(k) = (W_{\phi}f)(2^j, 2^j k)\)
- high pass: \(r_j(k) = (W_{\psi}f)(2^j, 2^j k)\)

where \(\psi\) is the wavelet and \(\phi\) is the scaling function.

Note that the FRDWT should not be confused with those wavelet-based methods which have become a very popular tool for image and signal compression. For these purposes, which require rather different properties of the wavelet transform, the non-redundant version with a discrete basis is used [1].

The basic idea of FRDWT is that it decomposes the signal into two sequences: the so-called “high pass” and "low pass". The number of observations in each sequence equals that in the original sample because of the redundancy of the transform and hence no information is lost about the localization of extrema. At each level of the decomposition the input signal is smoothed by removing the noise of a certain frequency and is placed into the low pass. The high pass contains the information on the features of the input signal [23]. The choice of scaling function \(\phi\) and wavelet \(\psi\) is defined only by general restrictions imposed on these functions and the type of features to be extracted. To localize the extrema, i.e. zeros of the first derivative, we apply basis functions that put the characteristics of the first derivative into the high pass. In solving the image registration problem the basic properties of FRWDT, such as a noise reduction and good localization of characteristic features, are required.

Malat et al. [9] have developed a method for extracting features of the first derivative of a signal on different scales using cubic B-spline wavelets. To apply this algorithm to the spatial registration of gene expression patterns the optimal level of signal decomposition is detected automatically. This modification gives the opportunity to use the wavelet technique more efficiently and to interpret the results in a straightforward way.
Wavelet decomposition: algorithm

Mallat’s algorithm is based on the fact that for a given discrete signal \( f(k) \) there exists a unique representation by a weighted sum of shifted B-splines

\[
f(x) = \sum_{k \in \mathbb{Z}} f_0(k) \beta_x(x - k),
\]

that provides the exact interpolation \( f(x) |_{x=k} = f(k) \) (see [23]). Here \( f_0(k) \) are the so-called B-spline coefficients, \( \beta_{0,x} \) is a discrete kernel of the B-spline of order \( n \). The B-spline of zero order is the box function defined as

\[
\beta_0(x) = \begin{cases} 
1, & \text{if } 0 \leq x \leq 1 \\
0, & \text{otherwise}
\end{cases}
\]

having the property that \( \beta_0^k(x) = \beta_0(x-k) \). The B-spline of order \( n \) is given recursively by the convolution with the B-spline of zero order

\[
\beta_n(x) = \int \beta_{n-1}(s) \beta_0(x-s) ds.
\]

The algorithm involves the following filters represented via their Z-transforms as:

Initialization filter

\[
B(z) = \sum_{k \in \mathbb{Z}} b(k) z^{-k} = \frac{z + 4 + z^{-1}}{6};
\]

Refinement filter

\[
H(z) = \sum_{k \in \mathbb{Z}} h(k) z^{-k} = \frac{z^2 + 4z + 6 + 4z^{-1} + z^{-2}}{8};
\]

1st derivative wavelet

\[
P(z) = \sum_{k \in \mathbb{Z}} p(k) z^{-k} = 1 - z,
\]

so that the wavelet function used in the algorithm is represented in terms of the B-spline basis of the third order as

\[
\phi(x) = \sum_{k \in \mathbb{Z}} p(k) \beta_3(x - k),
\]

where \( \beta_3(x)|_{x=k} = b_3(k) \)

The decomposition of a discrete signal is implemented by computing the consecutive convolutions of the signal with the above filters, involving into the summation only the terms with nonzero coefficients. To process the discrete signal \( f(k), k=0,\ldots,N-1 \) first we compute the sequence

\[
s_0(k) = \sum_{l=0}^{1} b(l) f(k + l).
\]

At \( j \)th level of decomposition the low pass of the input signal is given by scalar products of the signal and refinement filter

\[
s_j(k) = \sum_{l=-2}^{3} h(l) s_{j-1}(k + 2^{-j-1} l).
\]

The high pass is then calculated as the convolution of the low pass with the wavelet

\[
s_{j+1}(k) = \sum_{l=-2}^{3} \phi(l) s_j(k - 2^{-j-1} l).
\]
\[ r_j(k) = \sum_{l=0}^{1} p(l) s_j(k + 2^l I). \]

At the borders the periodic boundary conditions are used \( f(-1) = f(N-1), \ldots, f(N) = f(0) \).

We have chosen a wavelet function such that the high pass \( r_j(k) \) accumulates information about the first derivative of the signal. Our aim is to extract zeros of the first derivative from the high pass so that they can be used as GCPs in the spatial registration. The procedure of the feature extraction is described in detail in [13]. A brief illustration of the algorithm's behavior is presented in Figure 3. The extraction of the GCPs is done recursively, so that zeros of the derivative in a given iteration are extracted from the low pass output of a prior iteration. The original signals for temporal classes 3 and 8 are shown on panels 3a and 3d, respectively. Panel 3b displays an intermediate level of decomposition of the \textit{eve} gene expression pattern for temporal class 3. At this level the number of zeros of the first derivative (high pass) is greater than the characteristic number for temporal class 3 and so further recursions are required. Panel 3c shows the expression pattern of \textit{eve} gene for temporal class 3 at decomposition level that yields the required number of zeros of the first derivative in high pass and therefore is used for the extraction of GCPs. The consecutive decompositions of the \textit{eve} gene expression pattern for temporal class 8 are illustrated in panels 3(e) and 3(f). The maximal number of GCPs, which for this class is equal to 13, is achieved at the third level of signal decomposition (3f). At the intermediate level of decomposition (3e) several extra zeros of the first derivative are still present [8, 12].

![Fig. 3. An example of decomposition of \textit{eve} gene expression pattern. Panels a and d display one-dimensional expression patterns of \textit{eve} gene for temporal classes 3 and 8, respectively. The low pass (1) and the high pass (2) at intermediate and optimal levels of decomposition are presented on panels b and c for temporal class 3 and on panels e and f for temporal class 8.](image)

**Affine registration**

The registration of images is performed by resizing the two-dimensional expression patterns along the \( x \) axis by the affine transformation

\[ x' = x p_k + \Delta_k, \quad k = 1, \ldots, R \]

so that the total distance between \( x \) coordinates of all the extrema \( \{ X_{jy} \} \) and the average set of extrema
The purpose of spatial registration is to construct a map of all segmentation gene products from a set of embryos of the same age. Each protein in the map will have an "integrated pattern" defined by average concentrations of the protein over an age group. An ideal integrated pattern could be constructed in such a hypothetical situation where all embryos were built of a set of identically located nuclei with identical coordinates. In this case an average concentration of the protein could be computed in each nucleus independently. But in fact the number and location of nuclei constituting patterns are variable and therefore all the patterns should be brought to a unified discrete scale. We start by reconstructing the averaged nuclear structure of two-dimensional gene expression pattern which preserves the density of nuclei over an embryo image.

As a starting point in this work we use information about the coordinates of nuclear centroids. These coordinates were derived by averaging the $x$ and $y$ location of all the pixels over each nucleus of a binary nuclear mask. In our work the mask was generated from the brightest pixels collected among the three channels of the confocal microscope used to scan gene expression signal. At this point one meets with the problem that in completely nonexpressing areas the nuclear mask is highly inaccurate with incorrect number and shape of nuclei. Since there is no expression in these regions this inaccuracy has no effect on the measurement of levels of gene expression, however it could lead to serious errors if such embryos were to be used to infer the average nuclear distribution in an embryo image. To overcome this difficulty we calculate the average density of nuclei from data derived from embryos stained for the expression of even-skipped, bicoid and caudal. Even-skipped initially expresses protein in a single domain ranging from 20% to 80% of embryo length and restricts to seven narrow domains over a relatively short time interval [2, 5]. Caudal has a broad domain of expression extending through all but the most anterior part of the embryo and bicoid's expression domain is located in just anterior portion where caudal has the lowest expression. In that way the expression of bicoid, caudal, and even-skipped in total covers the whole embryo body and this makes it possible to construct accurate nuclear masks from embryos stained for these proteins. We have at our disposal 83 embryos with this combination of scanned genes.

As a first step the table of frequencies of nuclear centroids is constructed. For this purpose the image of each embryo is divided into rectangular cells whose sides are equal in length to a nuclear diameter (about 1% of an embryo length). Then all the nuclei from all embryos whose centroids fall into a given cell are minimized over all $R$ images. The number of extrema $M$ varies depending on the temporal class.

This approach was tested with the two cost functions: one based on the minimization of the Euclidean distance

$$S_E = \sum_{j=1}^{M} \left\{ X_{kj} \rho_k + \Delta_k - \overline{X}_j \right\}^2 \rightarrow \min, \quad k = 1, \ldots, R \quad (1)$$

and another defined by Mahalanobis distance

$$S_M = \sum_{j=1}^{M} \left\{ m_j X_{kj} \rho_k + \Delta_k - \overline{X}_j \right\}^2 \rightarrow \min, \quad k = 1, \ldots, R \quad (2)$$

where weights $m_j$ are the inverse variances of the $x$ coordinates of the extrema

$$m_j = \left( \frac{1}{R} \sum_{k=1}^{R} \left\{ X_{kj} - \overline{X}_j \right\}^2 \right)^{-1} \quad (3)$$

The purpose of spatial registration is to construct a map of all segmentation gene products from a set of embryos of the same age. Each protein in the map will have an "integrated pattern" defined by average concentrations of the protein over an age group. An ideal integrated pattern could be constructed in such a
counted and then divided by the number of embryos. We thus get a table of the average number of nuclei located in each cell. Next this table is converted into a table of average diameters of nuclei. If \( f_{i,j} \) is the frequency of nuclei in the \((i,j)\)th cell, the average diameter of a nucleus, the centroid of which belongs to this cell, is equal to

\[
d_y = 2 \sqrt{\frac{s}{\pi f_y}},
\]

where \( s \) is the area of a cell.

What we already have is a square grid for which we know average diameters of nuclei belonging to each cell. What we need now is to extend this knowledge to any point of an embryo body. In other words, it is necessary to interpolate estimates of diameters for any imaginary nucleus with the center at a given point. For this purpose the values \( d_{i,j} \) are approximated by the paraboloid \( D(x, y) = a(x^2 + y^2) + b(x + y) + cxy + d \) and \( D(x, y) \) is considered to be the estimated diameter of a nucleus with its center at \((x, y)\). The parameters of the paraboloid \( a, b, \) and \( c \) are obtained by the least squares method, minimizing

\[
\sum_{j} (D(x_i, y_j - d_{ij}))^2,
\]

where \((x_i, y_j)\) are the coordinates of the center of the \((i,j)\)th cell.

A nuclear structure of a pattern is modeled by constructing consecutively the one-dimensional chains of nuclei, beginning from the central A-P axis of an embryo and then moving up and down in dorsal and ventral directions. The nuclei are presented by small circles with the diameter equal to the value of the function \( D(x, y) \), computed in the center of the circle. This model will be referred to as a nuclear model of a pattern. The reconstructed nuclear model is shown in the Fig.5 together with an individual nuclear mask obtained by the segmentation procedure in Fig.4. Comparing these two images one can see that the nuclear model (Fig.5) preserves the right number, size and density of nuclei but the nuclear structure looks more regular and well-ordered than in the individual nuclear mask (Fig.4).

![Fig.4. The nuclear mask of an individual embryo obtained by the segmentation procedure.](image)

The two-dimensional integrated expression patterns of all the segmentation genes now can be mapped onto this model. The data assembly procedure is exactly the same as that used in construction of one-dimensional integrated patterns [12]. The registered two-dimensional patterns of all the embryos in the temporal class are averaged over the protein concentrations within each averaged nucleus. Each nucleus from the registered individual patterns is assigned to the least distant averaged nucleus and then the average protein concentrations of all the nuclei assigned to this nucleus are computed.
RESULTS

Embryos of approximately the same age stained for different combinations of proteins were subjected to spatial registration against the common \textit{eve} expression domain. An example of such registration is given in Figure 6.

![Fig. 6. An example of registration. Four two-dimensional patterns belonging to temporal class 8 are shown both before (upper panel) and after registration (lower panel). Seven expression domains of \textit{eve} gene expression are clearly seen on the lower panel. The lines correspond to an intensity level of approximately 50\% of maximum. Registration was done using a cost function based on the Euclidean distance.](image)

**Estimation of the accuracy of registration**

We performed the spatial registration of embryo images by extracting GCPs from the central 10\% strip of an embryo and applying the coordinate transformation to make GCPs in different images coincide as closely as possible. Previously we have shown that this method gives good accuracy of registration for the
data extracted from the central strip. To extend the one-dimensional registration method to two dimensions we apply the same coordinate transformation to the whole image.

Evidently to estimate the accuracy of the two-dimensional registration method it is necessary to measure a quality of registration of the data extracted from regions other than the central strip. To do this we have subdivided the area of an embryo below and above the central 10% strip into strips of 10% width. The data from each strip (located within the intervals of 20-30%, 30-40%, 40-50%, 60-70% and 70-80% of embryo width) were considered as one-dimensional patterns of *eve* gene expression. For each embryo 5 expression patterns were constructed and in each pattern the location of the extrema was estimated. The average *x* coordinates of all the extrema were computed over the temporal class together with their standard deviations. These standard deviations were used to estimate the accuracy of the spatial registration.

Tables 1–4 present the results of the accuracy of *eve* spatial registration in two dimensions for two temporal classes, namely temporal class 3, which is an early temporal class and temporal class 8, which is a late one. The locations of the extrema in each pattern, averaged over the time class, are given together with their standard deviations. Tables 1 and 3 present the accuracy of *eve* spatial registration for the data registered using the cost function based on the Euclidean distance, while results presented in tables 2 and 4 were obtained for the data registered using the cost function based on the Mahalanobis distance.

**Table 1**
The average values of all GCPs and their standard deviations after registration of embryos belonging to temporal class 8. The rows of the table correspond to the stripes and the columns correspond to the extrema of the patterns. The *x* coordinates of the extrema that were used as GCPs are marked out. Registration was done using cost function based on the Euclidean distance.

<table>
<thead>
<tr>
<th></th>
<th>max 1</th>
<th>min 1</th>
<th>max 2</th>
<th>min 2</th>
<th>max 3</th>
<th>min 3</th>
<th>max 4</th>
<th>min 4</th>
<th>max 5</th>
<th>min 5</th>
<th>max 6</th>
<th>min 6</th>
<th>max 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-30%</td>
<td>33.45</td>
<td>38.58</td>
<td>42.52</td>
<td>46.71</td>
<td>50.09</td>
<td>53.34</td>
<td>56.57</td>
<td>60.07</td>
<td>63.28</td>
<td>66.59</td>
<td>69.78</td>
<td>73.83</td>
<td>78.17</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.71</td>
<td>0.70</td>
<td>0.60</td>
<td>0.67</td>
<td>0.51</td>
<td>0.56</td>
<td>0.52</td>
<td>0.67</td>
<td>0.63</td>
<td>0.70</td>
<td>0.76</td>
<td>1.00</td>
</tr>
<tr>
<td>30-40%</td>
<td>33.31</td>
<td>38.30</td>
<td>42.15</td>
<td>46.52</td>
<td>49.75</td>
<td>53.22</td>
<td>56.36</td>
<td>59.96</td>
<td>62.98</td>
<td>66.48</td>
<td>69.70</td>
<td>73.92</td>
<td>78.04</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.54</td>
<td>0.47</td>
<td>0.45</td>
<td>0.48</td>
<td>0.40</td>
<td>0.37</td>
<td>0.44</td>
<td>0.58</td>
<td>0.56</td>
<td>0.59</td>
<td>0.60</td>
<td>0.80</td>
</tr>
<tr>
<td>40-50%</td>
<td>32.72</td>
<td>37.74</td>
<td>41.51</td>
<td>46.11</td>
<td>49.47</td>
<td>53.14</td>
<td>56.33</td>
<td>59.91</td>
<td>62.97</td>
<td>66.75</td>
<td>70.02</td>
<td>74.60</td>
<td>78.77</td>
</tr>
<tr>
<td></td>
<td>0.53</td>
<td>0.43</td>
<td>0.38</td>
<td>0.37</td>
<td>0.32</td>
<td>0.33</td>
<td>0.36</td>
<td>0.30</td>
<td>0.42</td>
<td>0.35</td>
<td>0.46</td>
<td>0.38</td>
<td>0.46</td>
</tr>
<tr>
<td>50-60%</td>
<td>31.64</td>
<td>36.83</td>
<td>40.69</td>
<td>45.62</td>
<td>49.04</td>
<td>53.06</td>
<td>56.42</td>
<td>60.22</td>
<td>64.37</td>
<td>67.54</td>
<td>71.01</td>
<td>76.00</td>
<td>80.28</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.32</td>
<td>0.26</td>
<td>0.31</td>
<td>0.22</td>
<td>0.27</td>
<td>0.30</td>
<td>0.29</td>
<td>0.38</td>
<td>0.30</td>
<td>0.30</td>
<td>0.33</td>
<td>0.39</td>
</tr>
<tr>
<td>60-70%</td>
<td>30.02</td>
<td>35.60</td>
<td>39.73</td>
<td>44.86</td>
<td>48.50</td>
<td>52.88</td>
<td>56.41</td>
<td>60.86</td>
<td>64.24</td>
<td>68.68</td>
<td>72.45</td>
<td>77.83</td>
<td>82.82</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.43</td>
<td>0.41</td>
<td>0.54</td>
<td>0.37</td>
<td>0.29</td>
<td>0.37</td>
<td>0.41</td>
<td>0.51</td>
<td>0.46</td>
<td>0.47</td>
<td>0.66</td>
<td>0.98</td>
</tr>
<tr>
<td>70-80%</td>
<td>28.78</td>
<td>34.00</td>
<td>38.41</td>
<td>43.91</td>
<td>47.86</td>
<td>52.54</td>
<td>56.28</td>
<td>61.04</td>
<td>64.76</td>
<td>69.68</td>
<td>73.73</td>
<td>78.85</td>
<td>83.78</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>0.55</td>
<td>0.64</td>
<td>0.69</td>
<td>0.57</td>
<td>0.54</td>
<td>1.07</td>
<td>1.97</td>
<td>2.38</td>
<td>2.33</td>
<td>2.49</td>
<td>3.74</td>
<td>4.42</td>
</tr>
</tbody>
</table>

For temporal class 8 the standard deviations of the *x* coordinates of the extrema are in the range of 0.22-0.47% in the central 10% strip which was used for registration. In the other strips their values are in the range of 0.29 - 4.42% and often less than 1% (Tables 1 and 2). A single nucleus is about 1% egg
length in diameter in the center of an embryo and 1.5 times larger at the borders, so this represents a high level of accuracy.

In general registration error increases from the center of the embryo to its borders. This is not surprising, since the GCPs used are in the central D-V strip. Moreover, close to the border the image loses focus due to curvature and we believe that this property could account for increase of the registration error from the center to the borders of an embryo.

Table 2
The average values of all GCPs and their standard deviations after registration of embryos belonging to temporal class 8. The rows of the table correspond to the stripes and the columns correspond to the extrema of the patterns. The x coordinates of the extrema that were used as GCPs are marked out. Registration was done using cost function based on the Mahalanobis distance.

<table>
<thead>
<tr>
<th></th>
<th>max 1</th>
<th>min 1</th>
<th>max 2</th>
<th>min 2</th>
<th>max 3</th>
<th>min 3</th>
<th>max 4</th>
<th>min 4</th>
<th>max 5</th>
<th>min 5</th>
<th>max 6</th>
<th>min 6</th>
<th>max 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-30%</td>
<td>33.45</td>
<td>38.59</td>
<td>42.52</td>
<td>46.71</td>
<td>50.09</td>
<td>53.34</td>
<td>56.57</td>
<td>60.07</td>
<td>63.28</td>
<td>66.59</td>
<td>69.78</td>
<td>73.83</td>
<td>78.17</td>
</tr>
<tr>
<td></td>
<td>1.01</td>
<td>0.71</td>
<td>0.69</td>
<td>0.59</td>
<td>0.67</td>
<td>0.50</td>
<td>0.56</td>
<td>0.52</td>
<td>0.67</td>
<td>0.63</td>
<td>0.70</td>
<td>0.76</td>
<td>1.00</td>
</tr>
<tr>
<td>30-40%</td>
<td>33.31</td>
<td>38.30</td>
<td>42.15</td>
<td>46.52</td>
<td>49.75</td>
<td>53.23</td>
<td>56.36</td>
<td>59.96</td>
<td>62.98</td>
<td>66.48</td>
<td>69.79</td>
<td>73.92</td>
<td>78.04</td>
</tr>
<tr>
<td></td>
<td>0.76</td>
<td>0.54</td>
<td>0.47</td>
<td>0.44</td>
<td>0.48</td>
<td>0.39</td>
<td>0.37</td>
<td>0.44</td>
<td>0.58</td>
<td>0.56</td>
<td>0.60</td>
<td>0.60</td>
<td>0.80</td>
</tr>
<tr>
<td>40-50%</td>
<td>32.72</td>
<td>37.74</td>
<td>41.51</td>
<td>46.11</td>
<td>49.47</td>
<td>53.14</td>
<td>56.33</td>
<td>59.91</td>
<td>62.97</td>
<td>66.75</td>
<td>70.02</td>
<td>74.60</td>
<td>78.77</td>
</tr>
<tr>
<td></td>
<td>0.55</td>
<td>0.44</td>
<td>0.38</td>
<td>0.36</td>
<td>0.31</td>
<td>0.32</td>
<td>0.35</td>
<td>0.29</td>
<td>0.42</td>
<td>0.35</td>
<td>0.46</td>
<td>0.37</td>
<td>0.46</td>
</tr>
<tr>
<td>50-60%</td>
<td>31.64</td>
<td>36.83</td>
<td>40.69</td>
<td>45.62</td>
<td>49.04</td>
<td>53.06</td>
<td>56.42</td>
<td>60.22</td>
<td>63.47</td>
<td>67.54</td>
<td>71.01</td>
<td>76.00</td>
<td>80.28</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>0.34</td>
<td>0.26</td>
<td>0.29</td>
<td>0.22</td>
<td>0.26</td>
<td>0.30</td>
<td>0.28</td>
<td>0.38</td>
<td>0.30</td>
<td>0.30</td>
<td>0.32</td>
<td>0.39</td>
</tr>
<tr>
<td>60-70%</td>
<td>30.02</td>
<td>35.60</td>
<td>39.73</td>
<td>44.86</td>
<td>48.50</td>
<td>52.88</td>
<td>56.41</td>
<td>60.86</td>
<td>64.24</td>
<td>68.68</td>
<td>72.45</td>
<td>77.83</td>
<td>82.82</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>0.44</td>
<td>0.42</td>
<td>0.53</td>
<td>0.36</td>
<td>0.28</td>
<td>0.36</td>
<td>0.40</td>
<td>0.51</td>
<td>0.46</td>
<td>0.47</td>
<td>0.66</td>
<td>0.97</td>
</tr>
<tr>
<td>70-80%</td>
<td>28.78</td>
<td>34.00</td>
<td>38.41</td>
<td>43.91</td>
<td>47.86</td>
<td>52.54</td>
<td>56.28</td>
<td>61.04</td>
<td>64.76</td>
<td>69.69</td>
<td>73.74</td>
<td>78.85</td>
<td>83.78</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.55</td>
<td>0.64</td>
<td>0.68</td>
<td>0.56</td>
<td>0.54</td>
<td>1.06</td>
<td>1.97</td>
<td>2.38</td>
<td>2.33</td>
<td>2.48</td>
<td>3.74</td>
<td>4.42</td>
</tr>
</tbody>
</table>

For temporal class 3 the standard deviations of the x coordinates of the extrema in the central 10% strip, which was used for registration, are larger than for temporal class 8, ranging from 0.51 – 2.01%. In the other strips their values are in the range of 1.06 – 5.98% (Tables 3 and 4). This difference from the behavior of temporal class 8 is a consequence of the imprecise shape of the early patterns.

For temporal class 3, registration of data using the cost function based on the Mahalanobis distance is more accurate in comparison with registration based on the Euclidean distance. The idea of the Mahalanobis distance is to assign different weights to the different terms in calculating the distance (2). Each term in the sum (2) corresponds to an extremum used in registration and by introducing weights the significance of the corresponding extremum in the registration procedure is either increased or decreased. In early patterns some of the stripes are not yet reliably formed and hence the location of their concentration extrema have major variance, while the small variance of an extremum on the contrary testifies that its location has already well stabilized in the expression pattern. It is clear that precise matching of the extrema with unstable locations makes not much sense. As the coefficients $m_j$ given by (3) represent the in-
verse variances of $x$ coordinates of the extrema they are greater for the extrema with “reliable” location, which makes them more significant in registration.

### Table 3

The average values of all GCPs and their standard deviations after registration of embryos belonging to temporal class 3. The rows of the table correspond to the stripes and the columns correspond to the extrema of the patterns. The $x$ coordinates of the extrema that were used as GCPs are marked out. Registration was done using cost function based on the Eucledian distance.

<table>
<thead>
<tr>
<th></th>
<th>max 1</th>
<th>min 1</th>
<th>max 2</th>
<th>min 2</th>
<th>max 3</th>
<th>min 3</th>
<th>max 4</th>
<th>min 4</th>
<th>max 5</th>
<th>min 5</th>
<th>max 6</th>
<th>min 6</th>
<th>max 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-30%</td>
<td>33.53</td>
<td>40.22</td>
<td>43.64</td>
<td>49.73</td>
<td>51.74</td>
<td>55.24</td>
<td>59.97</td>
<td>63.3</td>
<td>67.86</td>
<td>69.64</td>
<td>71.68</td>
<td>78.29</td>
<td>83.91</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
<td>1.4</td>
<td>1.11</td>
<td>1.94</td>
<td>2.56</td>
<td>2.16</td>
<td>1.59</td>
<td>1.8</td>
<td>2.1</td>
<td>2.18</td>
<td>2.56</td>
<td>2.16</td>
<td>2.59</td>
</tr>
<tr>
<td>30-40%</td>
<td>32.83</td>
<td>38.96</td>
<td>42.8</td>
<td>47.87</td>
<td>50.44</td>
<td>54.18</td>
<td>58.88</td>
<td>62.42</td>
<td>67.23</td>
<td>69.39</td>
<td>71.86</td>
<td>77.48</td>
<td>82.91</td>
</tr>
<tr>
<td></td>
<td>3.84</td>
<td>4.14</td>
<td>3.43</td>
<td>3.7</td>
<td>2.85</td>
<td>2.73</td>
<td>2.67</td>
<td>2.8</td>
<td>2.94</td>
<td>2.41</td>
<td>2.37</td>
<td>2.94</td>
<td>2.93</td>
</tr>
<tr>
<td>40-50%</td>
<td>32.81</td>
<td>39.05</td>
<td>42.97</td>
<td>48.08</td>
<td>50.66</td>
<td>54.42</td>
<td>59.16</td>
<td>62.62</td>
<td>67.29</td>
<td>69.26</td>
<td>72.09</td>
<td>78.21</td>
<td>83.7</td>
</tr>
<tr>
<td></td>
<td>2.62</td>
<td>2.66</td>
<td>2.13</td>
<td>2.58</td>
<td>2.3</td>
<td>2.24</td>
<td>2.23</td>
<td>2.37</td>
<td>2.94</td>
<td>2.89</td>
<td>2.57</td>
<td>2.91</td>
<td>3.46</td>
</tr>
<tr>
<td>50-60%</td>
<td><strong>32.49</strong></td>
<td>38.67</td>
<td>42.79</td>
<td>48.6</td>
<td>51.27</td>
<td><strong>55.31</strong></td>
<td>59.94</td>
<td><strong>63.39</strong></td>
<td>67.54</td>
<td>69.75</td>
<td><strong>72.86</strong></td>
<td>79.16</td>
<td><strong>84.58</strong></td>
</tr>
<tr>
<td></td>
<td>0.79</td>
<td>0.86</td>
<td>0.71</td>
<td>0.97</td>
<td>1.11</td>
<td><strong>1.37</strong></td>
<td>1.27</td>
<td><strong>1.36</strong></td>
<td>1.65</td>
<td>1.51</td>
<td><strong>0.85</strong></td>
<td><strong>1.22</strong></td>
<td><strong>1.31</strong></td>
</tr>
<tr>
<td>60-70%</td>
<td>31.27</td>
<td>37.41</td>
<td>41.75</td>
<td>47.85</td>
<td>50.78</td>
<td>55.29</td>
<td>59.88</td>
<td>63.65</td>
<td>68.14</td>
<td>70.93</td>
<td>73.52</td>
<td>80.13</td>
<td>85.03</td>
</tr>
<tr>
<td></td>
<td>1.66</td>
<td>1.34</td>
<td>1.15</td>
<td>1.09</td>
<td>1.15</td>
<td>1.25</td>
<td>1.47</td>
<td>1.74</td>
<td>2.61</td>
<td>3.48</td>
<td>3.19</td>
<td>4.68</td>
<td>5.76</td>
</tr>
<tr>
<td>70-80%</td>
<td>30.11</td>
<td>36.32</td>
<td>41.14</td>
<td>46.85</td>
<td>49.99</td>
<td>54.23</td>
<td>60.16</td>
<td>64.2</td>
<td>68.28</td>
<td>71.65</td>
<td>74</td>
<td>77.45</td>
<td>81.1</td>
</tr>
<tr>
<td></td>
<td>2.24</td>
<td>1.85</td>
<td>2.01</td>
<td>1.83</td>
<td>1.86</td>
<td>2.68</td>
<td>2.5</td>
<td>1.99</td>
<td>1.89</td>
<td>3.06</td>
<td>3.32</td>
<td>5.38</td>
<td>5.93</td>
</tr>
</tbody>
</table>

The two-dimensional map of segmentation gene expression

In creating the integrated patterns of segmentation gene expression we try to preserve the density and size of nuclei in an embryo image. Our approach is based on the fact that nuclei are not distributed uniformly over the embryo body. The nuclei are located most densely in the center with the density decreasing to the edges of the embryo. At the same time a nuclear diameter changes inversely by increasing towards the ends.

We use the knowledge about distribution of nuclei over an embryo image to construct the nuclear model of a pattern (see Fig.5). The number of nuclei in this model is equal to the mean value of nuclei in the embryo images.

To ensure reliable statistical properties, the integrated patterns were created only for those genes which have been stained in more than 9 embryos. These genes are $bcd$, $Kr$, $gt$, $kni$, $eve$, $run$, $h$, $ftz$, $odd$ and $cad$. Figure 7 presents the integrated patterns of $eve$ gene expression imposed on the nuclear model of a pattern. One of these patterns has been attributed by visual inspection to the temporal class 3, while the other one belongs to late temporal class 8. The integrated patterns are in good agreement with patterns from the individual embryos of the same class, as well as with the data about the evolution of $eve$ gene expression.
Table 4

The average values of all GCPs and their standard deviations after registration of embryos belonging to temporal class 3. The rows of the table correspond to the stripes and the columns correspond to the extrema of the patterns. The x coordinates of the extrema that were used as GCPs are marked out. Registration was done using cost function based on the Mahalanobis distance.

<table>
<thead>
<tr>
<th></th>
<th>max 1</th>
<th>min 1</th>
<th>max 2</th>
<th>min 2</th>
<th>max 3</th>
<th>min 3</th>
<th>max 4</th>
<th>min 4</th>
<th>max 5</th>
<th>min 5</th>
<th>max 6</th>
<th>min 6</th>
<th>max 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-30%</td>
<td>33.51</td>
<td>40.2</td>
<td>43.63</td>
<td>49.71</td>
<td>51.71</td>
<td>55.22</td>
<td>59.96</td>
<td>63.28</td>
<td>67.85</td>
<td>69.59</td>
<td>71.62</td>
<td>78.28</td>
<td>83.9</td>
</tr>
<tr>
<td></td>
<td>2.26</td>
<td>1.37</td>
<td>1.08</td>
<td>1.95</td>
<td>2.57</td>
<td>2.14</td>
<td>1.55</td>
<td>1.66</td>
<td>2</td>
<td>2.15</td>
<td>2.5</td>
<td>1.85</td>
<td>2.19</td>
</tr>
<tr>
<td>30-40%</td>
<td>32.8</td>
<td>38.94</td>
<td>42.78</td>
<td>47.85</td>
<td>50.41</td>
<td>54.16</td>
<td>58.85</td>
<td>62.39</td>
<td>67.22</td>
<td>69.33</td>
<td>71.77</td>
<td>77.47</td>
<td>82.95</td>
</tr>
<tr>
<td></td>
<td>3.88</td>
<td>4.19</td>
<td>3.47</td>
<td>3.73</td>
<td>2.86</td>
<td>2.73</td>
<td>2.63</td>
<td>2.71</td>
<td>2.86</td>
<td>2.36</td>
<td>1.96</td>
<td>2.26</td>
<td>2.83</td>
</tr>
<tr>
<td>40-50%</td>
<td>32.8</td>
<td>39.04</td>
<td>42.96</td>
<td>48.07</td>
<td>50.65</td>
<td>54.4</td>
<td>59.14</td>
<td>62.61</td>
<td>67.29</td>
<td>69.17</td>
<td>72.01</td>
<td>78.2</td>
<td>83.69</td>
</tr>
<tr>
<td></td>
<td>2.56</td>
<td>2.61</td>
<td>2.09</td>
<td>2.55</td>
<td>2.29</td>
<td>2.21</td>
<td>2.14</td>
<td>2.27</td>
<td>2.84</td>
<td>2.67</td>
<td>2.27</td>
<td>2.64</td>
<td>3.17</td>
</tr>
<tr>
<td>50-60%</td>
<td>32.47</td>
<td>38.65</td>
<td>42.77</td>
<td>48.59</td>
<td>51.26</td>
<td>55.28</td>
<td>59.92</td>
<td>63.36</td>
<td>67.54</td>
<td>69.73</td>
<td>72.84</td>
<td>79.18</td>
<td>84.62</td>
</tr>
<tr>
<td></td>
<td>0.57</td>
<td>0.75</td>
<td>0.65</td>
<td>0.98</td>
<td>1.11</td>
<td>1.35</td>
<td>1.1</td>
<td>1.06</td>
<td>1.6</td>
<td>1.5</td>
<td>0.74</td>
<td>1.61</td>
<td>2.01</td>
</tr>
<tr>
<td>60-70%</td>
<td>31.26</td>
<td>37.4</td>
<td>41.73</td>
<td>47.84</td>
<td>50.76</td>
<td>55.28</td>
<td>59.86</td>
<td>63.63</td>
<td>68.14</td>
<td>70.94</td>
<td>73.54</td>
<td>80.15</td>
<td>85.02</td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>1.22</td>
<td>1.11</td>
<td>1.06</td>
<td>1.12</td>
<td>1.2</td>
<td>1.41</td>
<td>1.58</td>
<td>2.53</td>
<td>3.46</td>
<td>3.15</td>
<td>4.53</td>
<td>5.66</td>
</tr>
<tr>
<td>70-80%</td>
<td>30.1</td>
<td>36.31</td>
<td>41.12</td>
<td>46.86</td>
<td>50.12</td>
<td>54.23</td>
<td>60.16</td>
<td>64.21</td>
<td>68.34</td>
<td>71.69</td>
<td>74.01</td>
<td>77.44</td>
<td>81.14</td>
</tr>
<tr>
<td></td>
<td>2.12</td>
<td>1.76</td>
<td>1.97</td>
<td>1.79</td>
<td>1.83</td>
<td>2.73</td>
<td>2.44</td>
<td>1.97</td>
<td>1.98</td>
<td>3.09</td>
<td>3.23</td>
<td>5.47</td>
<td>5.98</td>
</tr>
</tbody>
</table>

Fig. 7. The integral patterns of *eve* gene expression. One of these patterns has been attributed by visual inspection to temporal class 3 (upper panel), while the other one belongs to late temporal class 8 (lower panel).
known from the literature. At the stage of development corresponding to time class 3 this gene has a very broad domain of expression, while at late stages of cycle 14A its pattern is characterized by seven well defined stripes.

Some visual distortion in the shape of stripe borders, especially for the late pattern, is caused by the regular and discrete structure of the nuclear model. Nevertheless, the overall pattern of expression is well preserved.

The two-dimensional integrated patterns of different segmentation genes from one temporal class can be combined so as to make a map of segmentation gene expression at a given time point. It is not possible to visually perceive the expression domains of different genes on one two-dimensional map obtained by combining the integrated patterns of all the segmentation genes. We believe that combinations of three genes are optimal for this purpose. Figure 9 presents the map of the expression of \textit{gt}, \textit{Kr}, and \textit{eve} for temporal class 8. To make sure that the two-dimensional pattern preserves the main features of individual patterns we show an individual pattern with the same combination of stained proteins in Figure 8. The images are not identical due to individual variation of gene expression patterns among embryos but they look very similar and it is clearly seen that the model correctly reproduces the shape and size of expression domains. In Figure 10 the map of the expression of \textit{Kr}, \textit{gt}, \textit{kni} for temporal class 8 is presented. It is noteworthy to say that our dataset does not contain embryos stained for these three proteins simultaneously. By mapping integrated patterns of different genes on a nuclear model of a pattern it is possible to display and visualize the expression domains of segmentation genes in any possible combinations.

![Fig. 8. An individual embryo stained for the protein products of eve (red), Kruppel (green) and giant (blue) for temporal class 8.](image)

![Fig. 9. A map of the expression of eve (red), Kruppel (green) and giant (blue) for temporal class 8.](image)
CONCLUSIONS

We have developed a method for the two-dimensional registration of images of segmentation gene expression patterns. The image registration is performed against the extremal features of the pattern extracted using the FRDWT. The method is modified so that the level of decomposition optimal for the extraction of the GCPs is detected automatically. The algorithm provides high resolution and noise reduction of the decomposed signal.

We applied this method to the two-dimensional registration of segmentation gene expression patterns in 736 embryos minimizing two different cost functions, one defined as the Euclidean distance and the other one as the Mahalanobis distance between the sets of GCPs in embryo images. The registration method is highly accurate. The standard deviations of the $x$ coordinates of the extrema are reasonably small over the whole embryo image compared to the average size of a nucleus. For late temporal classes both cost functions give similar results. This similarity can be explained by the fact that all $eve$ gene expression domains are formed by this time. For early temporal classes the standard deviations of $x$ coordinates of the extrema are considerably larger than for late classes. Comparing the registration results we may conclude that using the Mahalanobis distance provides better accuracy of registration than the Euclidean one. The Mahalanobis distance is preferable in the case of early patterns because this method takes into account the dispersion of the GCPs before registration and at the early stage of development different GCPs may have different variances.

We have extended our method for construction of the integrated dataset to the two-dimensional case. The procedure for data assembly provides reasonable preservation of the nuclear structure of a confocal image of an embryo. The main imperfection of the nuclear model is that it is too regular, in some places approaching a rectilinear grid and in others hexagonal packing. The actual distribution of nuclei is approximately hexagonal, but some nuclei have only 5 neighbors and the nuclear spacing is somewhat irregular. This might be better approximated by adding a stochastic component, but doing this correcting would require an exhaustive statistical analysis of nuclear spacing. We believe the model used here to be about as good as can be achieved without adding a stochastic component.

In this work we have shown that nuclei are located most densely in the center with the density decreasing to the edges of an embryo image. We think it likely that this effect results from a combination of mechanical distortion of an embryo by a cover glass combined with some true variations in density. Separating these effects will require careful imaging and segmentation of embryos in three dimensions, which will be considerably more data intensive than the work reported here.
REFERENCES