Effects of hyperellipsoidal decision surfaces on image segmentation in artificial color

Jian Fu
Alabama A & M University
Computer Science Department
Normal, Alabama 35762
E-mail: jian.fu@aamu.edu

H. John Caulfield
Alabama A&M University Research Institute
P.O. Box 313
Normal, Alabama 35762

Dongsheng Wu
University of Alabama in Huntsville
Department of Mathematical Sciences
Huntsville, Alabama 35899

Trent Montgomery
Alabama A & M University
School of Engineering and Technology
Normal, Alabama 35762

Abstract. *Artificial color uses the projection of the spectrum into two or more broad, overlapping spectral bands to discriminate, pixel by pixel, among user-defined classes of objects. As initially practiced, it used a sequence of hyperspherical regions of the decision space to define class membership. Of course, a hypersphere is just a degenerate hyperellipsoid; thus, exploring the effect of loosening that degeneracy seemed appropriate. Initially, we use two-foci hyperellipsoids with a hyperellipsoidal distance metric to classify pixels with dramatic improvement in performance. We explore the work even further by allowing many foci and noting the effects of increased complexity of the decision surfaces. In the example case, three foci gave superior performance to one or two foci, but four added little improvement.* © *2010 SPIE and IS&T.*
[DOI: 10.1117/1.3377146]

1 Introduction

This work is part of our continuing effort to explore artificial color—the use of data from two or more broad, spectrally overlapping bands to discriminate pixels according to their membership in one or more user-defined sets.[1–4] The discrimination method of choice has been margin setting—a sequence of discriminants with extremely good generalization properties even if trained on very few fair samples.[5] Both methods have been published in numerous journals over the last several years.[6–8]

The idea of statistical pattern recognition is to regard the M measurements made as coordinates in an M-dimensional Euclidean hyperspace.[9] The goal becomes to erect some sort of M-dimensional surface that reliably classifies new inputs into their appropriate class.[10–14] Good surfaces are said to generalize well from the data on which they were trained. But how is good generalization to be obtained? The two factors most widely used for this purpose are margin and simplicity. Margin is a measure of the leeway for variation of new data from the training set data.[15] Simplicity reflects the idea that trying to use a complex surface for discrimination may wind up reflecting as much noise as real variations. The measure of complexity is the Vapnik–Chervonenkis (VC) dimension.[16,17] We want the lowest VC dimension that will do the job well. As one would expect, those factors are significantly counteracting, so trade-offs must be used to optimize any figure of merit.

Nevertheless, margin setting is a way to use discriminators. It is not itself a pattern recognition method. Figure 1 shows how this paper relates to pattern recognition.

The hyperspheres, often chosen as a tool in margin setting, may not be optimal if the data in the training set are far from hyperspherical. This paper studies hyperellipsoids and hyperspheres in margin setting. There are many other ways to do the pattern recognition steps, for instance, linear discriminants.[18,19] These simple classifiers can be used in margin setting, even for very complex problems. For the most part, only margin setting is being studied here and, thus, it is the background knowledge with which readers deal. Thus, most of the references are to it. Margin setting

Paper 09043RR received Apr. 3, 2009; revised manuscript received Jan. 22, 2010; accepted for publication Feb. 24, 2010; published online Apr. 8, 2010.

1017-9909/2010/19(2)/023003/10/$25.00 © 2010 SPIE and IS&T.

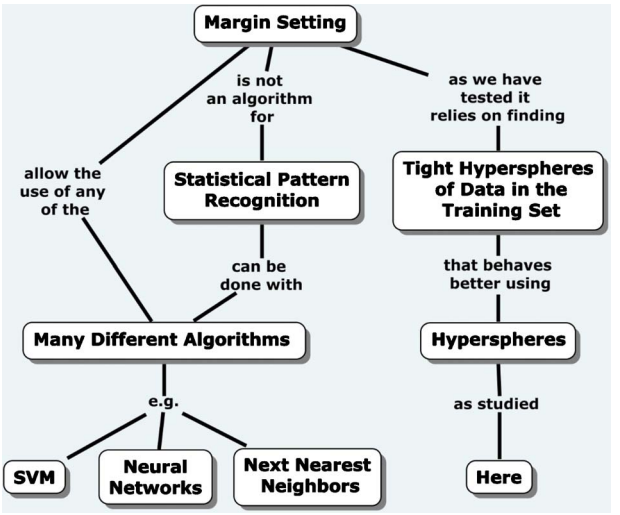

Fig. 1 Shows how this work relates to the field of statistical pattern recognition.

could be used with very advanced algorithms, such as SVDs (Support Vector Machine),[20] but that would be another paper. In this paper, we want simple-to-find discriminants for use in margin setting. Complex algorithms can be tested if we so choose, but that is not our choice right now.

The idea of margin setting is to use a simple discriminant that can classify many of the given samples correctly with a good margin. Then those samples not distinguished well would be subjected to further rounds of this procedure until some stopping criterion is reached. Big margins at every step make the corresponding decisions more likely to be correct but require more rounds of decisions than if the margins were smaller. Thus, again, some counterbalanced solution must be sought.

Because a linear discriminant has the lowest possible VC dimension[21,22] and because low VC dimensions can be expected to lead to good generalization, linear discriminants were used at every stage. This is equivalent to a hyperspherical decision surface in the data hyperspace.

This work explores a more complex situation in which the decision is made according to whether distance measure (the sum of the Euclidean distances from M foci) in that hyperspace is less than a threshold value when $M \geqslant 2$ (The $M=1$ case is the hypersphere case in our previous work.[2,6,7] In those cases, the distance measure is no longer quite so simple. The results both for artificial color and margin setting, in general, substantially improved when $M=2$. When $M>2$, in principle, there is no limit on how large M can be. But we expect the usefulness saturates quickly when M increases, therefore large M is of no practical value.

By studying the segmentation of various hard-to-segment images, we hope to arrive at some generally useful guidelines as to the number of foci likely to be most effective.

2 Decision Boundaries of Margin Setting

Margin setting uses supervised statistical pattern recognition algorithm. Let $\boldsymbol{V}$ denote a set of random vectors: $\boldsymbol{V}=\{\boldsymbol{v}_k: 1 \leqslant k \leqslant K\}$, where $\boldsymbol{v}_k$ is viewed as a potential prototype for the i'th separated sample subclass $S_i \subseteq \boldsymbol{S}$ of the training set $\boldsymbol{S}=\cup_{j=1}^{s} S_j$, including s many subclasses. Calculate the least Euclidean distance d_{kj} from $\boldsymbol{v}_k$ to every member of each subclass $S_j \subseteq \boldsymbol{S}$.

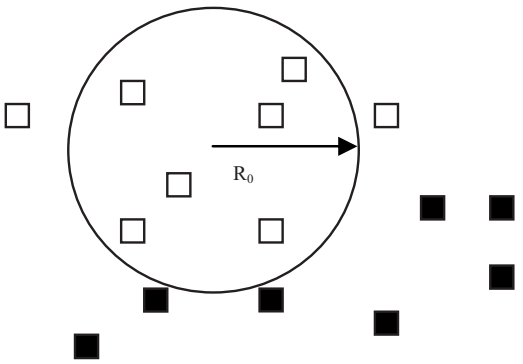

Fig. 2 Zero-margin radius R_0 is the largest radius if the hypersphere that contains only members of the target class.

$$d_{k,j} = \min\{\|\mathbf{v}_k - \mathbf{x}\| : \mathbf{x} \in S_j\} \tag{1}$$

Find the nearest member of another class $(S_j, j \neq i)$, $R_{k,0}=\min\{d_{k,j}, j \neq i\}$. That distance, $R_{k,0}$, is called the zero-margin radius. Consider the problem of classifying two types of patterns: white and black squares as shown in Fig. 2. The figure of merit F of the potential prototype is given by the cardinality of S inside the $R_{k,0}$ around that prototype. For instance, $F=6$ in Fig. 2. Then new potential prototypes are chosen based on the old ones. A fixed number of these prototypes is employed. To obtain the next generation of prototypes, we stochastically selected from the previous generation using some elitist method. The probability of selection of any prototype is proportional to the cardinality of data inside the hypersphere at that prototype. Those prototypes are mutated (perturbed) and placed into the set of prototypes for the next-generation distribution function governed by F's of the various prototypes. Then the prototype is mutated by choosing a perturbation from a normal distribution centered at the selected prototype.[23] This continues until no improvement in F is achieved. The highest scoring prototype is selected.

That process is repeated for each new stage. Each stage uses only those samples not classified in earlier stages in $\boldsymbol{S}$.

In classification, new data are tested on the first-stage classifiers. If those classifiers indicate a class, then it is accepted. If not, then we go to the second-stage classifiers, and so forth through all the predefined stages. $T=1$ when the new data point can be classified in the target class, or $T=0$ when the point can be classified in other classes or unclassified.

Now we shrink the radius in Fig. 2, $r=R_1$, and C_2 is the new decision surface, as shown in Fig. 3. The room between C_1 and C_2 is defined as margin. Margin is a measure of our ability to discriminate even in the presence of such "unfriendly" variations. New data (not in the training set) may vary in any direction from those in the training set. It is prudent to assume that some of those variations will carry the discriminants closer to the decision boundary than the training set points. Higher margin provides more room for new points to differ from those trained on without being misclassified. Zero margin means that some small varia-

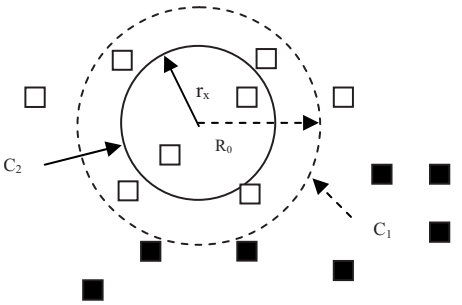

Fig. 3 Making the radius smaller than the zero margin radius creates more margin of safety in decisions made by that surface.

tions from the training set points near the decision surface will produce those misclassifications. The choice of margin has one obvious effect. Larger margins make fewer misclassification errors but leave more data unclassifiable. That trade-off may have different consequences in different circumstances. This method is strongly dependent on the actual distribution of training set data in the hyperspace. If that distribution is roughly hyperspherical, then adding foci accomplishes nothing. If the data occupy a very asymmetrical volume, then the improvement will be dramatic.

For the setting of the margin, we set a χ-percent margin by setting

$$r_x = (1 - 0.01\chi)R_{k,0}. \tag{2}$$

The zero-margin radius $R_{k,0}$ is not set by the user. It is found during the training period. It is the largest R that contains no members of the training set belonging to a call not intended for classification into the same set as all the other samples inside it. A 10% margin means that the decision is made at a radius that is $0.9R_{k,0}$ and so forth.

The ability to choose our margin is one we think useful, and we also suspect that exploratory analysis may yield the best results using the user's judgment, not ours. As the margin is increased, fewer false positives occur, but the number of iterations required increases.

Figures 4 and 5 present some results obtained in the image of a flagpole that appears to the human eye as very close in color to the ground. As the margin increased, the false alarm rate decreased until eventually, only flagpole pixels remained.

Fig. 4 An outdoor scene photographed with an ordinary color digital camera.

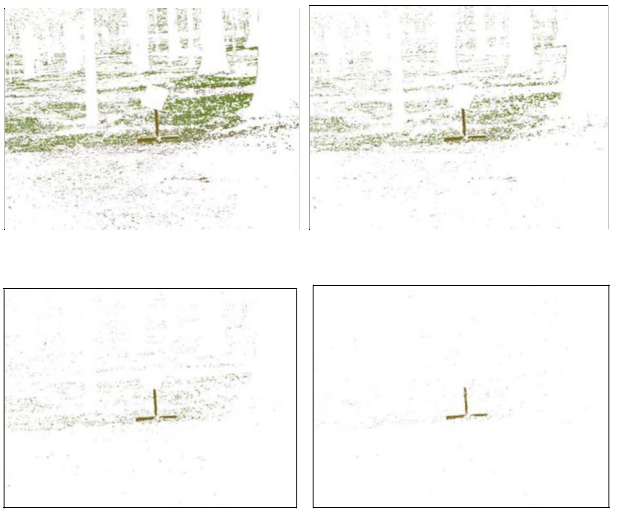

Fig. 5 Segmented results with different margins. The discrimination improves with margin size as all theories suggest. Note that the segmentation becomes more and more perfect as the margin increases.

In this paper, we extend the previous work by using a general hyperellipsoid with M foci. Going from a hypersphere to a hyperellipsoid involves only one step. It requires the distance metric to be not a single Euclidean distance, but rather the sum of the Euclidean distances of a data point from each focus of the hyperellipse.

Let $\boldsymbol{V}=\{\boldsymbol{V}_k: 1 \leqslant k \leqslant K\}$ be the foci set of class S_i, where $\boldsymbol{V}_k=\{\boldsymbol{v}_{m,k}: 1 \leqslant m \leqslant M\}$ represents the k'th foci (of S_i) in $\boldsymbol{V}$, and where $\boldsymbol{v}_{m,k}$ is a random vector. Instead of Eq. (1), calculate the least Euclidean distance $d_{k,j}$ from $\boldsymbol{V}_k$ to every member of each subclass $S_j \subseteq \mathbf{S}$ by

$$d_{k,j} = \min\left\{\sum_{m=1}^{M} \|v_{m,k} - x\| : x \in S_j\right\}. \tag{3}$$

The hyperellipse can have as many as M (the dimension of the hyperspace) and as few as one focus. The single-focus hyperellipsoid ($M=1$) is, of course, simply a hypersphere.

The improved multifoci margin setting is represented as follows:

Target class $S_1=\{\vec{x}_{11}, \ldots, \vec{x}_{1n_1}\}$

Sample set $S_j=\{\vec{x}_{j1}, \ldots, \vec{x}_{jn_j}\}$, where $j=1, \ldots, s$. Denote $S=\cup_{j=1}^{s} S_j$ and $n=\Sigma_{j=1}^{s} n_j$.

Notation:

$\#(A)$: the number of elements in set A

Ω: foci triplet set

K: designed number of foci in Ω

N_Ω: designed maximal number of steps to find K foci in Ω

N_Λ: designed maximal number of elements in Λ

Algorithm:

1. $\Lambda=\phi$ (the empty set).
2. Initialization: $i=1$, $t=1$, $\Omega=\phi$ (the empty set), $F=0$.
3. Generate an M-foci candidate randomly, that is, $V=(\vec{v}_1, \ldots, \vec{v}_M)$.

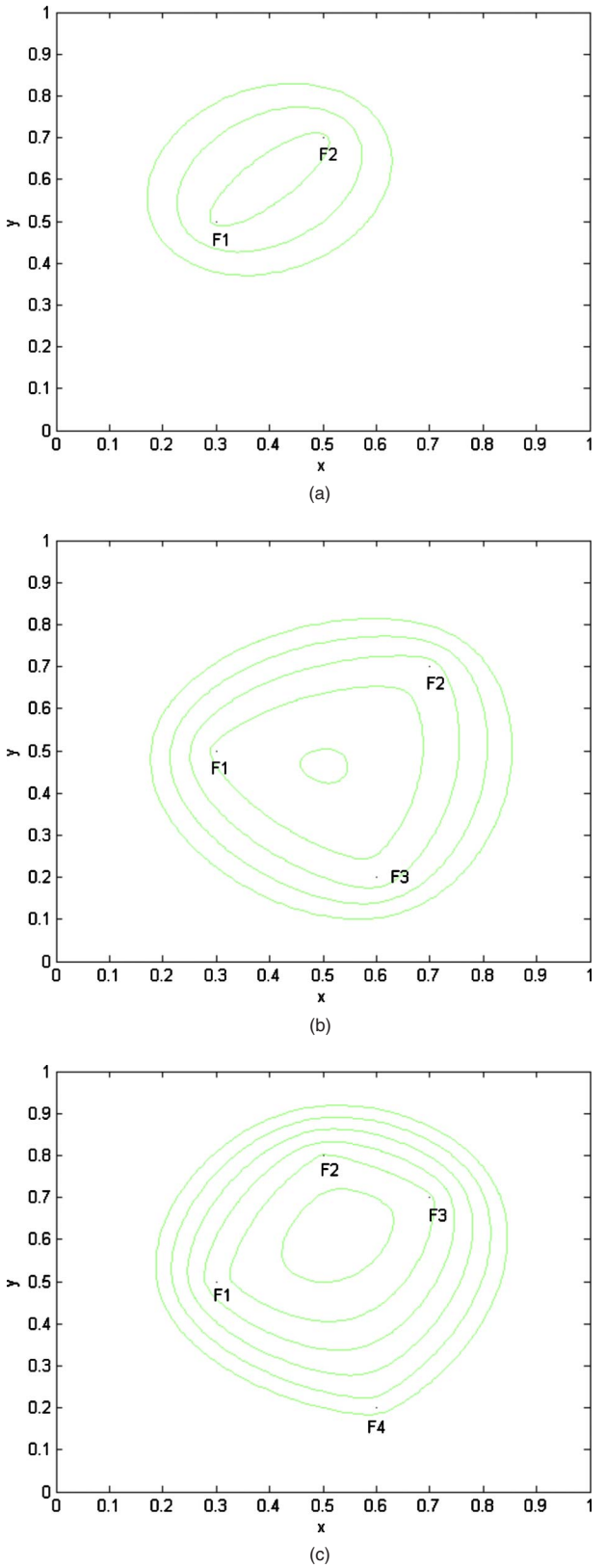

Fig. 6 Two-dimensional decision boundaries of multifoci margin setting. F_1, F_2, F_3, and F_4 are foci: (a) Two foci (The boundaries are ellipsoids.) (b) three foci, and (c) four foci. The three- and four-foci cases yield hyperellipsoids. Of course, the prior work was done with a single focus and produced hyperspheres.

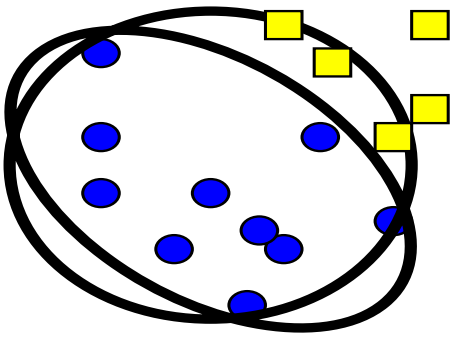

Fig. 7 Clearly, the ellipse can enclose the target pixels while leaving significantly less room for a member of another class to appear inside it in comparison to the circle. The tighter fit is thus likely to be less susceptible to false positives. The yellow pixels correspond to a different class. They show why the ellipse is better in this case. (Color online only.)

4. Compute the sum of the distances from the foci to every member of S and find

$$d_k = \min_{1 \leq l \leq n_k} \sum_{j=1}^{M} \|\vec{x}_{kl} - \vec{v}_j\|, \quad k = 1, \ldots, s.$$

5. If $d_1 = \min_{1 \leq k \leq s} d_k$, $V_t = V$, and define modulus: $R_t = \min_{k \neq 1} d_k$ and figure of merit F_t: $\#(S_1)$ in the region $\{\vec{y}: \Sigma_{j=1}^{M} \|\vec{y} - \vec{v}_{tj}\| \leq R_t\}$, $\Omega = \Omega \cup \{(V_t, R_t, F_t)\}$, $t = t+1$; else, $i = i+1$.

6. If $i > N_\Omega$, stop. Else if $\#(\Omega) < K$, go to step 2.

7. If $F < F_{t_0} := \max_{1 \leq t \leq K} F_t$, $W = V_{t_0}$, $R = R_{t_0}$, $F = F_{t_0}$; else, go to step 9.

8. Calculate $f_t = F_t / (\Sigma_{l=1}^{K} F_l)$. Choose a random number $Y (0 \leq Y \leq 1)$. Pick $(V_m, R_m, F_m) \in \Omega$ if

$$\sum_{l=1}^{m-1} f_l < Y \leq \sum_{l=1}^{m} f_l.$$

9. Randomly perturb V_m K times to generate K many new foci triplets $\{(V_{m1}, R_{m1}, F_{m1}), \ldots, (V_{mK}, R_{mK}, F_{mK})\}$, and mutate Ω: $V_t = V_{mt}$ $R_t = R_{mt}$ $F_t = F_{mt}$ for $t = 1, \ldots, K$, go to step 6.

10. For the triplet (W, R, F), assign a modulus $r = (1 - \chi)R$. If $\chi = 0$, it is the zero-margin modulus. Fix χ at some value, remove from S all the points in $T = \{\vec{y}: \Sigma_{j=1}^{M} \|\vec{y} - \vec{w}_j\| \leq r\} \cap S_1$, update $S_1 = S_1 \backslash T$ and $S = S \backslash T$, and store the triplet (W, r, T) in Λ.

11. If $S_1 = \phi$ (the empty set) or $\#(\Lambda) \geq N_\Lambda$, output Λ and stop; else, go to step 1.

In 2-D hyperspace, the boundaries of different margin for single focus $(M = 1)$ are a series of concentric circles. Figure 6 shows the boundaries of different margin for two, three, and four foci. When $M > 2$, the foci do not have to be inside of the enclosed boundary.

The sum of radii in 2-D generalizes to the sum of distance in 3-D. No claim is made that there is an optimum metric for distance. We chose the Euclidean norm because it is, by far, the most commonly used for any particular norm, and because the choice of norms may not impact the

Fig. 8 This is the original image we used to test segmentation as a function of number of foci of the discriminating hyperellipsoids.

Fig. 9 This hand-segmented frog image was used as the ideal with which all of the automatically color-segmented images were compared.

general effect we are seeking. It may be true that extreme cases, such as the infinity norm or the Manhattan norm, may give significantly improved or degraded performance, but still the reasoning put forth here will still apply.

The spectral classifier is drawn from a search procedure that seeks to enclose a large set of samples of one class without including any samples of the other set(s). If the training set data are not essentially hyperspherically disposed, then a hyperellipsoid will give a better fit.[24,25] This is illustrated crudely in Fig. 7.

Noise, if it has uniform expected value everywhere in the discrimination space, will be expected to be proportional to the volume of the discrimination surfaces (ellipse and circle in this 2-D space). The ellipse has the circle as a special case; thus, it will never give worse results than the circle. And, if the ellipse can be significantly smaller in volume than the circle, the ellipse will offer greater protection against noise.[26]

There is no algorithm for finding the maximum number of useful foci because that number is data dependent. Data-dependent algorithms are used often—iterations, optimization, iterative improvement—and they have the same problem and the same difficulty of finding a general theory. This is the subject of a wonderful book by J. Tukey entitled *Exploratory Data Analysis*.[27] Sometimes we must be content with a nonalgorithmic explanation.

When we do these calculations with a variable number of foci, we must start the algorithm from the beginning. The initial centers are computed in such a manner that there is a great variation between the centers at each stage. It should not be surprising if the results show variability and even an occasional situation in which a really good classifier using N foci gives results that are better than the ones obtained using $N+1$ foci. The results we show are raw and honestly gotten and shown. Sometimes we get unlucky. In the long term, the trends are as shown and explained.

3 Experiments and Experimental Setup

3.1 *Test Problem*

We chose a test problem that is very difficult, so we can see the improvements, if any, produced by increasing the number of foci of the hyperellipsoids systematically. Figure 8 shows the test image.

The frog's camouflage in the visible is so good that we expect to fail to separate the frog from its surround perfectly, because color is simply inadequate for that task. To make it possible to quantify our effects, we segmented the image by eye as best we could, see Fig. 9. This became our standard, allowing us to ask how many pixels classified as frog were truly frog pixels and, conversely, how many truly frog pixels were not identified as such.

3.2 *Training Procedure*

We took 15 pixels each from the frog, the dark green leaves, the light green leaves, the dark sticks, and shadow/sticks as the training set. The task was to separate frog pixels from pixels belonging to one of the other three classes on the basis of their RGB content. At each stage in the training, we found the foci of and what we call the largest modulus hyperellipsoids that contained as many frog pixels (volume elements in the 3D RGB hyperspace) as possible without containing any pixels from another class. The ellipsoid is defined as the locus of points such that the sum of their Euclidean distances from the foci is equal to the modulus. For example, a hypersphere has a single focus (center) and a modulus r, where r is its radius. In fact, it is a hypersphere. A two-focus hyperellipsoid uses

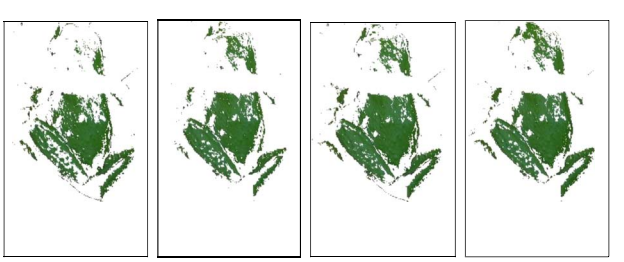

Fig. 10 Increasing the number of foci of the classifying hyperellipsoids had a quite noticeable effect on the frog segmentation as can be seen clearly here. But there is little subjective improvement in going from three to four foci.

Table 1 The ratios of image pixels suffering the three possible fates as a function of number of foci. The number of pixels in each fate is normalized to the baseline for better understanding.

Center	Correct ratio (inside)	Incorrect ratio (outside)	Unclassified ratio
1	1	1	1
2	1.11	0.41	0.89
3	1.23	0.70	0.77
4	1.24	1.08	0.76

as its modulus the sum of the distances from the two foci, and so forth. That maximum modulus is called the zero-margin modulus, because even a minutely larger one would take in at least one pixel from a nonfrog data point. Then we set a margin by allowing for a more conservative modulus than the zero-margin (R_0) by setting the modulus to a value less than the zero-margin value. In our experiments, we arbitrarily chose a margin of R_0. All pixels within the hyperellipsoid at any stage were removed from the training set (because they were already recognized). Then the procedure was applied again to the now-reduced training set. This continued until some stopping condition was met—in this case, we fixed the number of rounds at 4. The result was a set of hyperellipsoids such that any RGB values falling within any one of them was declared to be from a frog pixel. For display, we left the putative frog pixels unchanged and set all other pixels to 0.

3.3 *Test Results*

The subjective effect of increasing the number of foci can be viewed in Fig. 10. Using the comparison standard of Fig. 10 allowed us to assess these effects quantitatively as well. Table 1 shows the fate of image pixels for various numbers of foci. Some are correctly classified as belonging to the frog; some pixels outside the frog are attributed to the frog; and other pixels are simply not classified.

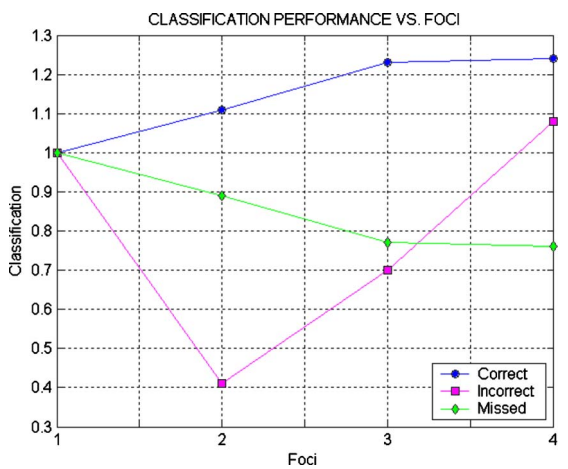

Fig. 11 This chart allows easy comparison of various pixels classification abilities as a function of the number of foci.

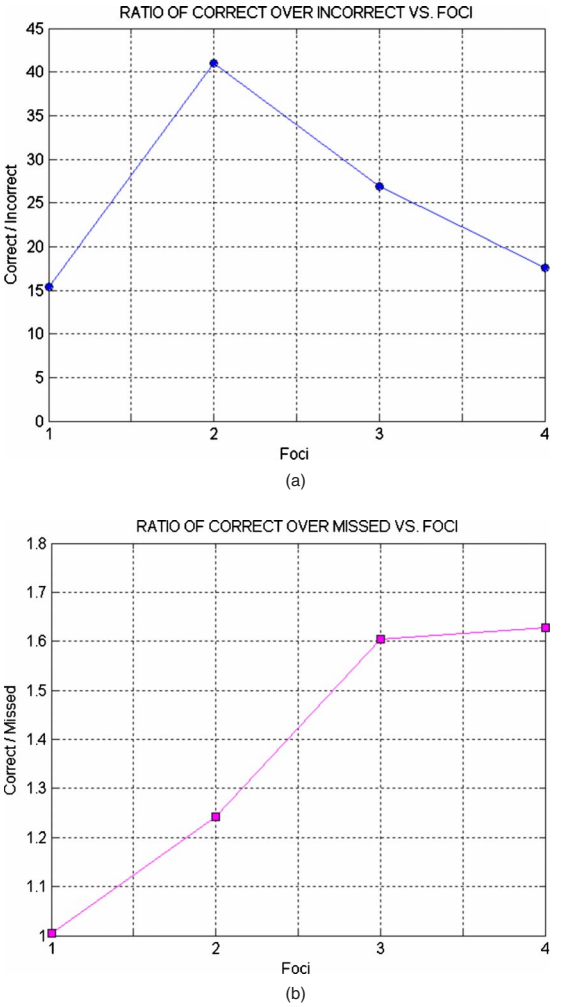

Fig. 12 (a) The ratio of pixels inside (Correct) over outside (Incorrect) of the frog body versus the number of foci and (b) The ratio of pixels inside (Correct) over unclassified inside (Missed) of the frog body versus the number of foci.

We can now plot these three ratios as functions of the number of foci to obtain Fig. 11. The data points used are not totally predictable because the evolution process used to calculate the parameters of a hypersphere involve stochastic effects that, by definition, are not repeatable. It is dangerous to infer minutiae in the data represented in these curves. What can be inferred is the general trends.

As expected, the number of correctly classified pixels increases with the number of foci but it appears to saturate at 3. The price of obtaining more correct classifications, however, appears to be obtaining more incorrect ones (pixels outside the frog identified as due to the frog). Increasing the number of foci appears to decrease the unclassified pixel rate monotonically, but again saturation appears to take place at ~3.

Another way to view these results is in terms of ratios. For example, Fig. 12(a) shows how the ratio of correct to erroneous frog pixel assignments varied with M. Figure 11 helps understand this result. The two-focus hyperellipsoid accomplished the indicated peak at the price of leaving sub-

Table 2 Despite the built-in stochastic elements and the problem-dependence, the results shown here seem to indicate that the method of multiple foci may have the expected superiority to other methods.

	Margin setting (4 foci)	Mahalanobis distance	Support vector machine	Maximum likelihood	Neural net	Minimum distance
Correct	1	0.99	0.92	0.83	0.72	0.72
Incorrect	1	2.49	0.19	0.34	0.01	2.52
Unclassified	1	0.69	0.96	1.18	1.58	1.54

stantial numbers of pixels unclassified. Figure 12(b) shows that after $M=3$, the ratio of correct to unclassified is much higher than for $M=2$ and appears to saturate at $M=3-4$.

Readers should recognize that the efficacy of this method is strongly dependent on the following:

1. The actual distribution of training set data in the hyperspace. If that distribution is roughly hyperspherical, then adding foci accomplishes nothing. If the data occupy a very asymmetrical volume, then the improvement will be dramatic.
2. The results we obtain are not necessarily repeatable in detail. There are random choices at every stage, and the results are therefore not expected to be the same every time or to show simple monotonic behavior as a function of number of foci.

The data shown in Table 2 reflect the expected variability associated with this approach but also show the expected improvements over other methods of attacking the same problem.

We ran three completely different analyses of the frog data and the results are shown in Table 3. The frog image was selected by hand, and all other pixels were called "background." The absolute value for the number of frog pixels is 22,194, as shown under the frog heading. We then calculated the number of frog pixels found and what the fraction of that total was. The fractional numbers in each box are important in control theory; this would be viewed as adaptive convergence. Whatever is not correct in this cycle remains present for possible correction in the next. Thus, in the end, roughly the same value is reached regardless of what starting conditions prevailed. Column 2, for instance, has a mean of 75.9 with a standard deviation of 0.2974.

The results from the three independent runs are exactly what one would expect. The results are very stable but, of course, slightly variable. The true frog and true background fractions are very stable. The error numbers look quite variable, but they are not. The number of errors itself is small and thus subject to what appear to be large variations. But that is exactly what is predicted. Large variations in small numbers are still small.

It is dangerous to draw general conclusions from a single example; therefore, we repeated this process for some other hard-to-discriminate problems.

Figure 13 shows an image of peppers. We seek to find one particular pepper using only a few samples from the huge number present.

Figure 14 shows the results with 1, 2, 3, and 4 foci. Once again, improvement is easy to see from the first three images and hard to find in the fourth. And, once again, saturation of improvement appears around the three-foci case.

Different classes in the same image may require different numbers of foci. Consider the target image of Fig. 15. The long green okra produced the results shown on Fig. 16.

For most purposes, such as identification, the single focus (hypersphere) is sufficient. But in the case of the orange-colored apricot, the results shown in Fig. 17 suggest easily noticeable improvement with number of foci, up to ~3, where saturation seems evident.

Seeking to be honest and avoiding the choice of images suitable for the hyperellipsoid cells, we showed some were situations where the result was not much of an improvement. We went on to say that the usefulness of going from hypersphere to hyperellipse depends on the distribution of the training set data in the decision hyperspace. If the data are essentially spherically disposed, then hyperellipsoids are no better than a hypersphere. But, if the distribution of training set data is quite aspherical, then hyperellipsoids were ideal.[28] Going from a hypersphere to a hyperellipsoid is probably not substantial.[29] Furthermore, it may require fewer hyperellipsoids than hyperspheres, and there are no general remarks to be made because of the previously noted problem dependence of the results.

4 Conclusions

Segmentation by artificial color methods can be very difficult as in this case, where there is little if any color distinc-

Table 3 Tabulation of results from three totally independent runs made on the frog image using two-foci hyperellipsoid voxels in the three-dimensional decision space as decision boundaries with zero margin.

Run no.	Frog (22,194 pixels)	Background (59,280 pixels)	Error
1	16895 (76.13%)	58649 (98.74%)	5930
2	16888 (76.09%)	57923 (97.01%)	6663
3	16751 (75.48%)	58572 (98.81%)	6152

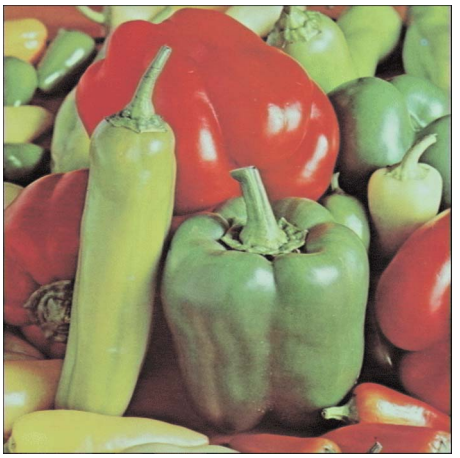

Fig. 13 Three different kinds of peppers. We used only 15 samples of each to try to segment out only one kind.

tion between classes and the number of samples from each class is very limited. The use of hyperellipsoidal classifiers with various numbers of foci offers a useful free variable. We can draw no universal conclusions from the limited number of experiments we did, but the general trends obtained and presented in Fig. 11 conform with our general expectation. For very hard problems, increasing the number of foci of the classifying hyperellipsoids may allow more correct classification but may do so at the price of producing more misclassifications as well. The reservoir of unclassified pixels is diminished to accommodate those effects,

Fig. 14 Improvement with number of foci is most noticeable with the almost saturated top of the pepper. Improvements are not uniformly distributed across the image. The circles show the same region with different numbers of foci. The arrow points from the single-focus image to the four-foci image. They can now bee seen to have produced noticeable improvement.

Fig. 15 In this simple image, there are three classes of vegetables that require different treatment for optimization.

and of course, we do not expect ever-increasing improvements. In this case, little change was made in going from three foci to four.

More generally, it appears that increasing the number of foci in the hyperellipse can sometimes be quite helpful, but there also appear to be prices in terms of other errors if we go to too many foci. Thus, each problem needs to be explored to determine which approach is optimal for a given data set and a given figure of merit.

This technique has been applied to and is being employed in many tasks, such as recognizing an object by its shape,[30] RGB values,[2,6] texture, and polarization.[31] It is also used for recognition in many different settings, e.g., iris (the colored part of the eye),[4] hyperspectral images,[30] easy-to-manipulate artificial problems, pills and capsules, passport covers, and Eurodollars. In all cases, a small number of samples is capable of achieving excellent discrimination.

There are many directions one might go from here, such as the following:

1. Explore different figures of merit. A reader suggested one very interesting approach, called "the normalized probabilistic rand,"[32] which was totally new to us but seems to have been rather widely used for evaluating image differences.
2. Study the effects of normalization, so all the hyperellipsoids become hyperspheres can be compared to

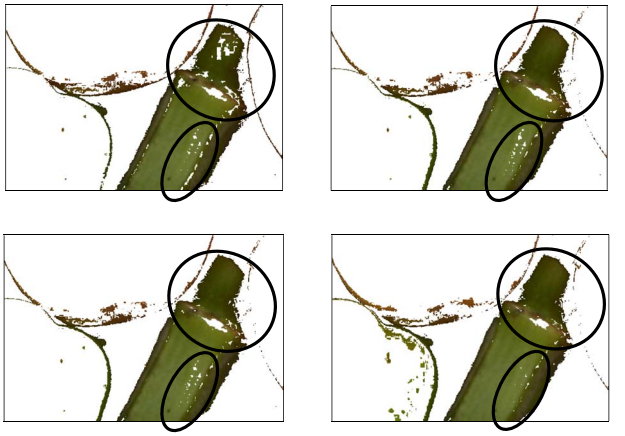

Fig. 16 This target class shows small but noticeable improvement with each increase in number of foci.

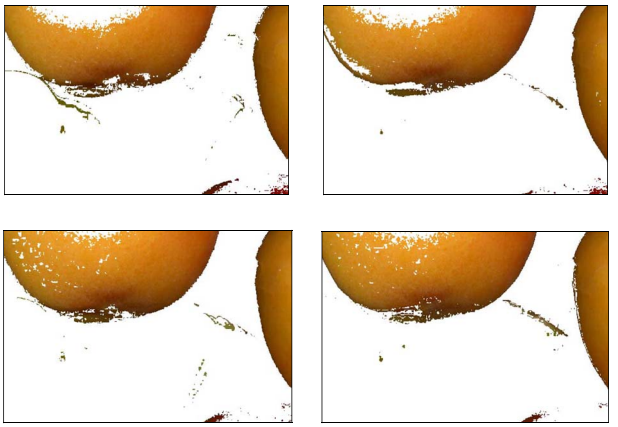

Fig. 17 Noticeable improvement with the number of foci.

treating the same problem with the unnormalized data used in this paper.

3. Look at the various distance metrics to find out if one works better than the other.
4. Compare a complex classifier (e.g., the support vector machines (SVM)) and a simple discriminant (e.g., a linear discriminant) for the same problem.

Acknowledgment

This work is supported by the Air Force Office of Scientific Research (AFOSR) under Award No. FA9550-07-1-0061.

References

1. H. J. Caulfield, "Artificial color," *Neurocomputing* **51**, 463–465 (2003).
2. H. J. Caulfield, J. Fu, and S. M. Yoo, "Artificial color image logic," *Inf. Sci.* **167**, 1–7 (2004).
3. J. Fu, H. J. Caulfield, and S. R. Pulusani, "Artificial color vision: a preliminary study," *J. Electron. Imaging* **13**(3), 553–558 (2004).
4. J. Fu, H. J. Caulfield, S. M. Yoo, and V. Atluri, "Use of artificial color filtering to improve iris recognition and searching," *Pattern Recogn. Lett.* **26**, 2244–2251 (2005).
5. H. J. Caulfield and K. Heidary, "Exploring margin setting for good generalization in multiple class discrimination," *Pattern Recogn.* **38**(8), 1225–1238 (2005).
6. J. Fu, H. J. Caulfield, and A. Bandyopadhyay, "Pairing mathematical morphology with artificial color to extract targets form clutter," *J. Imaging Sci. Technol.* **51**(2), 148–154 (2007).
7. J. Fu, H. J. Caulfield, and A. J. Bond, "Artificial and biological color band design as spectral compression," *Image Vis. Comput.* **23**, 761–766 (2005).
8. J. Fu, H. J. Caulfield, and T. Mizell, "Applying median filtering with artificial color," *J. Imaging Sci. Technol.* **49**(5), 498–504 (2005).
9. A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(1), 4–37 (2000).
10. T. Ho and M. Basu, "Complexity measures of suprised classification problems," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(3), 289–300 (2002).
11. J. Kim, J. R. Yu, and S. H. Kim, "Learning of prototypes and decision boundaries for a verification problem having only positive samples," *Pattern Recogn. Lett.* **17**(7), 691–697 (1996).
12. C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(4), 388–400 (1993).
13. K. Tumer and J. Ghosh, "Analysis of decision boundaries in linearly combined neural classifiers," *Pattern Recogn.* **29**(2), 341–348 (1996).
14. J. Zhang and Y. Liu, "SVM decision boundary based discriminative subspace induction," *Pattern Recogn.* **38**(10), 1746–1758 (2005).
15. G. Chechik, G. Heitz, G. Elidan, P. Abbeel, and D. Koller, "Max-margin classification of data with absent features," *J. Mach. Learn. Res.* **9**, 1–21 (2008).
16. A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, "Learnability and the Vapnik-Chervonenkis dimension," *J. ACM* **36**(4), 929–865 (1989).
17. V. Vapnik and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theor. Probab. Appl.* **16**(2), 264–280 (1971).
18. W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical performance analysis of linear discriminant classifiers," in *Proc. of IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 164–169 (1998).
19. B. Ji, C. I. Chang, J. L. Jensen, and J. O. Jensen, "Unsupervised constrained linear Fisher's discriminant analysis for hyperspectral image classification," *Proc. SPIE* **5546**, 344–353 (2004).
20. J. Zhang and Y. Wang, "A rough margin based support vector machine," *Inf. Sci.* **178**(9), 2204–2214 (2008).
21. S. Floyd and M. Warmuth, "Sample compression, learnability, and the Vapnik-Chervonenkis dimension," *Mach. Learn.* **21**(3), 269–304 (1995).
22. V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.* **10**(5), 988–999 (1999).
23. J. D. Farmer, N. H. Packard, and A. S. Perelson, "The immune system, adaptation and machine learning," *Physica D* **2**, 187–204 (1986).
24. C. Jou, Q. Z. Wu, S. C. Tsay, Y. J. Tsay, and S. S. Yu, "A hyperellipsoid neural network for pattern classification," in *Proc. of IEEE Int. Symp. on Circuits and Systems*, Vol. 2, pp. 1176–1179 (1991).
25. Q. Zhu, Y. Cai, and L. Liu, "A multiple hyper-ellipsoidal subclass model for an evolutionary classifier," *Pattern Recogn.* **34**(3), 547–560 (2001).
26. Z. Deng, F. L. Chung, and S. Wang, "A new minimax probability based classifier using fuzzy hyper-ellipsoid," *Proc. of Int. Joint Conf. on Neural Networks*, Orlando, pp. 2385–2390, International Neural Network Society, Orlando, FL (2007).
27. J. W. Tukey, *Exploratory data analysis*, Addison-Wesley, Reading, MA (1977).
28. N. Tsumura, K. Itoh, and Y. Ichioka, "Reliable classification by double hyperspheres in pattern vector space," *Pattern Recogn.* **28**(10), 1621–1626 (1995).
29. G. L. Peterson and B. T. McBride, "The importance of generalizability for anomaly detection," *Knowledge Inf. Syst.* **14**(3), 377–392 (2008).
30. H. J. Caulfield and J. Fu, "Simple algorithms to process the spectral and spatial information in a HyperSpectral Image (HSI)," *Pattern Recogn. Lett.* (submitted).
31. J. Fu and H. J. Caulfield, "Applying color discrimination to polarization discrimination in images," *Opt. Commun.* **272**, 362–366 (2007).
32. R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(6), 929–944 (2007).

Jian Fu received his PhD in computer science and engineering in 2005 from University of Alabama in Huntsville and an MS in computer sciences and another MS in physics in 1998 and 1996, respectively, from Alabama A & M University. He is an associate professor in the Department of Computer Sciences at Alabama A & M University. His research interests include pattern recognition, image processing, and computer vision.

H. John Caulfield is a multiple awardee of SPIE (Gabor, Fellow, President's, Governors'). Among his publications are 11 books, 22 book chapters, over 200 refereed journal papers, 29 U.S, patents, and numerous articles. His primary interests are signal processing, soft computing, holography, and optical metrology. He has been CTO of six companies of various sizes. He also serves as an assistant sheep farmer on Far Out Farm in Tennessee.

Dongsheng Wu received his PhD in mathematics from Michigan State University in 2006. He is currently an assistant professor in the Department of Mathematical Sciences at the University of Alabama in Huntsville. His research interests include stochastic processes and random fields, stochastic partial differential equations, statistical properties of Gaussian random fields, random fractal, and high-accuracy finite element analysis.

Trent Montgomery is a professor in electrical engineering and Dean of the School of Engineering and Technology at Alabama A & M University. His research interests include nonlinear electronics, digital systems, computer organization, and microprocessor architectures. He received his PhD in electrical engineering in 1976 from University of Texas at Austin.