Characterizing Patterns in Plant Somatic Embryo Cultures: Its Morphology and Development

Hugo Vits, Chung-Ming Chi and Wei-Shou Hu
Dept. of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455

E. John Staba
Dept. of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455

Todd J. Cooke
Dept. of Botany, Univerity of Maryland, College Park, MD 20742

While being developed, plant somatic embryos change shape and increase size. An effective kinetic description of growth and development of somatic embryos is important for process scale-up and optimization. An essential component of such a kinetic description is the developmental characterization of the individual embryos present in culture. Embryo morphological data obtained by image processing techniques were transformed into size and size-independent morphological descriptors. Qualitative relations between the descriptors and geometric properties of the embryos were established to interpret the results. For training, a branch-and-bound search technique was used to search for optimal subsets of descriptors, as determined by member clustering and class separability properties evaluated from within-class and between-class scatter matrices. In the classification mode, individuals were identified using a voting nearest neighbor classifier. This nonparametric nearest-neighbor classifier was trained on optimal projections of the feature space established from developmental stage discrimination (branch-and-bound algorithm). Using a test population, normal and abnormal embryos and callus were assigned to six morphological classes. The image-analysis-based classification was in 80–90% agreement compared to the results obtained through visual classification by an experienced operator.

Introduction

An angiosperm plant is reproduced naturally through development of embryo. Formation of the embryo begins with the division of the fertilized egg or zygote within the embryo sac of the ovule. Through an orderly progression of divisions, the embryo eventually differentiates, matures, and develops into the new plantlet. Alternatively, the embryo can be derived from the somatic cells from the vegetative part of the plant such as leaf, root and stem by manipulation. This regeneration process, different from the nature pathway, is called somatic embryogenesis. Recently, this technique has been extended to the regeneration of nonangiosperm species (Ahuja, 1993; Bajaj, 1992).

The development of new crop and woody plant varieties has been an unrelenting pursuit of humankind over time. Nowadays, the classical breeding methods can be supplemented with genetic engineering approaches. Many of the DNA manipulations are performed at the single-cell or small-tissue scale and, therefore, require additional procedures to regenerate whole plants. Currently, sterile micropropagation of shoot meristems, whole shoots and somatic embryos is either applied commercially or in the process development stage (Kurtz et al., 1991). Somatic embryos follow the morphological pathways similar to the zygotic processes occurring in plant seeds after fertilization. Somatic embryogenesis can be performed in suspension and thus offer the advantage of easier processing

Correspondence concerning this article should be addressed to W.-S. Hu.
Current address of H. Vits: Environmental Research, Shell Research Ltd., Sittingbourne, Kent ME9 8AG, England.

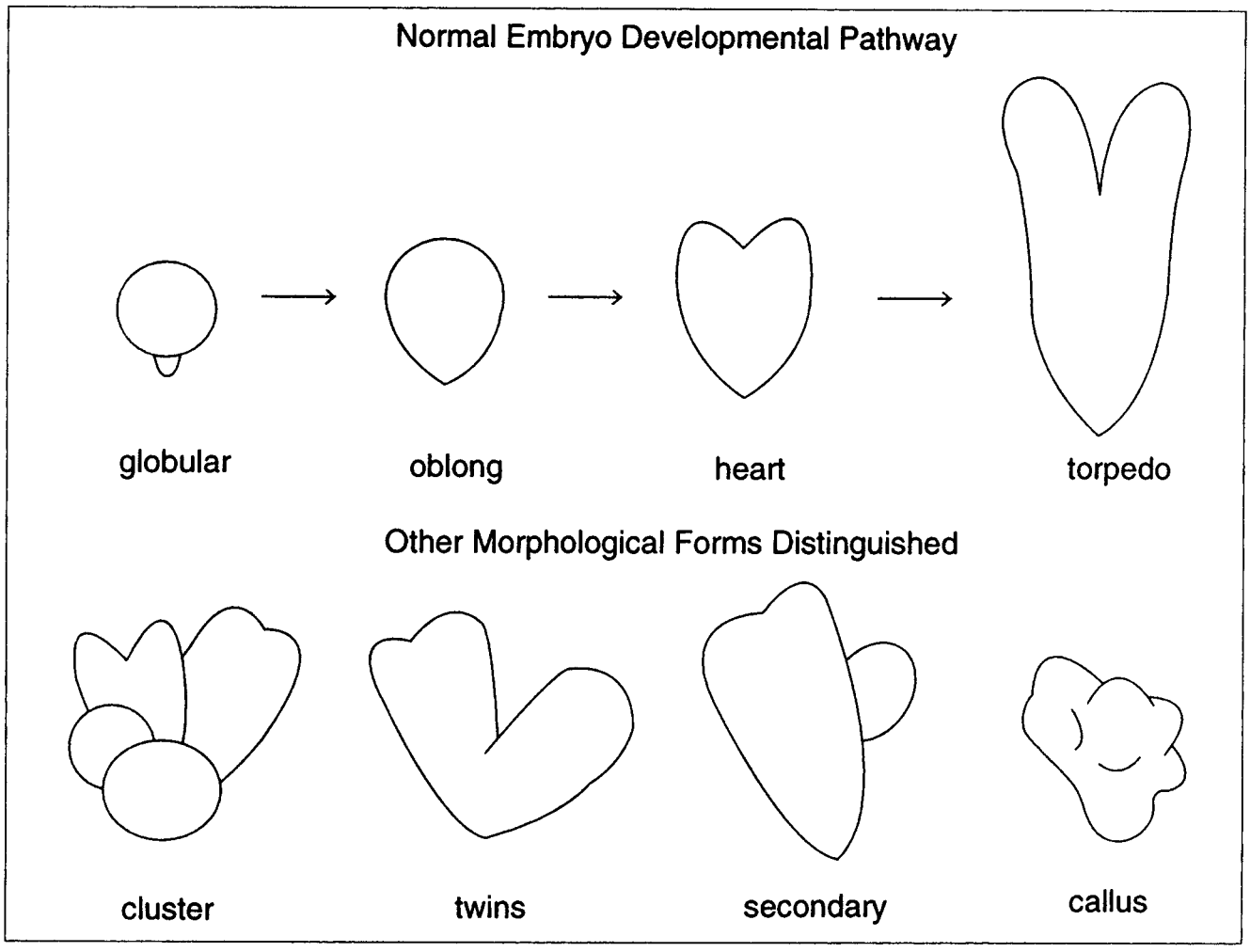

Figure 1. Developmental stages of carrot somatic embryos and abnormal embryos.

over other micropropagation methods. Customarily, the normal embryo developmental pathway is described qualitatively as a time sequence of globular, heart and torpedo stages (Figure 1) (Steeves and Sussex, 1989). Somatic embryogenesis is usually initiated by a proper conditioning of pluripotent cells with plant growth regulator, plant growth regulator analogs, and/or physicochemical stresses (Raghavan, 1986).

In many cases, the technologies developed for microbial fermentation can be readily adapted for large-scale cultivation of somatic embryos in suspension cultures (Ammirato and Styer, 1985; Ammirato, 1987; Preil, 1991). At maturity, the torpedo embryos can be embedded in gel matrices for mechanical field delivery as synthetic seeds (Redenbaugh, 1993; Redenbaugh et al., 1991). Current high process costs due to inefficiencies have limited the introduction of these synthetic seeds as direct competitors to traditionally produced seeds. These inefficiencies are usually identified with high frequencies of abnormally developing embryos, and lack of culture reproducibility and developmental synchrony. Thus, understanding the developmental mechanisms through improved kinetic description and state identification is important to improve embryo culture performance.

The kinetic information derived from the carbon and nitrogen-source uptake and biomass time profiles is insufficient to properly characterize heterogeneous embryo populations (Huang et al., 1993). The larger size of the embryo, compared to single cells or small-size aggregates, precludes the current extension of flow cytometry. Alternatively, the detection of differential gene expression and cell activities associated with different developmental stages (Borkird et al., 1988; Crouch, 1982) have not yet been implemented beyond the single embryo level. Current embryo population characterizations derive from the application of image processing technologies to acquire morphological and size data. The resulting correlation with intrinsic developmental phenomena at the metabolic and mechanical levels is indirect. Therefore, the experimental verification of proposed mechanistic hypothesis (Vits et al., 1992) employing image processing techniques must rely on statistical inference methods (Chi et al., 1994).

An effective characterization method should be capable of analyzing sufficiently large populations in a short time. In addition, it has to be able to separate the information of the overall growth rates and the spatio-temporal distributions of growth rates reflected as morphogenetic changes into inde-

pendent regimes (Vits et al., 1992). Valuable mechanistic and process information could also be obtained from the comparison of frequency and nature of aberrant development. Hence, the classification method should cover both normal and abnormal morphological pathways.

Previous embryo classification efforts have concentrated on selection of features describing the different embryo stages and process automation. Schiavone and Cooke used geometric descriptors such as diameter, axial length, cotyledon length, and various combinations of these variables to classify normal developing embryos (Schiavone and Cooke, 1985). They observed that the heart and torpedo shapes formed a continuum, instead of two discrete classes. Cazzulino et al. (1990a) implemented an embryo shape recognition software on an image analysis station based on geometric ratios describing "elongation," "roughness" and "form" of the embryos. This system was used to study the kinetics of carrot embryo development in shaker flasks (Cazzulino et al., 1990b). Circularity and the features proposed by Schiavone and Cooke (1985) have been implemented for birch embryo recognition (Hamalainen et al., 1993). This recent embryo classifier, much like the earlier ones, did not discriminate among different abnormalities or provide information on the size distribution. In addition, some employed features depended on the position of the centroid in the projected embryo image, and as shown elsewhere (Mitchell and Grogan, 1984), slight alterations in the embryo contours may have significant effects in this extraction step. On-line callus growth estimation has been attempted from particle-size measurements (Harrell et al., 1992). In all the classifiers developed so far, the embryos were univocally assigned to one of a set of morphological classes. However, developmental changes are continuous and shape transitions are set somewhat arbitrarily (Schiavone and Cooke, 1985). It may be relevant to analyze the developmental information as continuous variables, especially for studies of population dynamics. In this article, a nonparametric classifier accounting for the developmental continuity between classes allows the interpretation of positional effect within classes.

As discussed, it is desirable to utilize size-independent morphological descriptors. Morphological Fourier descriptors possess the appropriate mathematical properties and a proven track record in pattern recognition encompassing from aircraft types (Wallace and Wintz, 1980) to handwritten characters (Granlund, 1972). Complementarily, geometric similitudes can be found between the harmonic series and dicotyledoneous embryo shapes. Harmonic series result in a total number of features proportional to the number of discretization points. It is thus essential to determine what subset of these features conveys sufficient information required to keep the results within manageable bound. Optimal clustering methods provide for selection of such a subset of features by considering the scatter within and between training populations. The morphological and/or size properties which provide the most relevant differentiation among the different embryo classes are then determined using statistical criteria. Since some features are expected to have a class separability effect only among a subset of adjacent classes, it may be appropriate to organize the clustering algorithm in a treelike structure. The morphologically similar classes are grouped, and each class can then be separated into individual morphologically unique embryo classes during the successive clustering process. This methodology coupled to the reduction in the number of features involved results in a more efficient morphological delineation of embryo classes. The organization of the information is important in the decision-making process, especially if some features do not carry relevant information for a particular subset of classes. The quality of the decision-making step has to be evaluated in relation to one-step decisions made with a similar number of features in terms of class separability, result interpretation, and computational effort. This article presents a kinetic descriptor model to characterize the embryo development via implementation of an algorithm which generates a large set of morphological and size features for embryos and callus.

Experimental Materials and Methods

Carrot cultures

Carrot callus and suspension cultures (*Daucus carota* L. cv. Danvers) were initiated from carrot seedling hypocotyls. Suspension cultures were maintained in a modified Murashige and Skoog (m-MS) medium (Murashige and Skoog, 1962) supplemented with 1.0 mg/L (4.5 μM) 2,4-dichlorophenoxyacetic acid (2,4-D) (m-MSA medium). The modification consisted of replacing sucrose with an equivalent weight of glucose. The pH of medium was adjusted to 5.8 with potassium hydroxide (KOH) prior to autoclaving. Cell suspensions were cultivated in 500-mL Erlenmeyer flasks each with 100-mL working volume and were subcultured at a fourfold dilution in fresh m-MSA medium every 12 days. The cultures were incubated in the dark at 24 ± 1°C on a rotary shaker (New Brunswick, Edison, NJ) at 80 rpm. Stock callus cultures were maintained on 10-cm petri dishes on the solidified medium (m-MSA medium plus 10 g/L agar) and were transferred every six weeks.

Embryo cultures

Carrot cell clusters for embryogenesis were obtained from 12-day-old suspension cultures. To limit the number of cell clusters originating polyembryos, the suspension was passed through a sieve series (297, 105, 74, 63 and 44 μm openings). The fraction of cell clusters retained on the 44-μm sieve was chosen as inoculum for embryo cultures. Residual 2,4-D was removed by washing the cell clusters three times with m-MS medium without 2,4-D (m-MSE medium), resuspending them into 50-mL fresh m-MSE medium, and incubating for 3 days in 250-mL Erlenmeyer flasks. Clusters were then washed again with 50-mL m-MSE medium and resuspended in m-MSE medium. The final cell cluster concentration was enumerated with a hemocytometer, and only highly cytoplasmic clusters or single cells were counted. Cluster concentration was adjusted to 600–900 clusters per mL with m-MSE medium for embryo initiation, and 100-mL of suspension was dispensed into each 500-mL Erlenmeyer flask and incubated in the dark on a shaker at 80 rpm and 24 ± 1°C.

Data Processing and Methodology

Image acquisition and preprocessing

Samples from the embryo cultures were placed on a petri dish and scanned using an inverted microscope (Carl Zeiss Inc., Thornwood, NY) equipped with a motorized stage and a video camera (Hamamatsu Photonics, Hamamatsu, Japan).

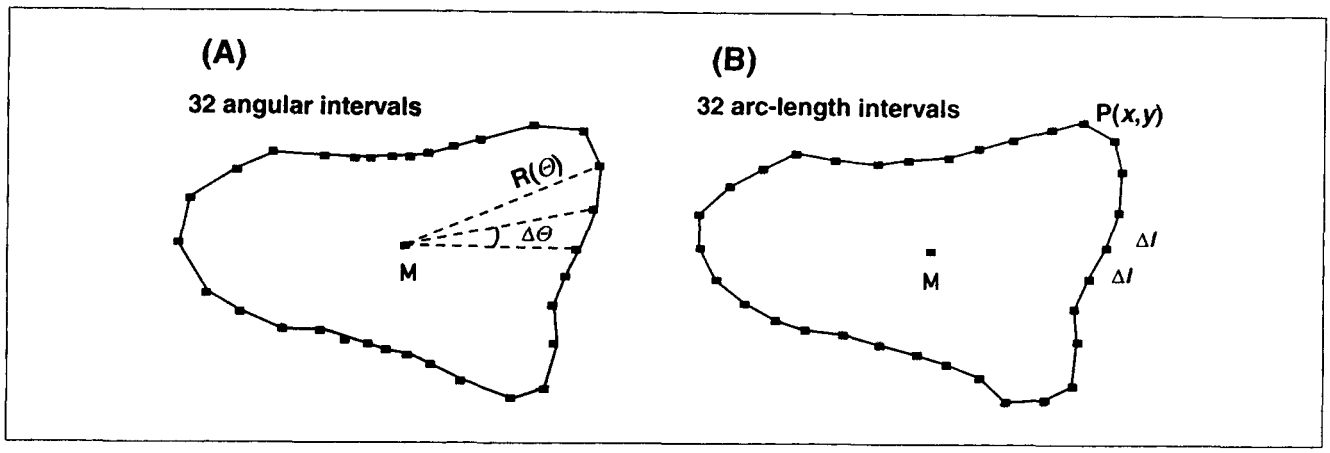

Figure 2. Contour discretization methods: (A) constant angular increment method; (B) constant arc-length increment method.

The acquired images (with resolution 768 × 480 and pixel aspect ratio 1:1) were stored on optical disks in a CD-ROM recorder (Matsushita, Osaka, Japan). Image preprocessing was performed on a microcomputer with array processor (Kontron Bildanalyse, Eching, Germany). Open contours were discarded to avoid the identification of embryos intersected by the image edges. A data file containing the coordinates of one contour point and the centroid for each object was created. Processed images and data files were transferred concurrently to a workstation (Apollo Computer Inc., Chelmsford, MA) for further classification and/or training steps.

Embryo classification

Normalized Fourier Morphological Descriptors. In a previous effort, Cazzulino et al. (1987) computed Fourier spectra for carrot somatic embryos at the three canonical developmental stages, for the function defined as the distance from the centroid to projected perimeter, sampled at a constant angular increment (Figure 2A). Later, it was suggested (Cazzulino et al., 1991) that image analyzers were imprecise in their centroid determination and that the information provided by standard software packages was inadequate for reliable Fourier analysis. The method, as presented by Cazzulino et al. (1987), has three limitations: (1) the unavailability of protocols to ensure consistency (similarity in the Fourier spectra) between similar shaped embryos independent of size, orientation, starting point and chirality effects; (2) the possibility of multivalued dependencies for complex shapes such as abnormally developing embryos, embryo clusters and large callus particles; and (3) the nonuniformity associated with perimeter sampling at constant angular increment.

These limitations can be overcome by first projecting the embryo contour onto a complex plane (Granlund, 1972). Contour is then represented as a complex periodic function $z(t) = [x(t), y(t)]$ with t a parameter. The periodicity of $z(t)$ results from "traveling" with a uniform velocity repeatedly over the contour, while the time required to traverse the contour is 2π. $z(t)$ can then be expanded into a Fourier series:

$$z(t) = \sum_{n=-\infty}^{\infty} A(n)e^{jnt}; \quad A(n) = \int_0^{2\pi} z(t)e^{-jnt}dt \qquad (1)$$

where $A(n)$ is the Fourier coefficient of order n. The contour is sampled at N points, separated by perimeter arcs of equal length (Figure 2B). A fast Fourier transform (FFT) (Press et al., 1989) is implemented to provide $N+1$ Fourier coefficients $(-N/2, \ldots, 0, \ldots N/2)$. Such a sampling protocol assures a well-weighted representation of the contour. Currently, we employ a 32-point sampling scheme as a compromise between contour resolution and computational time.

A normalization procedure similar to that described by Wallace and Wintz (1980) is used to reduce the effects of size, orientation and starting point on the computed Fourier coefficients. Consistency in going over the embryo contours in a counterclockwise manner is achieved by computing the cross product of the vectors joining the starting point to the centroid and the starting point to a contour point located in the neighborhood of the starting point. In consideration of possible variations in the convexity of embryo projections, this operation is repeated a few times to assure a correct result. The descriptor normalization procedures are based on the linearity of the Fourier transform. First, each object is translated so that its centroid coincides with the origin of the coordinate system, a step equivalent to setting $A(0) = 0$. All coefficients are divided by the magnitude of the largest coefficient $[A(-1)]$, thus scaling the object size by the magnitude of $A(-1)$. $A(-1)$ is always the largest coefficient when the contour is traced counterclockwise, and the contour does not cross itself. Starting point effects are minimized by requiring that the phases of the two coefficients of largest magnitudes be zero. This is attained by changing the starting point on the boundary and rotating the coordinate basis and is equivalent to multiplying each coefficient $A(i)$ by:

$$e^{j\left[\frac{(i-k)u+(-1-i)v}{k+1}\right]} \qquad (2)$$

with u and v denoting the phases of $A(-1)$ and $A(k)$ which is the coefficient of second largest magnitude; j is the imaginary number $\sqrt{-1}$. However, this condition is generally satisfied by more than one contour configuration. The phases of $A(-1)$ and $A(k)$ coincide with $k+1$ different starting points. Through the rotation operation, both phases can be reduced to zero, so the total number of combinations is given by:

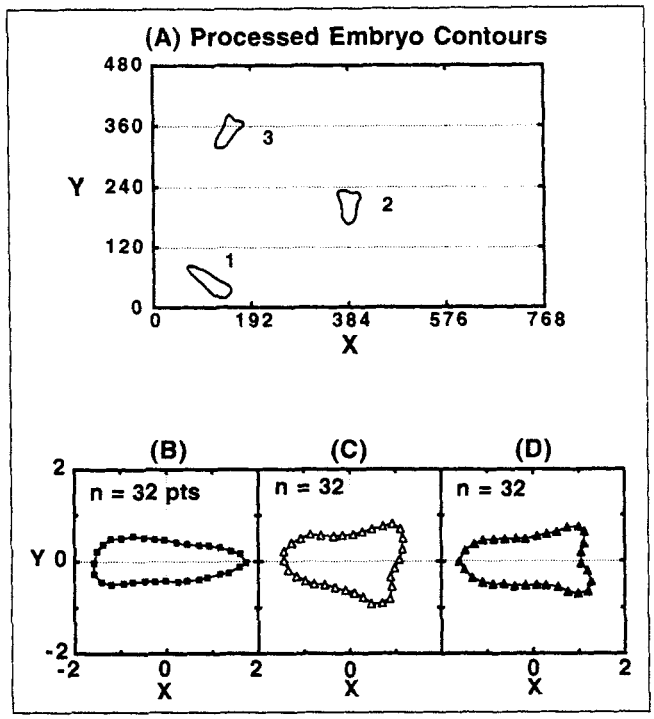

Figure 3. Demonstration of processed image and normalized embryo contours for feature extraction: (A) raw image of embryo 1, 2 and 3; (B) normalized embryo 1; (C) normalized embryo 2; and (D) normalized embryo 3.

Note that the principal axis of the embryo is aligned with the real axis.

$$m = |k+1| \tag{3}$$

To assure consistency in the selection of the configuration, the configuration maximizing the function F was selected (Wallace and Wintz, 1980):

$$F = \sum_{i=-N/2}^{N/2} \Re[A(i)]\,|\Re[A(i)]| \tag{4}$$

where $\Re[A(i)]$ denotes the real part of $A(i)$. As shown (Figures 3A-3D), these procedures align the axis of maximum symmetry of the embryo with the real axis of the complex plane. However, when normalizing embryo shapes, criterion 4 does not always provide a consistent orientation of the embryos. For the embryo as shown in Figure 3D, Eq. 3 yields two configurations (rotated 180° with respect to each other). However, the numerical values of Eq. 4 given by the two sets of Fourier coefficients obtained from Eq. 3 are almost indistinguishable. This limitation is only of minor concern.

Size and Volume Descriptors. The contour is segmented and represented by a set of eight-neighbor chain codes (Freeman, 1961). The codes represent right, up, left and down by 0, 2, 4, 6 and the intervening diagonal directions by 1, 3, 5, 7. The contour perimeter is approximated by the number of even digits plus the square root of two times the number of odd digits. Another easily computable size descriptor is the area enclosed by the contour. The projected area of a closed contour can be determined by integrating its corresponding chain codes (Freeman, 1961). Both perimeter and area are expressed in pixel units.

The orientation of the object is used to compute an estimate of the particle volume. The estimated volume is less erroneous when the object is axisymmetric. This condition is satisfied during early embryo development and lost with progressive development of cotyledons during the heart and torpedo stages. It also degrades with the appearance of abnormalities in embryo morphology. The average volume, in voxels, is computed as:

$$V = A(-1)\frac{(V_+ + V_-)}{2} = \frac{\pi}{2}A(-1)\left(\int_a^b y_+^2\,dx + \int_a^b y_-^2\,dx\right); \quad a \le x \le b \tag{8}$$

where y_+, y_- are the portions of the normalized $z(t)$ curve above and below the real axis, respectively; a and b are the abscissas of the contour intersection with the real axis. The integrals were approximated numerically by applying the trapezoidal rule to the normalized, discrete data set $z(i)$. The actual volume is then obtained by multiplying V by the voxel size. Similar, but simpler, geometric considerations have been utilized to compute cell volumes of filamentous microorganisms (Packer et al., 1992). For nonaxisymmetric objects, the area enclosed by the contour offers an advantage over estimated volume. Although these size descriptors may not always correlate well with biomass, they are included for completeness and used for embryo classification. They are also useful to test which 'size' variable is more representative for description of developmental kinetics.

Combination of Descriptors. It is possible that, due to the complexity of embryo morphology, a combination of descriptors rather than several individual descriptors may be better suited to distinguish different embryo morphologies. For instance, callus particles and cluster embryos usually have highly irregular shapes that are difficult to relate with any individual morphological or size descriptor. To address this problem, additional features are formed by combining basic morphological or size descriptors. Feature M_1 denotes the mean of medium-order Fourier features ($4 \le |i| \le 7$) and S_1 is the reciprocal of the characteristic radius. Both M_1 and S_1 capture the roughness of a contour and are used to discriminate callus aggregates and clusters from embryos. Feature S_2 measures the circularity of contour and describes its deformation from a circle (Veillon, 1986). S_2 is particularly useful in classification of clusters:

$$M_1 = \frac{1}{8}\sum_{i=4}^{7} -\{\log[|A(-i)|] + \log[|A(i)|]\} \tag{5}$$

$$S_1 = \frac{\text{Perimeter}}{\text{Area}} \tag{6}$$

$$S_1 = \frac{\text{Perimeter}}{\text{Area}^{0.5}} \tag{7}$$

Selection of a Subset of Embryo Descriptors. By calculating a Fourier spectrum, a set of N-points in the complex

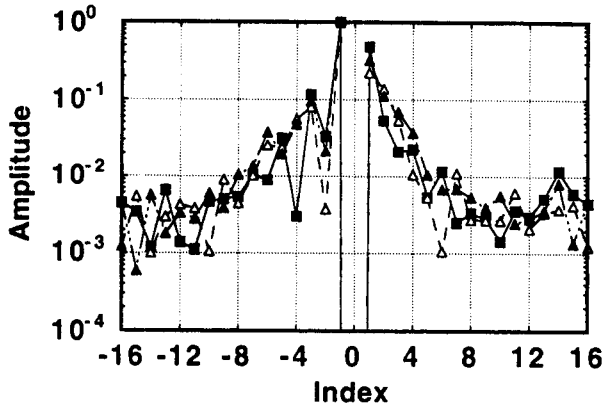

Figure 4. Fourier spectra of the three embryos in Figure 3.

(■) Embryo 1; (△) Embryo 2; (▲) Embryo 3.

plane C is transformed into one point in the complex space, C^N. While the Parseval identity can be used to compare, in a least-squares sense, two contours in the plane (Wallace and Wintz, 1980), minor shape changes may affect the location of the centroids in two otherwise comparable figures (Mitchell and Grogan, 1984). As a consequence, a larger-than-desired least-squares error may result for two otherwise similar contours. Additionally, some of the features may be the net result of perturbations or noise derived from the extraction process; others may be highly correlated during certain stages of morphogenetic development of embryos and thus redundant. On the other hand, these correlated variables, if it is possible to interpret them geometrically, may offer insights into the kinetics of embryo development.

While it may be conceptually easy to compute distances in $C^{N'}$ to establish proximal relations for classification, the scheme may be simplified computationally by determining what combination of features provides sufficient classifying power (N' is the sum of the number of morphological, size and combining features). This can be reformulated as: What combination of features, chosen from the total, N', would provide the best classification once the subset size, m_s, was specified? To avoid the possibility of overlooking a combination of features that could have a 'hidden' separating power, a systematic search method was adopted to search the relevant features (Fukunaga, 1990). Typical Fourier spectra, evaluated for the embryos in Figure 3, show how difficult it would be to search for an appropriate combination, solely based on inspection (Figure 4). A training population which contained representative individuals, each being assigned to a class by an experienced "trainer," was established for selecting relevant features for classification. The classes we chose are based on the previous observation and the need to describe the developmental kinetics. Schiavone and Cooke (1985) observed the nonsimultaneous development of cotyledons and axis in somatic embryogenesis instead of an associated event in zygotic embryogenesis. Characterization of abnormal classes also provides a crucial input to understand the developmental mechanisms.

The previously delineated process can be systematized by optimizing a functional, J, representing dense clustering and cluster separability properties. The functional J increases when the scatter within classes decreases and/or when the scatter between classes increases (Fukunaga, 1990):

$$J = tr(S_w^{-1} S_b) \tag{9}$$

S_w and S_b are within-class scatter and the between-class scatter matrices defined as:

$$S_w = \sum_{i=1}^{L} P_i E\{(X_H - M_i)(X_H - M_i)^T | \omega_i\} \tag{10}$$

$$S_b = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T \tag{11}$$

where L is the number of classes; P_i is *a priori* probability of ω_i and equal to $1/L$; M_i and M_0 are the ω_ith class and the overall mean (expected) feature vectors; X_H is any feature vector in the training database H. J is invariant under any nonsingular coordinate transformation in the feature space. To ensure that all features are of a similar order of magnitude to avoid numerical problems when inverting scatter matrices, the components of the morphological feature vector X are taken to be the negative logarithms of the amplitudes of the normalized Fourier coefficients.

$$X = [-\log|A(-N/2)|, \ldots, -\log|A(-1)|, -\log|A(1)|, \ldots, -\log|A(N/2)|]^T \tag{12}$$

This vector can be extended with the size and combined descriptors.

A branch-and-bound algorithm is employed to take advantage of the monotonicity of J (a subset of features has no better discriminating properties than any larger set containing it), while limiting the search of features. The search process for selecting two optimal features from four features is exemplified in Figure 5. At each level, one of the features is removed and J is computed (Figure 5B). The feature subset (1,2) with the highest J value among the three computed in level 2 is the optimal subset for classification in this example. The computational advantages of this method involved in searching the optimal subsets become evident for large search sets (>15 features). The inversion of the scatter matrices along a branch is performed efficiently using a sequential method (Fukunaga, 1990).

Training procedure

Somatic embryos of different morphologies were collected and analyzed using the procedures described above. While it is clear that embryo development is a continuous evolution process, discrete classes have a qualitative appeal and can 'freeze' development in a morphological sequence. Along with the globular, oblong, heart and torpedo classes of the normal pathway, the following classes were used for abnormal embryo classification: cluster, twins, secondary and callus (Figure 1). During the training stage, embryo data were assigned to different morphological classes by a "trainer." Based on the optimal features determined by the branch-and-bound algorithm, the classes were projected onto the resulted spaces and stripped of individuals fall in the regions of significant super-

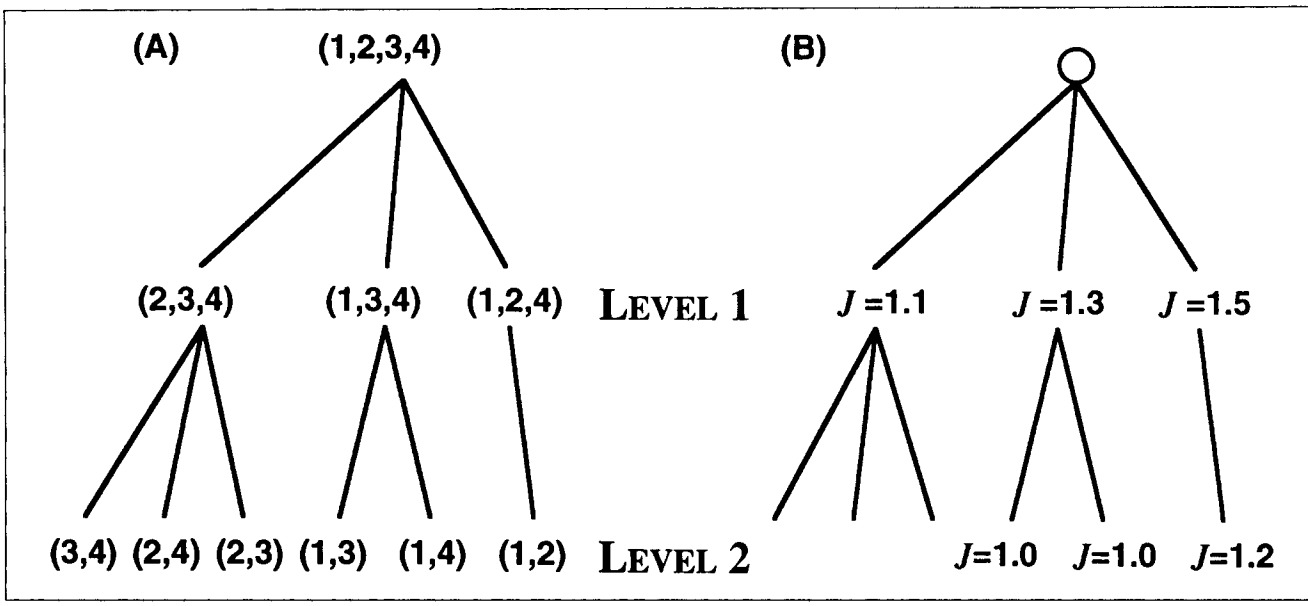

Figure 5. Selection of two optimal features of branch and bound algorithm from a total of four: (A) all six possible combinations considered; (B) at each level, one feature removed and *J* computed at each level.

The computation of *J* values for level 2 begins with the branch that has highest *J* value at level 1 [branch (1,2,4)]. Since *J* value of feature subset (2,3,4) is less than *J* value of feature subset (1,2), there is no need to further compute *J* values of the feature subsets (3,4), (2,4) and (2,3). The feature subset (1,2) with highest *J* value among the three computed in level 2 is the optimal subset for classification in this example.

imposition. These procedures were repeated in terms of the distribution properties of the class until its defined region was achieved. The training database was subsequently formed by combining all the morphological classes, and each class consisted of 200-500 representative individuals. The separation between classes in training database was then optimized by the branch-and-bound procedures, and the selected features were marked for further use in automatic classification.

Automatic classification

We employ a hierarchical classification procedure realizing that not all the features used in a one-step scheme may be necessary to separate certain objects. If the number of features is maintained constant, certain classifications may be performed suboptimally (Park and Sklansky, 1990). In addition, it is possible to follow the classification geometrically if at each step the feature vectors are projected onto a plane or three-dimensional spaces to construct scatter plots. Some heuristics including the assumed closeness of classes based on well-known developmental pathway and application of the branch-and-bound algorithm were used to construct our classification scheme. The resulting procedures are basically a binary separation process (Figure 6), thus the number of features needed at each classification stage can be reduced. The callus particles are first separated from the embryos, which are subsequently classified into three groups. Each group is further classified into different embryo classes.

The embryos are assigned class memberships according to a *k*-nearest neighbor (*k*-NN) method (Fukunaga and Narendra, 1975). Implementing such a nonparametric procedure directly could be computationally prohibitive for large training sets composed of high-dimensional feature vectors. Effectively, no

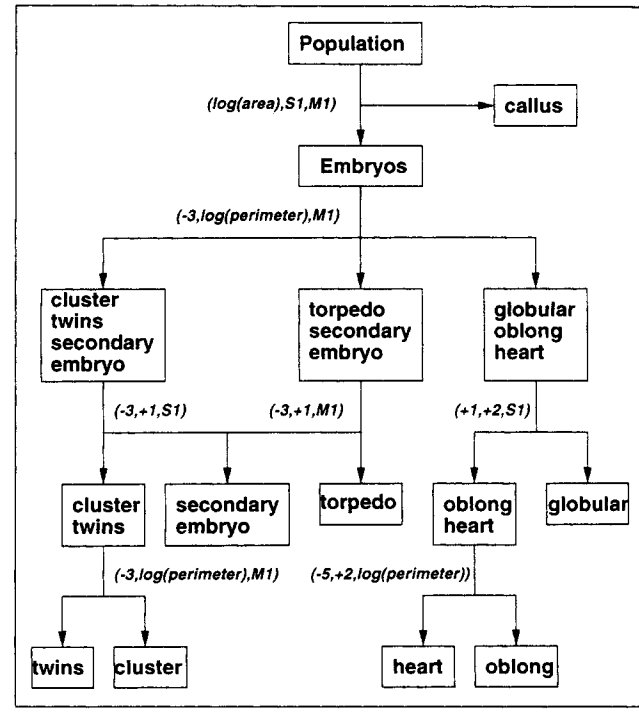

Figure 6. Sequential scheme for automatic classification of embryos.

In level 1, the embryo and callus are separated. In level 2, the embryos are classified into three groups according to group morphological and size characteristics, followed by the discrimination of secondary embryos from each group. Each group is resolved into individual morphological classes.

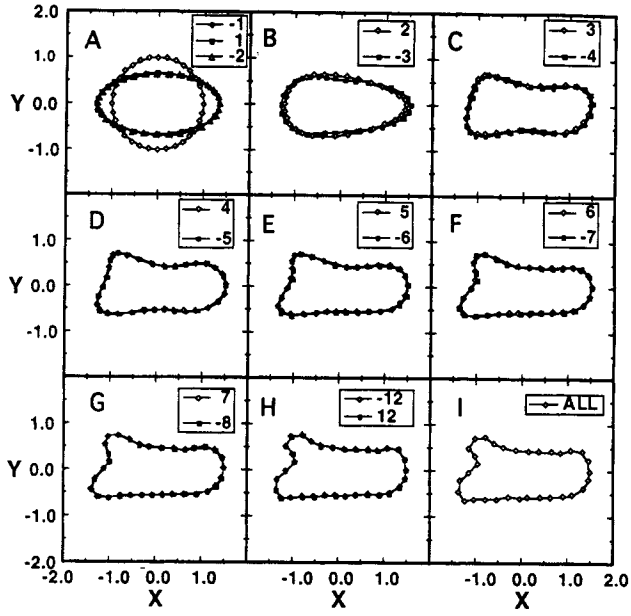

Figure 7. Fourier term contributions to embryo shape reconstruction.

The features added for the reconstruction are shown in the upper right to each reconstructed contour.

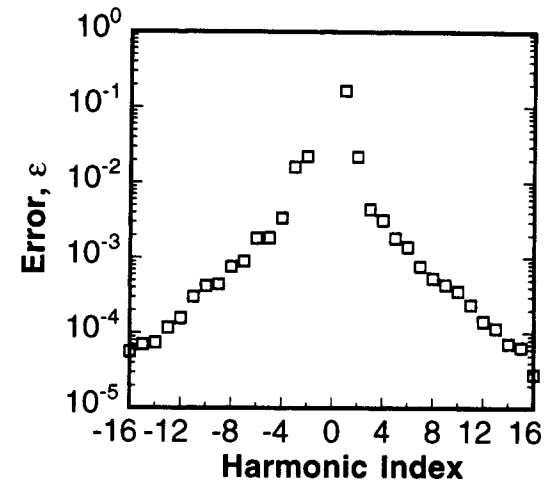

Figure 8. Original contour vs. contours generated from subsets of Fourier coefficients by a least-squares approximation based on Parseval's identity.

The error ϵ results from the elimination of the harmonics of corresponding and higher orders. So the $\epsilon(+1)$ is larger than the $\epsilon(-2)$ (Figure 7). Also there is no need to compute the $\epsilon(-1)$ because of no corresponding contour.

high dimensionality was used in the classification, but it is preferred to keep the flexibility. An alternative approach is to enclose the embryo to be classified in a region determined by intersecting three hyperspheres (Sethi, 1981). To perform this approach, the unknown embryo is projected on a feature subspace based on the optimal features obtained from the branch-and-bound algorithm. The distances are computed from the unknown embryo to three reference points, each generating a hypersphere. (These reference points are arbitrarily chosen, but should not be aligned.) By successive comparison, the neighbors confined in the intersection space of the three hyperspheres are retained. The region is expanded until k neighbors are obtained. In the training database H, there are k feature vectors projecting on the intersecting region). The unknown embryo is assigned to the class that has the maximum number of embryos in the intersecting region.

Results and Discussion

Geometric interpretation of Fourier descriptors

To find a qualitative physical meaning of the Fourier harmonics, we reconstructed a torpedo embryo (Figure 7) step-by-step and performed a sensitivity analysis. The embryo contour was compared to contours generated from subsets of Fourier coefficients by a least-squares approximation based on the Parseval identity (Figure 8). The overall error ϵ remained small if harmonics of order higher than say -6 and 6 are deleted. It increased exponentially with every deletion of lower-order terms, indicating that the more salient shape features were captured by the lower-order terms and that "compensatory" effects between terms were unlikely. As globular and heart embryos are geometrically simpler, this analysis can be applied conservatively and intuitively to them. More complex, abnormal embryos may require higher numbers of terms to attain the same ϵ.

The shape reconstruction started with the transformation of the circle $A(-1)$ into an ellipse $[A(-1)+A(1)]$ with the major (long) axis coincident with the real axis (Figure 7A). $A(-2)$ was comparatively small but provided asymmetry with respect to both the real and imaginary axes. For a given abscissa in the region close to the cotyledons, the negative coordinate was larger than the positive one, suggesting a bulging toward the bottom of the plane. In relation to final shape, this can be interpreted as providing for larger cotyledons. The addition of $A(2)$ increased the "weight" at negative abscissas; the bulging effect of $A(-3)$ made the "girth" of the central portion of the embryo more uniform and slightly elongated the embryo (Figure 7B). The large $A(3)$ coefficient rendered the embryo more "quadrilateral" starting from a "trilateral" structure. This marked the onset of the upper bulge at the radicle end of the embryo (Figure 7C). Note that the point-by-point differences between successive curves became less significant, a fact reflected in the nature of the ϵ dependence. The interpretation of higher-order terms was less evident, but they provided for the deeper cleft between cotyledons and increased contour roughness. Since the magnitude of these contributions was relatively small, their contributions were more likely to slightly alter the local curvature of the contour rather than to introduce extra bulges or sharp angles. Addition of Fourier coefficients of order greater than 12 only slightly affected the contour (Figures 7H-7I), indicating that 32 coefficients were sufficient to describe the contour for our purposes.

Simulations were performed for different combinations of harmonics. The first term, $A(-1)$, corresponded to a circle in the complex plane. The addition of $A(1)$ transformed the circle into an ellipse, thereby altering the shape symmetry from radial to bilateral. As $A(1)$ increased, the shape elongated. $A(3)$, $A(4)$, and other positive index coefficients provided for trilateral, quadrilateral and higher-order polygonal symmetries, respectively. By increasing the magnitude of the coefficients (the

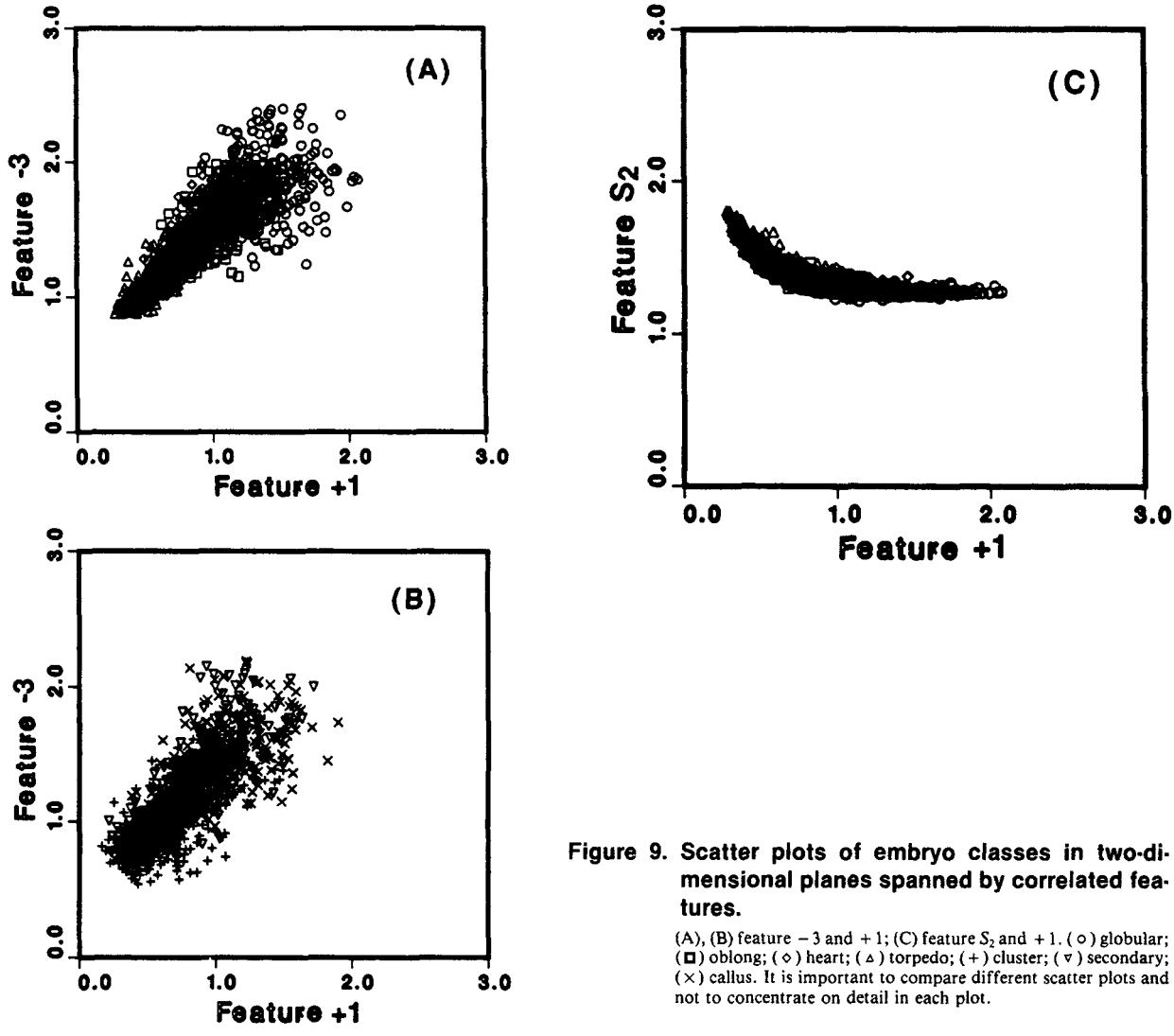

Figure 9. Scatter plots of embryo classes in two-dimensional planes spanned by correlated features.

(A), (B) feature −3 and +1; (C) feature S_2 and +1. (○) globular; (□) oblong; (◇) heart; (△) torpedo; (+) cluster; (▽) secondary; (×) callus. It is important to compare different scatter plots and not to concentrate on detail in each plot.

relative real and imaginary parts determining the phase), the shape can lose convexity and become starlike. The series of negative indices provided for bulges that resembled accreted circles, which mimicked shapes physically similar to embryo clusters. Wallace and Wintz (1980) provide some of these figures (with inverted index notation).

It is expected that with advancing normal development, the magnitude of $A(1)$ will become larger. Globular to heart transitions should be accompanied by an increase in the value of $A(2)$. Abnormal development of extra bulges should be associated with larger values in higher-order contributions of the $A(-3)$, $A(-4)$, $A(-5)$, . . . series. These observations can facilitate the interpretation of classification results based on descriptor subsets, indicating what geometric characteristics of the contours are being used.

Correlated descriptors

Scatter plots at classifier-training stage show that some features correlated strongly for some morphological classes (Schiavone and Cooke, 1985). For instance, in the scatter plot of the feature plane constructed for feature +1 and −3, normally developing embryos fall in a narrow band (Figure 9A). Reference to Figure 7 suggests that elongation (feature +1) is associated with polarity development (feature −3), representing formation of the cotyledon radicle. As the abnormal morphologies mostly scatter below this band, it suggests that there is a rather narrow band of growth pattern compatible with normal transition (Figure 9B). Feature S_2, related to the circularity of embryo contour, also shows correlation with feature +1 (Figure 9C), reflecting the deformation of embryo contour from the circle with the progression of development. These results show that embryo elongation and cotyledon formation may be associated events during the development of normal somatic embryos.

Determination of core classes

For a classification scheme associated with the conventionally characterized stages, the regions corresponding to each class need to be defined in the feature space. Although conventionally characterized by several discrete stages, embryo

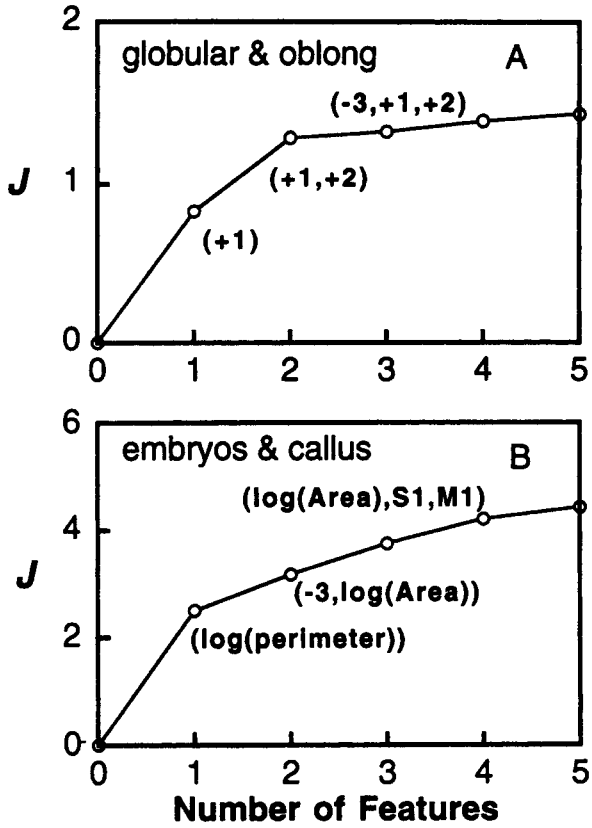

Figure 10. Effect of number of features on class separability: (A) globular and oblong stages; (B) embryos and callus.

development is, in fact, a continuous process. By projecting these embryo classes onto the feature subspace (specified by branch-and-bound algorithm) that inhibited the best separation power, the individuals that exhibited "intermediate" properties (those individuals locating at the overlapping regions among classes) were stripped. These procedures were repeated iteratively until desirable levels of class separation were achieved. As expected, the largest overlaps (nearly half of them were superimposed) occurred for oblong-heart classes. These embryos were difficult to discriminate in an unbiased fashion at the training level.

Effect of feature number on class separability

Limiting the number of features for class separation reduces the computational requirement. Figure 10 shows the effect of number of features on class separability. The rate of convergence to asymptotical separability depends on the particular classification being executed. For example, the separation of globular embryos from oblong embryos was not significantly improved when more than two features were employed. The separation between these classes was exclusively based on morphological features, associated with the elongation and shoot-root polar asymmetry. To discriminate between embryos and callus particles, using only one feature, the perimeter was selected (Figure 10B). At the two-feature level, a combination of area and polar asymmetry provided the best separability. The results also show that the sequence of features formed by adding one extra dimension to the projected subspace need not be inclusive. Also, while some of these selected subsets are rather obvious, other combinations are not. Therefore, the general search method based on branch-and-bound optimization of J constitutes a robust and natural approach. By comparison, other embryo classifiers (Cazzulino et al., 1990a; Hamalainen et al., 1993) look rather arbitrary or their feature set has to be prespecified.

Hierarchical classification

To evaluate the quality of the k-NN algorithm, a population composed of 522 particles (including only embryos and callus not belonging to the training database) was classified by a

Table 1. k-Nearest Neighbor (k-NN) Classification to Operator Microscopic Classification*

	Microscopic Observation	k-Nearest Neighbor Classification							
Embryo Class		Globular Embryo	Oblong Embryo	Heart Embryo	Torpedo** Embryo	Cluster Embryo	Twin Embryo	Secondary Embryo	Callus Particle
Globular Embryo	**87**	**71**	8	2	0	0	0	0	6
Oblong Embryo	**45**	8	**28**	8	0	0	0	0	1
Heart Embryo	**47**	0	17	**27**	3	0	0	0	0
Torpedo** Embryo	**127**	0	0	6	**118**	0	2	1	0
Cluster Embryo	**54**	0	0	0	1	**34**	14	5	0
Twin Embryo	**44**	0	0	0	0	7	**35**	2	0
Secondary Embryo	**38**	0	0	0	5	3	1	**29**	0
Callus Particle	**80**	3	2	0	0	0	0	0	**75**
Total	**522**	**82**	**55**	**43**	**127**	**44**	**52**	**37**	**82**

*The tested population was obtained from a sample having not been analyzed before.
**Torpedo with cotyledons and torpedo without cotyledons.

Table 2. Performance of Embryo Classifiers

Sources	Feature Origin	Embryo and Nonembryos Developmental State (P_1, P_2)	Normal Embryos Developmental State (P_1, P_2)	Abnormal Embryos Developmental State (P_1, P_2)
Cazzulino et al. (1990a) Cazzulino et al. (1991)	morphology size	embryo (0.67, 0.77) nonembryo (0.87, 0.80)	globular (0.89, 0.69) heart (0.49, 0.73) torpedo (0.94, 0.98)	none
Hamalainen et al. (1993)	morphology size	embryo (0.86, **) nonembryo (1.00, **)	globular (0.83, 1.00) *(HT1 + HT2)-heart (0.88, 0.60) (HT2 + HT3)-torpedo (0.88, 0.70)	none
Vits et al.	morphology size	embryo (0.98, 0.99) callus (0.94, 0.91)	globular (0.82, 0.87) oblong + heart (0.87, 0.82) torpedo (0.93, 0.93)	cluster + twins (0.91, 0.98) secondary (0.76, 0.78)

*HT1: heart, HT2: intermediate, HT3: torpedo
**Data not available

trainer and automatically by the k-NN classifier. The results in Table 1 are grouped into eight classes, as shown in Figure 1. The consistency of classification for individuals by the two methods was generally around 60–80%. However, if oblong and cluster are pooled with heart and twins, respectively, the consistency is better than 80%. The classification of globular, torpedo and callus are better defined than classifications of the intermediate stages (oblong and heart stages).

It is interesting to compare the performance of our nearest-neighbor classifier to the performance of the other classifiers. (However, since each work has its own data set, comparison of their performance may not be inclusive.) To do so, we define two frequencies P_1 and P_2.

P_1 = frequency of a particle previously classified as state θ by a trainer, which was also classified as state θ by a pattern classifier

P_2 = frequency of a particle previously classified as state θ by a pattern classifier, which was also classified as state θ by a trainer

By this definition, P_2 will be affected by the classification of other classes. Larger values of P_1 and P_2 imply less misclassification and better performance of the classifier. According to these criteria, the performance of three classifiers is shown in Table 2. The classifier by Cazzulino et al. (1990a) is limited by high probability of misclassification in embryo-nonembryo recognition. Additionally, it lacks the description of abnormal developmental states. The misclassification of heart embryos into globular embryos is also significant. These discrepancies may significantly affect the estimated population distribution and kinetic description of embryo development. The classifier proposed by Hamalainen et al. (1993) shows improved performance in embryo-nonembryo recognition. However, the distinction between heart and torpedo embryos is ambiguous and the abnormal developmental states are not included. Our classifier not only improves the embryo-callus, globular-heart, and heart-torpedo classifications, but also includes the adept classification of several abnormal developmental states. This ability is important among others for the statistical interpretation of the environmental effects. The advantages of our classifier can be attributed to the employment of size-independent morphological descriptors and statistical optimization algorithms.

The results of our automatic classification are satisfactory, and the method is being currently employed to support our experimental work on the linkage between environmental factors and embryo development. However, further improvement is expected through the inclusion of a fuzzy class membership. The fuzzy membership assignment is preferable over a statistically discrete assignment since it reflects better on the continuity of development (Schalkoff, 1992).

Estimated volume distributions

Estimated volume distributions of normal embryo subpopulations are shown in Figure 11. The volume distribution of an embryo population may provide a good approximation for estimation of biomass, thus allowing for rapid computation of biomass distributions among population classes automatically identified by the k-NN classifier. The biomass distribution may then be used to estimate kinetic parameters in population balance models, though the correlation between biomass and estimated biovolume need to be established. Figure 12 shows the estimated volume distribution in a batch somatic embryo culture at day 10 and 20 after initiation. The estimated particle concentrations (including developing embryo and callus) were around 50 and 400 per mL, respectively. While there was a

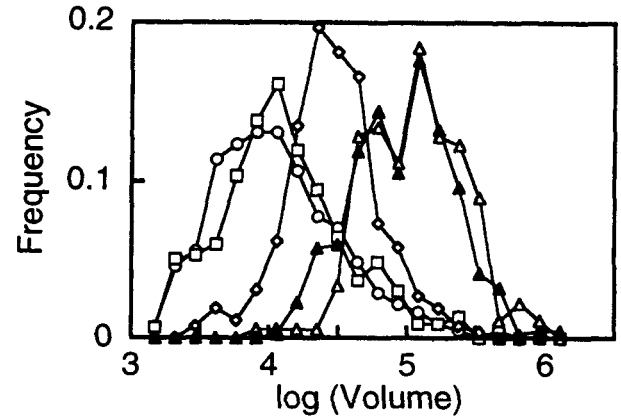

Figure 11. Estimated volume distributions of normal embryo classes.

(○) globular; (□) oblong; (◇) heart; (△) torpedo with cotyledons; (▲) torpedo without cotyledons.

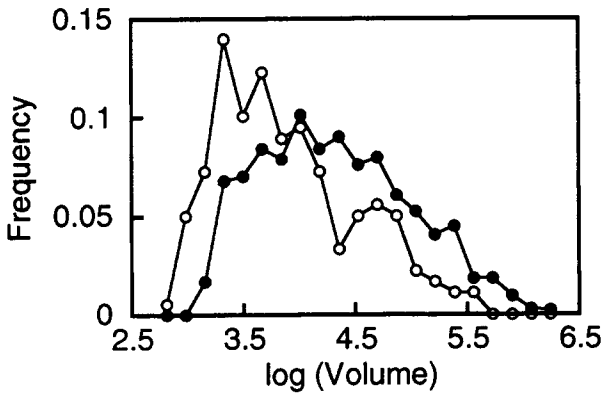

Figure 12. Estimated particle volume distribution in samples from batch embryo culture.

(○) day 10; (●) day 20.

shift toward larger particle sizes over time, continuous generation of smaller individuals remained active. Huang et al. (1993) reported that embryo concentration exceeded the inoculum concentration after 15 days of cultivation. By extending the characterization of the estimated volume distribution (or number distribution) schematized in Figure 10 to abnormal embryo classes, it may be possible to estimate the relative rate of callus formation, secondary embryogenesis, and so on. Once these are established for different environmental regimes, it may be possible to optimize the normal development by manipulating experimental conditions.

Conclusions

Currently, the image processing is performed off-line and manually; however, they can be further automatized. In the image acquisition step, the sample is properly diluted to avoid embryos touching each other. Sample size, an important factor in statistical evaluation of population properties, is discussed by Chi et al. (1994). Our results indicate that image processing can efficiently extract and separate the size and morphological information from the geometry of somatic embryos. Thus, an embryo can be quantitatively described by its morphology and size. By using the branch-and-bound algorithm, it becomes possible to select relevant features with optimization and statistical procedures. The contribution of these features provides the most significant information for class discrimination. The nearest neighbor algorithm allows us to automatically classify the developmental states of individual embryos and populations objectively, thus significantly reducing labor and bias. The image analysis technique in conjunction with statistical methods constitutes a pattern recognition system which can be extended to compare populations under different treatments. Since this system has the flexibility of investigating embryo populations at the individual and/or population levels, it provides a feasible means of effectively monitoring the development of embryo cultures and objectively evaluating the effect of experimental variables.

Acknowledgments

This work was supported in part by a grant from the National Science Foundation (BCS 9015817) and a grant from the Minnesota Supercomputer Institute. The technical assistance from Meheretab Abraha is gratefully acknowledged.

Notation

a = left abscissa of contour and real axis
$A(i)$ = ith order coefficient of Fourier series expansion of function $z(t)$
b = right abscissa of contour and real axis
j = imaginary number $\sqrt{-1}$
J = separability between classes
L = number of classes
m = total number of combinations for normalization
m_s = number of features in subset
M_1 = mean of medium-order Fourier coefficients
M_i = mean vector of embryo class ω_i
M_0 = mean vector of all embryo classes
N = number of points sampled in boundary
N' = number of total features
P_1 = frequency of a particle previously classified as state θ by a trainer, which was also classified as state θ by a pattern classifier
P_2 = frequency of a particle previously classified as state θ by a pattern classifier, which was also classified as state θ by a trainer
S_1 = reciprocal of the characteristic radius
S_2 = circularity factor
S_b = scattering matrix between classes
S_w = scattering matrix within class
u = phase of $A(-1)$
v = phase of $A(k)$
V = estimated volume of embryo
$x(t)$ = real part of function $z(t)$
X = feature factor
X_H = feature vector in training database H
$y(t)$ = imaginary part of function $z(t)$
y_+ = portion of the normalized $z(t)$ above real axis
y_- = portion of the normalized $z(t)$ below real axis
$z(t)$ = complex periodic function

Greek letters

ϵ = overall error between original and reconstructed contours
ω_i = class i

Literature Cited

Ahuja, M. R., ed., *Micropropagation of Woody Plants*, Kluwer Academic Publishers, Boston (1993).

Ammirato, P. V., "Organizational Events During Somatic Embryogenesis," *Plant Tissue and Cell Culture*, C. E. Green, ed., Alan R. Liss Inc., p. 57 (1987).

Ammirato, P. V., and D. J. Styer, "Strategies for Large-Scale Manipulation of Somatic Embryos in Suspension Culture," M. Zaitlin, P. Day, and A. Hollaender, eds., *Biotechnology in Plant Science: Relevance to Agriculture in the 1980s*, Academic Press, New York, p. 151 (1985).

Bajaj, Y. P. S., ed., *Biotechnology in Agriculture and Forestry: High-Tech and Micropropagation III and VI*, Springer-Verlag, Berlin (1992).

Borkird, C., J. H. Choi, and Z.-H. Jin, "Developmental Regulation of Embryonic Genes in Plants," *Proc. Acad. Sci. USA*, **85**, 6399 (1988).

Cazzulino, D. L., H. Pedersen, and C.-K. Chin, "Characterization of Plant Somatic Embryo Development Using Fourier Shape Analysis," *Ann. N.Y. Acad. Sci.*, **506**, 190 (1987).

Cazzulino, D. L., H. Pedersen, and C.-K. Chin, "Kinetics of Carrot Somatic Embryo Development in Suspension Culture," *Biotechnol. Bioeng.*, **35**, 781 (1990a).

Cazzulino, D. L., H. Pedersen, and C.-K. Chin, "Monitoring Somatic Embryo Development in Carrot Cell Cultures," *Dev. Ind. Microbiol.*, **31**, 285 (1990b).

Cazzulino, D. L., H. Pedersen, and C.-K. Chin, "Bioreactors and Image Analysis for Scale-Up and Plant Propagation," *Scale-Up and Automation in Plant Propagation*, I. K. Vasil, ed., Academic Press, New York, p. 147 (1991).

Chi, C.-M., H. Vits, E. J. Staba, T. J. Cooke, and W.-S. Hu, "Morphological Kinetics and Distribution in Somatic Embryo Cultures," *Biotech. Bioeng.*, **44**, 368 (1994).

Crouch, M. L., "Non-Zygotic Embryos of *Brassica napus* L Contain Embryo Specific Storage Proteins," *Planta*, **156**, 520 (1982).

Freeman, H., "On the Encoding of Arbitrary Geometric Configurations," *IRE Trans. Electron. Comput.*, **EC-10**, 260 (1961).

Fukunaga, K., *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, New York (1990).

Fukunaga, K., and P. M. Narendra, "A Branch and Bound Algorithm for Computing *k*-Nearest Neighbors," *IEEE Trans. Comput.*, **C-24**, 750 (1975).

Granlund, G. H., "Fourier Preprocessing for Hand Print Character Recognition," *IEEE Trans. Comput.*, **C-21**, 195 (1972).

Hamalainen, J. J., U. Kurten, and V. Kauppinen, "Classification of Plant Somatic Embryos by Computer Vision," *Biotechnol. Bioeng.*, **41**, 35 (1993).

Harrell, R. C., M. Bieniek, and D. J. Cantliffe, "Noninvasive Evaluation of Somatic Embryogenesis," *Biotechnol. Bioeng.*, **39**, 378 (1992).

Huang, L.-C., C.-M. Chi, H. Vits, E. J. Staba, T. J. Cooke, and W.-S. Hu, "Population and Biomass Kinetics in Fed-Batch Cultures of *Daucus Carota* L. Somatic Embryos," *Biotech. Bioeng.*, **41**, 811 (1993).

Kurtz, S. L., R. D. Hartman, and I. Y. E. Chu, "Current Methods of Commercial Micropropagation," *Scale-Up and Automation in Plant Propagation*, I. K. Vasil, ed., Academic Press, New York, p. 7 (1991).

Mitchell, O. R., and T. A. Grogan, "Global and Partial Shape Discrimination for Computer Vision," *Opt. Eng.*, **23**, 484 (1984).

Murashige, T., and F. Skoog, "A Revise Medium for Rapid Growth and Bioassays with Tobacco Tissue Cultures," *Physiol. Plant*, **15**, 473 (1962).

Packer, H. L., E. Keshavarz-Moore, M. D. Lilly, and C. R. Thomas, "Estimation of Cell Volume and Biomass of *Penicillium Chrysogenum* Using Image Analysis," *Biotechnol. Bioeng.*, **39**, 384 (1992).

Park, Y., and J. Sklansky, "Automated Design of Linear Tree Classifier," *Pattern Recognition*, **23**, 1393 (1990).

Preil, W., "Application of Bioreactors in Plant Propagation," *Micropropagation*, P. C. Debergh and R. H. Zimmerman, eds., Kluwer Academic Publishers, Amsterdam, p. 425 (1991).

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, New York (1989).

Raghavan, V., *Embryogenesis in Angiosperms*, Cambridge University Press, Cambridge (1986).

Redenbaugh, K., ed., *Synseeds-Application of Synthetic Seed to Crop Improvement*, CRC Press, Boca Raton (1993).

Redenbaugh, K., J. A. Fujii, and D. Slade, "Synthetic Seed Technology," *Scale-Up and Automation in Plant Propagation*, I. K. Vasil, ed., Academic Press, New York, p. 36 (1991).

Schalkoff, R., *Pattern Recognition: Statistical, Structural and Neural Approaches*, Wiley, New York (1992).

Schiavone, F. M., and T. J. Cooke, "A Geometric Analysis of Somatic Embryo Formation in Carrot Cell Culture," *Can. J. Bot.*, **63**, 1573 (1985).

Sethi, I. K., "A Fast Algorithm for Nearest Neighbor," *IEEE Trans. Syst. Man. and Cybern.*, **SMC-11**, 245 (1981).

Steeves, T. A., and I. M. Sussex, *Patterns in Plant Development*, Cambridge University Press, Cambridge (1989).

Veillon, F., "Study and Comparison of Certain Shape Measures," *Signal Process.*, **11**, 81 (1986).

Vits, H., W.-S. Hu, E. J. Staba, and T. J. Cooke, "Formulation of a Morphogenetic Model for Embryo Development in Dicotyledonous Plants," *J. Theor. Biol.*, **157**, 221 (1992).

Wallace, T. P., and P. A. Wintz, "An Efficient Three-Dimensional Aircraft Recognition Algorithm Using Normalized Fourier Descriptors," *Comput. Graphics and Image Process.*, **9**, 99 (1980).

Manuscript received July 7, 1993, and revision received Oct. 13, 1993.