

Automatic Recognition and Annotation of Gene Expression Patterns of Fly Embryos

Jie Zhou¹ and Hanchuan Peng²*

¹Department of Computer Science, Northern Illinois University, DeKalb IL 60115

²Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn VA 20147

[†] Equal contribution.

ABSTRACT

Motivation: Gene expression patterns obtained by *in situ* mRNA hybridization provide important information about different genes during *Drosophila* embryogenesis. So far, annotations of these images are done by manually assigning a subset of anatomy ontology terms to an image. This time-consuming process depends heavily on the consistency of experts.

Results: We develop a system to automatically annotate a fruitfly's embryonic tissue in which a gene has expression. We formulate the task as an image pattern recognition problem. For a new fly embryo image, our system answers two questions: 1) Which stage-range does an image belong to? 2) Which annotations should be assigned to an image? We propose to identify the wavelet-embryo features by multi-resolution 2D wavelet discrete transform followed by min-Redundancy max-Relevance feature selection, which yields optimal distinguishing features for an annotation. We then construct a series of parallel bi-class predictors to solve the multi-objective annotation problem since each image may correspond to multiple annotations.

Supplementary Information: The complete annotation prediction results are available at: <http://www.cs.niu.edu/~jzhou/papers/fruitfly> and http://research.janelia.org/peng/proj/fly_embryo_annotation/. The data sets used in experiments will be available upon request to the correspondence author.

Contact: jzhou@cs.niu.edu and pengh@janelia.hhmi.org

1. INTRODUCTION

Analysis of *in situ* gene expression patterns sheds new light on understanding the complicated relationship of genes. Recent work on automating this process has been reported for several model systems including mouse (e.g. Carson, 2005), fruitfly (e.g. Peng and Myers, 2004), etc. For fruitfly (*Drosophila melanogaster*), gene expression pattern images during embryogenesis obtained by *in situ* mRNA hybridization provide important spatial-temporal functional information. These images contain body structures that emerge during certain developmental stages. For investigation of genetic regulatory elements, automatic retrieval and clustering these images have been found very useful (Peng and Myers 2004; Peng et al. 2006; Pan et al, 2006.).

Annotations of these structures are important for the study of *Drosophila* embryogenesis. So far, the annotations of these fly embryo images are done by manually assigning anatomy ontology terms to the images (Tomancak et al. 2002). This time-consuming process depends heavily on the consistency of experts. With the availabil-

ity of a large number of pattern images in databases such as the Berkeley *Drosophila* Genome Project (BDGP), an automatic and systematic annotation approach that can increase the efficiency and consistency of the analysis becomes highly desirable.

We develop a system to automatically annotate fruitfly's embryonic tissue in which a gene has expression. We formulate the task as an image pattern recognition problem. For a new fly embryo image, our system answers two questions: 1) Which stage-range does an image belong to? 2) Which annotations should be assigned to an image?

One key issue is that the embryonic tissues in which a target gene is expressed are often unknown beforehand; therefore an arbitrary input image of gene expression pattern may have many different ontological annotations associated. From the viewpoint of pattern recognition, this means that each input sample may correspond to multiple class labels. Given the set of K class labels $\Phi = \{c_1, \dots, c_K\}$, automatic annotation of a fly embryo gene expression image is *multi-objective*: an input images x_i corresponds to a target set $a_i \subset \Phi$. This problem demands a special design for feature extraction and classification. There are three major challenges:

- For the multi-objective problem of embryonic image annotation, features corresponding to an unknown number of tissue structures co-exist in the same image. As in the image data each image pixel is often regarded as a dimension, the total dimensionality equals the number of image pixels, which is often very large. How to effectively extract features is critical for identifying the discriminating features correlated with different structures.
- The sample distribution of our data is heavily skewed. Typically some ontology terms are commonly used to annotate our image samples while many other terms are associated with a relatively small number of images. The actual percentages for different annotations vary greatly from less than 1% to about 20%.
- The quality and morphology of images vary greatly. Variation in embryo morphology, expression pattern staining and image orientation increase the difficulty of effective preprocessing and registration of the images. In addition, some inconsistent or missing annotations make it a nontrivial effort to generate an objective evaluation dataset.

This paper proposes the wavelet-embryo feature extraction and selection method for fly embryonic images so that various gene-expressed structures within the same image can be effectively decomposed and automatically recognized. We use the wavelet de-

*To whom correspondence should be addressed (pengh@janelia.hhmi.org).

composition scheme to project the original pixel-based embryonic images to a new feature domain in order to reveal features at different resolutions and frequency bands. Then we apply the min-Redundancy Max-Relevance (mRMR) feature selection algorithm to identify the optimal distinguishing features for a specific annotation. With this scheme, we then construct a series of parallel bi-class predictors to solve the multi-objective annotation problem.

We have tested our approach using *in situ* mRNA expression patterns of 463 fly genes. We compared our method against several others and found that the combination of wavelet-embryo-features and mRMR feature selection yields promising features for recognition of gene expression patterns, despite various challenges in the input image data. We have generated comprehensive prediction tables over the entire course of the embryogenesis for these 463 fly genes. Comparison of our results with the expert manual annotations indicates that our paradigm is successful. Our predictions are available on the authors' websites.

2. METHODS

Our method has three major portions, namely wavelet-embryo feature decomposition, feature selection, and recognition of gene expression patterns.

2.1 Wavelet-Embryo Features

With a large number of features coexisting in a single embryonic gene expression image, the task of feature extraction is to produce a representation that can best characterize this embryo image.

One option to extract features from digital images is to use pixels or combination of pixels as candidate features. With the high dimensionality of embryonic images, pixel combinations have been considered due to its smaller computational load and information redundancy (Peng and Myers 2004; Peng et al. 2006; Pan et al, 2006). For example, eigen-embryo analysis uses principal component analysis (PCA) to conduct linear combinations of pixel intensity values and extract the most prominent image features (Peng et al. 2006). While eigen-embryo features in conjunction with graph partition methods have led to very interesting results in clustering gene expression patterns in an unsupervised way, the method seems to be less appropriate for the purpose of automatic annotation where features with different levels of prominences need to be revealed to characterize various structures coexisting in the same image.

In this paper, we propose using the multi-resolution wavelet representation for embryo images. Wavelet representation lies between the spatial and Fourier domains in the sense that wavelets are localized in both space and frequency whereas the standard Fourier transform is only localized in frequency (Daubechies 1992; Mallat 1989, Mallat 1999). Multi-resolution representations based on wavelet decomposition are effective for identifying and analyzing local and multi-scale features from signals or images and have been used in other pattern recognition tasks such as face recognition and image retrieval (Chien and Wu 2002; Manjunath and Ma 1996). With wavelet-decomposition, a fly embryo image is projected to a feature space where information is decomposed into various frequency bands and different levels of resolutions so that features of different structures may be effectively separated.

We use two-dimensional discrete wavelet transform (DWT) to extract multi-resolution image features. 2D DWT decomposes an

image using orthonormal wavelet basis functions. Let the set of wavelet basis functions be $\{\Psi_{k,n}(r_1, r_2), k, n \in \mathbb{Z}\}$, where n is the translation factor (when n increases, the wavelet shifts right), and k is the dilation factor, which denotes a particular resolution level (when k gets smaller, the resolution increases). Let $x(r_1, r_2)$ represent the intensity values of embryo image pixels indexed by the location vector (r_1, r_2) . We can obtain the wavelet transform coefficient $d_{k,n}$ as below (Mallat 1999; Resnikoff and Wells 1998):

$$d_{k,n} = \iint \Psi_{k,n}(r_1, r_2) x(r_1, r_2) dr_1 dr_2$$

In order to construct the two-dimensional orthonormal wavelet basis function $\Psi_{k,n}(r_1, r_2)$ for multi-resolution analysis, we start from the following one-dimensional (1D) counterpart.

In multi-resolution analysis theory, the wavelet function $\psi(r)$ has a companion, the scaling function $\varphi(r)$, which approximates the signal/image at a specific level of resolution. The sets of scaling and wavelet functions are built by translation and dilation:

$$\begin{cases} \varphi_{k,n}(r) = 2^{-k/2} \varphi(2^{-k}r - n) \\ \psi_{k,n}(r) = 2^{-k/2} \psi(2^{-k}r - n) \end{cases} \quad k, n \in \mathbb{Z}$$

Intuitively, the scaling function indicates the trend of the signal, while the wavelet functions code the details of the images that can be added to reconstruct the original signal. Mallat (1989) linked orthogonal wavelets and scaling functions to quadrature mirror filters in signal processing theory: scaling function $\varphi(r)$ produces a smoothed signal as a low-pass filter; wavelet functions $\psi(r)$ catch the high frequency details as high-pass filters. Correspondingly, the 2D orthonormal wavelet basis function $\Psi(r_1, r_2)$ is factorized using the scaling function $\varphi(r)$ and mother wavelet function $\psi(r)$ as below :

$$\begin{cases} \Psi^1(r_1, r_2) = \varphi(r_1)\varphi(r_2) \\ \Psi^2(r_1, r_2) = \varphi(r_1)\psi(r_2) \\ \Psi^3(r_1, r_2) = \psi(r_1)\varphi(r_2) \\ \Psi^4(r_1, r_2) = \psi(r_1)\psi(r_2) \end{cases} \quad (1)$$

Schematically, each base function in Eq. (1) represents a two-step transform process: 1) perform a 1D wavelet transform on each row (indexed by r_1) of the embryonic image; and then 2) perform the transform on each column (indexed by r_2) of the result of step one. In our experiments, $\varphi(r)$ and $\psi(r)$ of Daubechies-1 wavelet and scaling functions (Daubechies 1992) are used.

With the connection between wavelet base functions and filters established by Mallat (1989), we can also view each subspace in Eq. (1) as a two-dimensional filter that transforms the original image into one of the four components: the low-low (LL), low-high (LH), high-low (HL) and high-high (HH) parts of the transform. In other words, the wavelet decomposition yields 4 sub-images at each resolution level. LL filtering is a "smoothing" of the original image. The other three at each resolution level are the detailed sub-images. Then at level 2, we apply the same analysis to the LL subimage, where the wavelets now reveal coarser-grained details of the image, and thus achieve the *multi-resolution* feature decomposition.

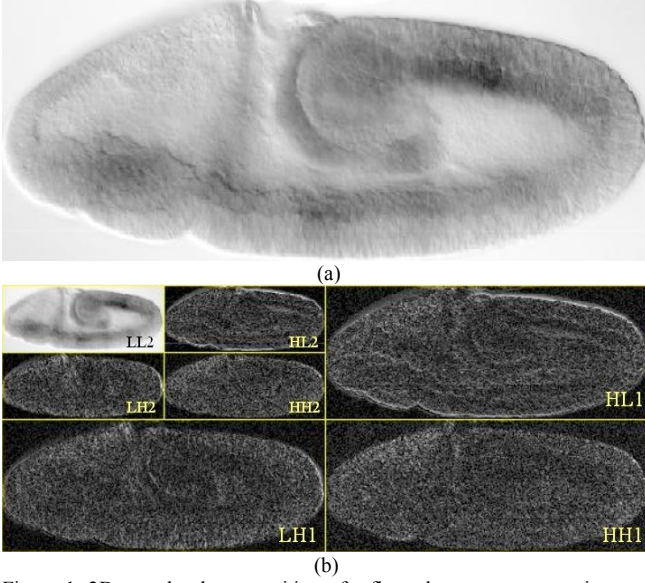


Figure 1. 2D wavelet decomposition of a fly embryo gene expression pattern image. (a) The original *in situ* mRNA expression pattern image of gene CG3400 in embryonic stage 7-8. (b) The wavelet-embryo features obtained by applying 2-Level 2D wavelet decomposition.

Figure 1 shows an example of the wavelet-embryo decomposition. The LL2 quadrant in the upper left corner of (b) is a smoothing of the original image in (a). The other three parts, LH, HL, and HH, are detail images at two different resolution levels. The set of coefficients $\{d_{k,n}\}$ of all sub-images at all resolution levels are used as features to characterize an gene expression pattern. In our experiment, $k = 0, 1$ (correspond to levels 1 and 2 in Figure 1); $n = 1, \dots, |r_1| \cdot |r_2|$, where $|r_1|$ and $|r_2|$ are width and height of the sub-image at a specific resolution level. As seen in Figure 1, a subsampling by 2 is done by DWT on both directions of r_1 and r_2 , so the number of coefficients in each subimage is one quarter of that of the input image. As a result, the total number of coefficients (features) is about the same as the number of pixels of original image. If the number of pixels is not exactly a power of 2, then the number of coefficients after subsampling is $\lceil |r_1| / 2 \rceil$ or $\lceil |r_2| / 2 \rceil$. For example, in the case of an image of 50×100 , the total number of coefficients is $5050 = 25 \times 50 \times 3 + 13 \times 25 \times 4$.

In the rest of the paper, we will use *wavelet-embryo features* to refer to features generated from embryonic images by wavelet decomposition.

2.2 min-Redundancy Max-Relevance Feature Selection

The dimensionality of the wavelet-embryo features is equivalent to the number of pixels in an image. It is found that using the full set of wavelet coefficients may often lead to inaccurate results for several other problems (Mallet et al. 1997; Pitter and Kamarhthi 1999). In our method, the next step is to select a most characterizing subset of features that can best discriminate the patterns and help annotating embryo images. We consider generating a compact feature set with strong discriminating strength for specific annotations using the min-redundancy max-relevance (mRMR) feature selection method (Peng et al. 2005). The mRMR feature selection algorithm is theoretically formulated as that features should be selected to maximize the statistical dependency between the annotation-distribution of image samples and the joint distribution of

the selected features. Based on information theory, this can be factorized as finding features that are mutually far away from each other (minimum redundancy) but also individually most similar to the distribution of annotations (maximum relevance).

We use mutual information to measure the level of similarity between features. Let S denote the features subset that we are seeking and Ω the pool of all candidate features $\{f_i\}$ (they equal $\{d_{k,n}\}$ in the case of wavelet-embryo features). The minimum redundancy condition is

$$\min_{S \subset \Omega} \frac{1}{|S|^2} \sum_{i,j \in S} I(f_i, f_j)$$

where $|S|$ is the number of features in S and $I(f_i, f_j)$ is mutual information between f_i and f_j ,

$$I(f_i, f_j) = \iint p(f_i, f_j) \log \frac{p(f_i, f_j)}{p(f_i)p(f_j)} df_i df_j$$

where $p(\cdot)$ is the probabilistic density function.

We again use mutual information $I(c, f_i)$ between the target class $c \in \{c_1, \dots, c_K\}$ and the feature f_i to quantify the relevance of f_i for the classification task. The maximum relevance condition is to maximize the total relevance of all features in S :

$$\max_{S \subset \Omega} \frac{1}{|S|} \sum_{i \in S} I(c, f_i)$$

To obtain mRMR features, we optimize these two conditions simultaneously, in quotient form by

$$\max_{S \subset \Omega} \left\{ \sum_i I(c, f_i) / \left[\frac{1}{|S|} \sum_{i,j \in S} I(f_i, f_j) \right] \right\} \quad (2)$$

The solution of Eq. (2) can be computed efficiently in $O(|S| \cdot N)$ time, with N being the total number of features in Ω .

2.3 Design of the Recognition/Annotation System

Our automatic annotation system has two tiers. At the first tier, an incoming image is assigned to a specific developmental stage in the course of fly embryogenesis. Follow the convention of Berkeley Drosophila Genome Project (<http://www.fruitfly.org>), there are six stage-ranges: stage 1-3, stage 4-6, stage 7-8, stage 9-10, stage 11-12, and stage 13-16.

At the second tier, image ontology annotation terms are assigned to an input image. As explained earlier, the problem of annotating images is multi-objective, we solve it by decomposing it into multiple binary subproblems, for each of which we train a classifier to predict a particular annotation term based on the *one-vs-others* method. In another word, for a particular annotation term, the training samples are associated with labels 1 or 0 depending on whether this image sample has this annotation or not in the training data. In the testing phase, this trained classifier is responsible for predicting whether or not an input image should be associated with this image ontology annotation term. In this way, a series of parallel classifiers will determine if an image has any of the annotations in the training set.

The trained classifiers also calculate a probabilistic confidence score in order to give users a quantitative measure of predictions. Since classifiers for different annotations are trained independently, we are able to add annotation-specific classifiers based on

user requirement and/or system complexity constraints, without altering previously trained classifiers.

For our problem, one challenge in designing classifiers is that an annotation term usually corresponds to only a small portion of images. For example, in the evaluation set of genes we collect, even the most common annotation terms are associated with only about 20% image samples. As a result, for the one-vs-others binary classifier, the sample numbers for class 1 (with the annotation) and class 0 (without the annotation) are very unbalanced, because most samples are labeled as class 0. Therefore, we need to choose a classifier that is able to predict accurately for the small-sample-number class instead of being biased toward the large class.

We use Linear Discriminant Analysis (LDA) to implement the binary annotation classifier. LDA is a commonly used statistical algorithm for separating samples of two or more categories (Webb 2002). It is reported in literature that linear discriminant analysis may give better performance than the Quadratic Discriminant Analysis (QDA) when true covariance matrices are unknown and the sample sizes are small (O'Neill 1992; Webb 2002). Our empirical comparison with the Support Vector Machine (Vapnik 1995; Chang and Lin 2001) shows LDA has a relatively balanced performance on both large and small classes. These observations help decide our choice of classifier. (Results are omitted due to the space limitation.)

Denote the posterior probability of an image x having an annotation as $p(y=1|x)$, where y ($y \in \{0,1\}$) is the class label of the image. It can be shown that, based on the assumption of multivariate normal density of LDA, we have

$$\log p(y=1|x) = g_1(x) + b$$

where b is class-independent, and

$$g_1(x) = \log p(y=1) - \frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1 + x^T \Sigma^{-1} \mu_1$$

with μ_1 is the sample mean of class 1, $\Sigma = \frac{1}{n_1+n_0-2} (n_1 \Sigma_1 + n_0 \Sigma_0)$ is the pooled within-class covariance matrix, and n_1 and n_0 are numbers of samples for classes 1 and 0. $g_0(x)$ for class 0 can be calculated similarly. Using

$$p(y=1|x) + p(y=0|x) = 1 \quad \text{and}$$

$$\log p(y=1|x) - \log p(y=0|x) = g_1(x) - g_0(x),$$

we can then compute $p(y=1|x)$ and $p(y=0|x)$.

For embryonic image annotation, the body part structure with gene expression is considered to be present if 1) $p(y=1|x) > p(y=0|x)$; and 2) $p(y=1|x) > TH$, where TH is a threshold set to 0.6 in our experiment. $p(y=1|x)$ is the confidence score of the annotation.

Because a gene may be associated to multiple annotations, we can aggregate all individual annotation confidence scores as an overall rank score,

$$rankscore = \frac{\sum_i^M p_i \times f(p_i - TH)}{\alpha + \sum_i^M f(p_i - TH)} \quad (3)$$

where M is the number of total possible annotations, p_i is the confidence score for a particular annotation term, $f(\cdot)$ is the hard-limiting function with $f(u)=1$ if $u > 0$ and otherwise $f(u)=0$; the parameter α on the denominator is mainly for avoiding division by zero when no annotation is found to be associated to the image.

When α is 0, the *rankscore* is simply the average of the individual probabilities for various annotations.

The *rankscore* of the annotation predictions provides two uses: 1) The predicted annotation results can be ranked/sorted based on these scores. When the annotation system processes a large number of images, most confident predictions with interesting annotation results can be quickly identified; 2) Instead of having only a decision on individual annotations, users will have additional global quantitative control of the confidence level.

3. EXPERIMENTS AND DISCUSSIONS

In this study, we made use of the mRNA gene expression pattern data set of 463 genes generated through our earlier work of gene clustering (Peng et al. 2006), which also explains the details the image preprocessing methods. In most cases, for each gene, there is one representative mRNA expression pattern image determined by experts for each of the 6 stage-ranges (stages 1-3, 4-6, 7-8, 9-10, 11-12, and 13-16) over the entire course of fly embryogenesis. We addressed two questions about our computational approach, as detailed in Sections 3.1 and 3.2, respectively.

- (1) How good is the performance of pattern recognition with mRMR wavelet-embryo-features, compared against several other possible feature extraction methods?
- (2) How can the automatic annotation system be applied to real applications?

3.1 Experiment 1 - Evaluation of the mRMR Wavelet-Embryo Features

We compared our scheme with two feature extraction methods:

- 1) Eigen-embryo decomposition (Peng et al. 2006) that uses principal component analysis to extract prominent image features.
- 2) LeNet features. LeNet, an artificial neural network that can extract local image features, is considered as one of best machines for pattern recognition (LeCun 1998). We used the common structure of LeNet with four hidden layers. Outputs of the 4th hidden layer were used as the extracted features of an embryo image.

In order to evaluate our feature extraction/selection algorithm, we built 10 testing sets using real images drawn from various stage ranges. For each stage range, we built 2 different data sets using two annotations: in one set, each image has only one of the two possible annotations; these annotations will never correspond to one image sample at the same time. In another word, these two annotations are mutually exclusive (M.E.); thus we call the first set the "M.E. set". In the second data set, some images may correspond to two annotations at the same time, i.e., annotations overlap and the problem is multi-objective (M.O.). It is thus called the "M.O. set", which models the more typically situation in our problem.

In addition, as we need a scheme working well with unbalanced sample sizes, the synthetic sets were built in the way that one annotation class has a lot more images than the other: the larger class usually doubles or triples the number of images of the smaller class. This can be seen from the sample numbers shown in both Tables 1 and 2. Overall, with both the M.E set and the M.O. set on

each of the 5 stage ranges (stages 4-6, 7-8, 9-10, 11-12, 13-16), we produced 10 testing sets listed in Tables 1 and 2. (Note: here we did not show results for stage 1-3 because the gene expression

pattern for stage 1-3 is very easy to classify, as it corresponds to only 2 annotations, "maternal" or "none".)

Table 1. Recognition Rates (%) of Feature Extraction/Selection Methods on Synthetic Data Sets with *Mutually Exclusive (M.E.)* Annotations. Top 20 eigen features are used. Top 20 features are selected when using mRMR feature selection. R_s : Recognition rate on the smaller class only. R_A : Overall recognition rate on the whole data set for all classes. (Annotation Short Terms: CB - cellular blastoderm; TMA - trunk mesoderm anlage; PEA - posterior endoderm anlage; TMP - trunk mesoderm primordium; PEP - procephalic ectoderm primordium; HPP - hindgut proper primordium; PMP - posterior midgut primordium; EBNS - embryonic central nervous system; VNC - ventral nerve cord.)

Stage Range	Set Size	Set Details	Score	Eigen-Embryo	LeNet	LeNet + mRMR	Wavelet-Embryo	Wavelet Embryo + mRMR
4-6	98	78CB+20Subset	R_s	70	75	70	70	100
			R_A	82	85	83	82	98
7-8	69	49TMA+20PEA	R_s	60	50	55	60	100
			R_A	67	59	80	67	100
9-10	69	49TMP+20PEP	R_s	65	50	60	65	100
			R_A	71	52	83	71	100
11-12	61	41HPP+20PMP	R_s	45	50	90	55	95
			R_A	59	69	90	64	98
13-16	32	20EBNS+12VNC	R_s	83	58	67	83	100
			R_A	84	59	81	84	100

Table 2 Recognition Rates (%) of Feature Extraction/Selection Methods on Synthetic Data Sets with *Multi-Objective (M.O.)* Annotations. Short terms and testing parameters are the same as Table 1.

Stage Range	Set Size	Set Details	Score	Eigen-Embryo	LeNet	LeNet + mRMR	Wavelet-Embryo	Wavelet Embryo + mRMR
4-6	152	78CB+20Subst+54Both	R_s	49	42	55	49	76
			R_A	64	63	71	64	82
7-8	89	49TMA+20PEA+20Both	R_s	45	50	58	45	83
			R_A	40	46	62	40	85
9-10	87	49TMP+20PEP+18Both	R_s	42	34	42	39	79
			R_A	49	33	55	47	80
11-12	75	41HPP+20PMP+14Both	R_s	38	32	59	38	88
			R_A	33	36	55	37	89
13-16	97	20EBNS+12VNC+65Both	R_s	57	53	70	57	87
			R_A	55	48	70	55	86

We compared wavelet-embryo features against eigen-embryo features and LeNet features. Input images have the size 50*100 pixels. For eigen-embryo features, we used the top 20 principal components based on the best performance tested in the range of 10-50 principal components. Without feature selection, dimensions of raw LeNet and wavelet features are 3036 and 5050, respectively. For a fair comparison, we also selected 20 top LeNet and wavelet-embryo features using mRMR. These comparison schemes are shown in Tables 1 and 2, where we report recognition rates using Leave-One-Out Cross-Validation (LOOCV) and the LDA classifier (Section 2.3). LOOCV uses each image as a testing sample while the remaining images are used to train the classifier, and sums up all errors in testing.

Tables 1 and 2 list the recognition rates of these 5 feature extraction/selection methods on M.E. and M.O. sets, respectively. Since the data sets are unbalanced, we report both the recognition rate on the small annotation class only (R_s in Tables 1 & 2) and the overall recognition rate (R_A in Tables 1 & 2).

We have several major observations from Tables 1 and 2:

1) Both result tables consistently show that, in both M.E. and M.O. situations, mRMR selected wavelet features delivers a superior performance than all 4 other types of features. For

example, in Table 1 we see that while eigen-features, LeNet features, and wavelet features without mRMR selection lead to 50%~80% recognition accuracy, the mRMR wavelet-embryo features consistently yield close to 100% accuracy. This indicates the combination of wavelet-embryo features and mRMR feature selection is excellent to obtain the most distinguishing feature sets. We thus used mRMR wavelet-embryo features in our annotation system.

- 2) Feature selection is essential for the success of the proposed used of wavelet-embryo features. Using all wavelet coefficients turned out to be redundant and led to less accurate results, which is comparable to using eigen-embryo features, than using mRMR wavelet-embryo features. On the other hand, applying mRMR feature selection on LeNet features did not achieve as good results, which shows that wavelet decomposition does capture more discriminative information from embryo images.
- 3) Our results confirm that the multi-objective (M.O.) recognition problem is harder than the mutually exclusive (M.E.) situation. Recognition rates in Table 2 are in general lower than those in table 1. This shows the difficulty of building the automatic annotation system for fly embryo pattern images that are usually multi-objective.
- 4) Comparing the R_A and R_s recognition rates of the mRMR wavelet-embryo features we can see that the two numbers are often very close to each other. This indicates with the

good features, the LDA classifier works effectively on unbalanced data sets without being biased toward the larger class significantly. Thus we used LDA as the classifier in our automatic annotation system.

3.2 Experiment 2 - Evaluation of the overall automatic annotation system

We applied the best performing features, mRMR-selected wavelet-embryo features, to evaluate our two-tier automatic annotation system for gene expression images. The evaluation set consists of expression patterns of all 463 genes extracted from the BDGP gene expression database (<http://www.fruitfly.org>).

- Tier 1: An input gene expression pattern image is automatically assigned to a stage-range using our algorithm.

This is a mutually exclusive six-class recognition problem since an image belongs to one and only one stage range out of six possibilities – stages 1-3, 4-6, 7-8, 9-10, 11-12 and 13-16. For each image, we used top 20 wavelet-embryo features selected by mRMR. 10-fold cross validation was used to report the recognition rate. This commonly used method randomly separates the our data set as 10 portions, each is iteratively used as a test set while the remaining 9 subsets are used in training the classifier. The final recognition rate is computed based on the total error tested using all portions.

As shown in the confusion matrix of prediction and detailed prediction table in the supplement materials, for all our pattern images, we achieved the recognition rate of 99.55% using wavelet-embryo features and LDA classifier. This number indicates that our system can assign an incoming image to one of the six stage-ranges in a very reliable way.

- Tier 2: Once an image is assigned to a stage range, our system automatically assigns annotation terms to it. This is a multi-objective problem for which we developed binary classifiers for every annotation of interest.

Top 20 mRMR-selected wavelet-embryo features were used in training each LDA classifier. As a quantitative assessment of the results, LOOCV is used to produce the prediction accuracy.

The annotation system produces three pieces of information for each testing image:

- 1) The decision whether the specific annotation is considered "present" in this image;
- 2) The estimation confidence score of each annotation;
- 3) The *rankscore* of all annotations given to this image.

These results are presented In Table 3. Genes are sorted so that embryo images with the most confident annotation predictions are shown at the top. Entries with a probabilistic confidence score lower than 0.6 are marked with a dash "-" to indicate that our system did not predict the respective annotation is "present" in this image. The complete annotation tables for all embryogenesis stages are available on authors' websites as supplements. Table 3 shows partial annotation results of stages 11-12. The top 30 ranked genes with their expression pattern images are listed together with their probabilities of having any of the most popular 5 annotations "present" at the specific stage range. The supplementary materials include prediction results of a larger set of annotations varying from 10 to 18 annotations per stage range.

It can be seen from the result table that the majority of our automatic annotations are consistent with the expert's annotations stored in the BDGP database. This is evident that our system is a meaningful effort to address this challenging problem.

Our system can successfully recognize patterns and annotations for some pretty poor images in term of the blur and deformation, such as the patterns of gene CG12157 (row 27), CG4608 (row 29), all of which are successfully annotated without any error.

It is also worth noting that the BDGP database may have some missing annotations such as row 12 in Table 3 (gene *bow1*, CG10021) for stages 11-12. In this case, our system is still capable of predicting some gene expression image-structures, which are likely to be correct if we visually compare them with other genes with similar annotations. For example, for this gene *bow1* our system predicts it has two annotations PMP and AMP; when we check other genes like CG10535, CG33071, etc, with the same annotations, we can see similar patterns visible in the respective images. This suggests that the automatic annotation system may be used to fill missing annotations for existing databases besides being used for annotating newly collected pattern images.

4. CONCLUSIONS

In this paper we have proposed a practical paradigm to recognize and automatically annotate *in situ* gene expression patterns of fly embryos by combining the minimum redundant wavelet-embryo features and a two-tier classification system. We have achieved promising results using patterns of the entire embryogenesis course of 463 fly genes.

REFERENCES

- Carson, J.P., Ju, T., Lu, H.C., Thaller, C., Xu, M., et al. (2005) A Digital Atlas to Characterize the Mouse Brain Transcriptome. *PLoS Comput Biol* 1(4): e41
- Chang, C. and Lin C., (2001) LIBSVM: a library for support vector machines, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Chien, J.T., Wu C.C. (2002) Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(12):1644-1649.
- Daubechies, I. (1992) Ten Lectures on Wavelets, Science for Industrial and Applied Mathematics. Society for Industrial and Applied Mathematics
- LeCun, Y. et al. (1998) Gradient-based learning applied to document recognition. *Proc. IEEE*.
- Mallat, S.G. (1989) A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11, 674-693.
- Mallat, S. G.(1999) A wavelet tour of signal processing. Academic Press.
- Mallet Y., Coomans D., Kautsky J., and De Vel O. (1997) Classification Using Adaptive Wavelets for Feature Extraction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 1,058-1,066.
- Manjunath B. S., Ma W. Y. (1996) Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):837-842.
- Pan, J.Y., Balan, A., Xing, E.P., Traina, A., and Faloutsos, C., (2006) Automatic Mining of Fruit Fly Embryo Images, *Proc. ACM SIGKDD 2006*.
- Peng, H. and Myers, E.W. (2004) Comparing in situ mRNA expression patterns of Drosophila embryos. *Proc. RECOMB 2004*, pp.157-166.
- Peng, H., Long, F. and Ding, C. (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 27(8), 1226-1238.
- Peng, H., Long, F., Eisen, M. and Myers, E. (2006) Clustering gene expression patterns of fly embryos. *Proc. ISBI 2006*, pp.1144-1147.
- Pitter S. and Kamarthi S. V. (1999) Feature extraction from wavelet coefficients for pattern recognition tasks. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(1): 83-88.
- O'Neill, T.J. (1992) Error rates of non-Bayes classification rules and the robustness of Fisher's linear discriminant function. *Biometrika*, 79(1), 177-184.

Resnikoff, H.L. and Wells, R.O. (1998) *Wavelet Analysis, The Scalable Structure of Information*, Springer New York.

Tomancak P., Beaton, A. et al. (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis," *Genome Biology*, 3(12).

Vapnik, V (1995) *The nature of statistical Learning theory*. Berlin: Springer-Verlag.

Webb, A (2002) *Statistical Pattern Recognition*, 2nd Edition, John Wiley & Sons.

Table 3. Predicted annotations for images at stage 11-12. Short Terms: HPP: hindgut proper primordium; PMP: posterior midgut primordium; AMP: anterior midgut primordium; PP: protocerebrum primordium; DEP: dorsal epidermis primordium.

	Gene Name	Pattern Image	Rank Score	Annotation Terms					Our Automatic Annotations	Manual Annotations from BDGP Experts
				HPP	PMP	AMP	PP	DEP		
1	CG10535		0.95	-	1	1	-	-	PMP; AMP	PMP; AMP
2	CG3224		0.94	-	0.98	1	-	-	PMP; AMP	PMP; AMP
3	CG33071		0.94	-	0.97	0.99	-	-	PMP; AMP	PMP; AMP
4	CG9322		0.94	0.99	-	-	-	0.97	HPP; DEP	HPP; DEP
5	DI (CG3619)		0.94	0.97	1	1	0.95	0.9	HPP; PMP; AMP; PP; DEP	HPP; PMP; AMP; PP; DEP
6	Toll-6 (CG7250)		0.94	-	-	-	0.99	0.99	PP; DEP	PP; DEP
7	Wnt4 (CG4698)		0.94	0.99	-	-	0.95	0.96	HPP; PP; DEP	HPP; PP; DEP
8	yellow-e3 (CG17045)		0.94	0.99	-	-	0.94	1	HPP; PP; DEP	HPP; PP; DEP
9	CG32685		0.93	-	-	0.96	0.98	-	AMP; PP;	AMP; PP;
10	CG9836		0.93	0.99	0.94	0.94	-	-	HPP; PMP; AMP;	Annotations are missing in BDGP
11	TXBP181-like (CG2072)		0.93	-	0.98	0.97	-	-	PMP; AMP	PMP; AMP
12	bowl (CG10021)		0.93	-	0.98	0.98	-	-	PMP; AMP	Annotations are missing in BDGP
13	bys (CG1430)		0.93	-	0.99	1	-	0.88	PMP; AMP; DEP	PMP; AMP
14	CG6361		0.92	-	0.94	0.95	-	0.97	PMP; AMP; DEP	PMP; AMP; DEP
15	CG7993		0.92	0.89	0.95	0.99	-	-	HPP; PMP; AMP	PMP; AMP
16	GATAe (CG10278)		0.92	-	0.97	0.97	-	-	PMP; AMP	PMP; AMP
17	Ugt36Bc (CG17932)		0.92	-	0.94	1	-	-	PMP; AMP	PMP; AMP
18	car (CG12230)		0.92	-	0.95	0.99	-	-	PMP; AMP	PMP; AMP
19	dbo (CG6224)		0.92	0.97	-	-	0.96	-	HPP; PP	HPP; PP
20	CG13126		0.92	-	0.97	0.95	-	-	PMP; AMP	PMP; AMP
21	CG33099		0.91	0.85	0.97	1	-	-	HPP; PMP; AMP	AMP
22	CG10440		0.91	-	-	-	0.96	0.95	PP; DEP	PP
23	Ip259 (CG5277)		0.91	0.82	1	0.99	-	-	HPP; PMP; AMP	HPP; PMP; AMP
24	NetA (CG18657)		0.91	-	-	-	1	-	PP	PP
25	Pfrx (CG3400)		0.91	-	0.96	0.95	-	-	PMP; AMP	PMP; AMP
26	SP555 (CG14041)		0.91	-	-	-	1	-	PP	PP
27	Tom40 (CG12157)		0.91	1	-	-	-	-	HPP	HPP
28	Ugt37b1 (CG9481)		0.91	-	-	-	1	-	PP	PP
29	bnl (CG4608)		0.91	-	-	-	-	1	DEP	DEP
30	fkh (CG10002)		0.91	0.89	0.94	0.98	-	-	HPP; PMP; AMP	HPP; PMP; AMP; PP