MULTIPLE SHAPE MODELS FOR
SIMULTANEOUS OBJECT CLASSIFICATION AND SEGMENTATION

By

Federico Lecumberry
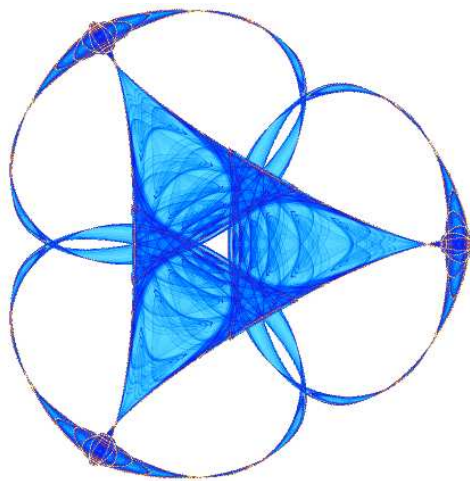
Álvaro Pardo

and

Guillermo Sapiro

INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **FEB 2009**	2. REPORT TYPE	3. DATES COVERED **00-00-2009 to 00-00-2009**
4. TITLE AND SUBTITLE **Multiple Shape Models for Simultaneous Object Classification and Segmentation**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Minnesota,Institute for Mathematics and Its Applications,Minneapolis,MN,55455-0436**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT **see report**

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT **Same as Report (SAR)**	18. NUMBER OF PAGES **5**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

MULTIPLE SHAPE MODELS FOR SIMULTANEOUS OBJECT CLASSIFICATION AND SEGMENTATION

Federico Lecumberry   *Álvaro Pardo*   *Guillermo Sapiro*

IIE, Universidad de la República   DIE, Universidad Católica   ECE, University of Minnesota
fefo@fing.edu.uy   apardo@ucu.edu.uy   guille@umn.edu

ABSTRACT

Shape models (SMs), capturing the common features of a set of training shapes, represent a new incoming object based on its projection onto the corresponding model. Given a set of learned SMs representing different objects, and an image with a new shape, this work introduces a joint classification-segmentation framework with a twofold goal. First, to automatically select the SM that best represents the object, and second, to accurately segment the image taking into account both the image information and the features and variations learned from the on-line selected model. A new energy functional is introduced that simultaneously accomplishes both goals. Model selection is performed based on a shape similarity measure, determining which model to use at each iteration of the steepest descent minimization, allowing for model switching and adaptation to the data. High-order SMs are used in order to deal with very similar object classes and natural variability within them. The presentation of the framework is complemented with examples for the difficult task of simultaneously classifying and segmenting closely related shapes, stages of human activities, in images with severe occlusions.

Index Terms— Shape priors, image segmentation, object modeling.

1. INTRODUCTION

Object segmentation is one of the most fundamental tasks in image processing, still lacking a completely automatic solution. The main idea is to find a set of features that describes and discriminates the object of interest from the rest of the image. Pixel color is a low level feature that can be used as such descriptor, although its discrimination capacity is often insufficient in real images. Using shape as a high level feature is a common approach to augment such low level features.

The shape of the desired object is added as a descriptor, constraining the set of possible solutions to regions of the image that simultaneously "match" this shape and the low level features (intensity, edges, etc.). The most common way to add this shape prior information is in the form of a weighted linear combination of functionals concerning, on one hand, the low level features and, on the other hand, the shape priors. This leads to a minimization problem where the solution is a compromise between the shape of the final contour and the information constrained by the image. The actual minimization techniques vary, including gradient descent methods [1, 2, 3, 4, 5] and graph-cuts [6]. The used shape representations also vary, including signed distance functions (SDF) [1, 2, 3, 6, 5], quadratic splines [7], characteristic functions [4], and landmark points [8].

When M different objects can appear in an image, a single shape prior (model) is not sufficient, and multiple shape priors must be considered. A possible, but not elegant, approach is to run the process with each one of the shape priors separately, and then choose the best solution. In [5, 6] the authors define M possible labels for each pixel on the image, and a segmentation energy includes the optimization of these labels in order to determine where to apply each prior. In [7] the authors perform density estimation in a non-linear feature space, where different objects are separable. The proposed energy is then minimized considering both the curve's control points and the image.

Considering the natural deformations and the variability of objects in a class, high-order shape models (SMs) should be included in the segmentation. Leventon et al. [1] compute PCA on a set of registered shapes (see also Tsai et al. [3]), fitting a Gaussian probability distribution to the coefficients of the reconstruction. This allows to compute the probability of a certain shape, included along with geodesic active contours for low level features, in an MAP estimation of the object in the image. Cootes and Taylor [8] compute, using PCA, a point distribution model of landmarks points defining a shape. More recently, Charpiat et al. [4] proposed a framework to compute non-linear shape statistics based on the Hausdorff distance between shapes, and then model distributions similarly to [1].

In this work, a new framework for image segmentation with multiple high-order shape models is introduced, addressing at the same time the selection of the model and its image-driven adjustment to the modeled deformations. In particu-
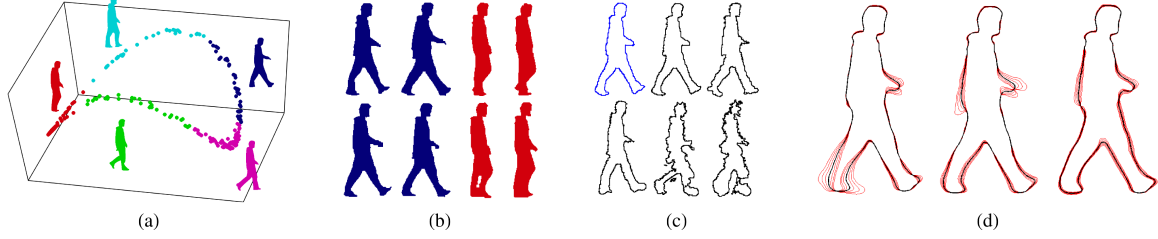
Fig. 1. (a) Each point correspond to the first three coordinates of the mapping obtained with diffusion maps, [9], for the shapes of a walking person. Clustering into five groups and one sample shape from each cluster (walking-cycle position) are shown. (b) Four shapes from two different clusters (one in blue and one in red). (c) Original shape (in blue) and its projections $\mathcal{P}_k\phi$ (in black) to $M = 5$ different models in the walking sequence, one for each cluster. Note how the projection is completely deformed when using the wrong model. (d) First three modes of variation for one of the models in the walking sequence; the thick black line is the mean shape, the red lines are obtained varying the amplitude (each figure is then the addition of the average and a constant times the first, second, or third eigenmode respectively).

lar, the high-order SMs are computed using PCA in a similar way as [1, 3], obtaining a set of eigenmodes of variations. The selection of the model is obtained with a binary selection coefficient, on-line learned based on a shape similarity measure between shapes. The proposed functional combines two terms, the first one is a region-based segmentation term [10]. The second term is a combination of the multiple high-order SMs, addressing the model selection and constraining the local features to the high-order shape information coming from the on-line selected model. While the framework is presented for planar curves, it can be easily extended to data in higher dimensions.

2. HIGH-ORDER SHAPES MODELS

Consider M sets Φ_k, $k = 1, \ldots, M$, each with N_k registered shapes $\Phi_k = \{\phi_k^1, \ldots, \phi_k^{N_k}\}$, where each $\phi_k^j \in \mathbb{R}^D$ is a SDF, whose zero level-set curve, $\mathcal{C}_k^j$, represents a shape from the k-th class of objects. The shape deformations in the set Φ_k are relatively small (compared with the deformations across different classes k). Let $\mathcal{M}_k^d$ be a d-order model that captures the intrinsic deformations of the training set Φ_k for the class k. A model $\mathcal{M}_k^d$ generates a representation of a new incoming shape $\hat{\phi}$, whose accuracy depends on the similarity between $\hat{\phi}$ and the shapes in Φ_k. In this work, $\mathcal{M}_k^d$ is derived from a PCA decomposition, and $\hat{\phi}$'s representation is given by the d-projection $\mathcal{P}_k^d\hat{\phi}$, $\mathcal{P}_k^d\hat{\phi} = \mu_k + U_k^d\alpha_k$, where $\mu_k \in \mathbb{R}^D$ is the mean shape of Φ_k, $U_k^d \in \mathbb{R}^{D \times d}$ is a matrix containing the first d modes of variation (eigenmodes), and $\alpha_k \in \mathbb{R}^d$ are the corresponding reconstruction coefficients, which of course depend on $\hat{\phi}$ (see [1, section 2.1] for details). Constraining small shape variations in the class Φ_k allows to obtain accurate representations using a linear approximation like PCA. If a class has large, non-linear deformation, a set of clusters may be considered, and the deformations in each cluster are linearly approximated. Finally, let $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_M\}$ be a set of SMs for the M different classes of objects (for simplicity, we omit the order d in the notation from now on). Figure 1 shows SMs for a walking person.

3. PROPOSED FRAMEWORK

Given an input image $\mathcal{I} : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ containing one or more shapes generated by the shape models $\mathcal{M}_k \in \mathfrak{M}$, an energy E is defined to simultaneously select the best model(s) and obtain a segmentation of the corresponding objects in $\mathcal{I}$ (a single object is considered now for simplicity), $(\mathcal{M}^*, \phi^*) := \arg\min_{\mathfrak{M},\phi} E(\mathcal{I}, \phi, c_0, c_1, \mathfrak{M})$. This energy includes two terms linearly combined with the constant λ, $E(\mathcal{I}, \phi, c_0, c_1, \mathfrak{M}) = E_{\mathrm{CV}}(\mathcal{I}, \phi, c_0, c_1) + \lambda E_{\mathrm{SM}}(\phi, \mathfrak{M})$.

E_{CV} is the energy introduced in [10], splitting the input data into two different regions of approximately piecewise constant values (c_0 and c_1). The term E_{SM} adds an additional force aiming at maximizing the similarity between the evolving shape ϕ and its projection onto *only one* of the d-order models from $\mathfrak{M}$. Which one of the M models is used depends on the evolution of the shape and its projection to each model. The proposed term is

$$E_{\mathrm{SM}}(\phi, \mathfrak{M}) = \sum_{k=1}^{M} \beta_k \int_\Omega \|H(\phi(p)) - H(\mathcal{P}_k\phi(p))\|^2 \, \mathrm{d}p, \quad (1)$$

where H is the Heaviside function, β_k is a binary coefficient that (on-line) selects which of the M models is used, and $\mathcal{P}_k\phi$ is the projection of ϕ to this model $\mathcal{M}_k$. Only one of the β_k must be different from zero in (1), since it is not fair to penalize for models that do not correspond to the object in the image. Which is the non-zero β_k is computed based on a shape dissimilarity measure (Υ) between two shapes,

$$\Upsilon(\phi_1, \phi_2) = \int_\Omega \left[\frac{|\phi_1(p)|\delta(\phi_2(p))}{\mathrm{length}(\mathcal{C}_2)} + \frac{|\phi_2(p)|\delta(\phi_1(p))}{\mathrm{length}(\mathcal{C}_1)} \right] \mathrm{d}p. \quad (2)$$

This is a length-normalized variation of the measure introduced in [11]. This measure evaluates the sum of Euclidean distances corresponding to moving the contour of the first shape to points in the contour of the second shape, and viceversa, scaled by the curves lengths. In Figure 1(c), the projected shapes are ordered according to increasing values of $\Upsilon_k(\phi) := \Upsilon(\phi, \mathcal{P}_k\phi)$ (1.35, 2.83, 3.59, 5.87, and 7.83 respectively).
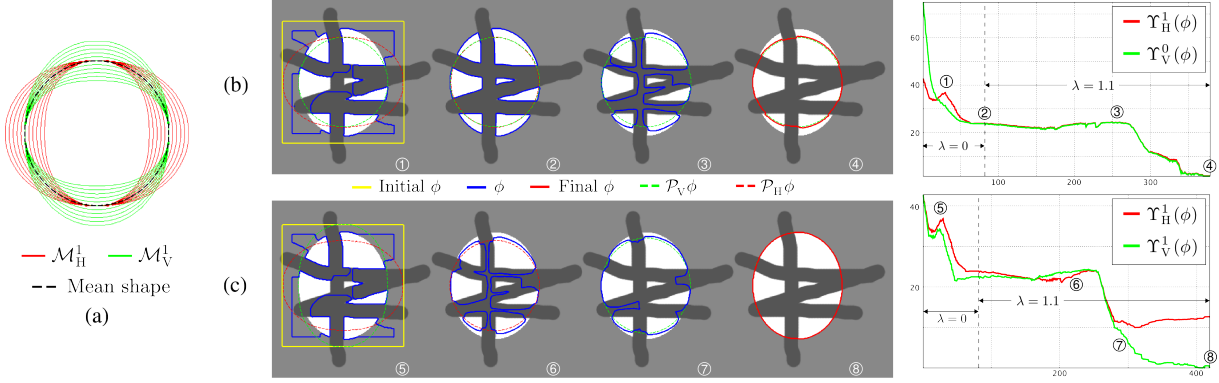
Fig. 2. (a) Mode of variation for the two ellipses models ($\mathcal{M}_\text{V}^1$ in green and $\mathcal{M}_\text{H}^1$ in red), the mean shape of both models is the same and is plotted in black dash line. (b) Results for experiments with $\mathcal{M}_\text{H}^1$ and $\mathcal{M}_\text{V}^0$ (only mean shape). Some steps in the segmentation (see text) and the evolution of the shape dissimilarity measure are also shown. (c) Same for experiments with $\mathcal{M}_\text{H}^1$ and $\mathcal{M}_\text{V}^1$ (complete model).

Based on (2), a normalized shape similarity measure $\bar{\xi}_k(\phi)$ between a shape ϕ and its projection $\mathcal{P}_k\phi$ to the d-order k-th model is computed as $\bar{\xi}_k(\phi) = \frac{\xi_k(\phi)}{\sum_{l=1}^{M} \xi_l(\phi)}$, where $\xi_k(\phi) = \exp\left(-\Upsilon_k(\phi)\right)$. This normalized similarity measure $\bar{\xi}_k(\phi)$ is close to one for the model that better represents ϕ. Finally to force the binary value in β_k soft thresholding, based on a sigmoid function, is performed. Note that a unique coefficient is used as model selector, instead of one coefficient in each pixel as in [5, 6]. This encourages shape consistency and significantly simplifies the optimization.

Shape validation. With the proposed method, one model is always selected and a segmentation is obtained, even if the shape in the image has no appropriate model that provides a good representation. The following measure permits to discard a segmentation ϕ_f given the model selected. First, the mean $\bar{\Upsilon}_k$ and variance $\sigma_{\Upsilon_k}^2$ of $\Upsilon(\phi_k^j, \mu_k)$ are computed, $\forall \phi_k^j \in \Phi_k$. Then if $\Upsilon(\phi_f, \mu_k) > \bar{\Upsilon}_k + 2\sigma_{\Upsilon_k}^2$, the segmentation is discarded, and the shape can not be recognized.

Energy minimization. The proposed energy is minimized using a classical gradient descent method. For the gradient descent of E_CV, the expression is given in [10, Equation (9)]. For E_SM, the obtained expression is

$$\frac{\partial E_\text{SM}}{\partial \phi} = -2\sum_{k=1}^{M} \beta_k \left\| H(\phi) - H(\mathcal{P}_k\phi) \right\| \left(\delta(\phi) - \delta(\mathcal{P}_k\phi)W \right),$$

where $W = U_k U_k^T$. Although the model selector β_k depends of ϕ, is treated as static, as a first order approximation for the gradient descent, since it affects the model selection and only indirectly the evolution of the curve.

The first steps of the optimization are performed with $\lambda = 0$, until stationarity, then E_SM is added with $\lambda \neq 0$ (manually obtained) until a new stationary point is reached (a similar idea is considered in [6]).

4. EXPERIMENTAL RESULTS

The first example is a "toy example" with two models of ellipses, where the only difference is that the first (and only) eigenmode is rotated $\frac{\pi}{2}$ (this already exemplifies the importance of high-order models). Let us name $\mathcal{M}_\text{V}^1$ the model with vertical deformations and $\mathcal{M}_\text{H}^1$ the one with horizontal ones. Figure 2(a) shows the mode of variation for both models. The input image contains an occluded vertical ellipse, not present in the training set. Two different experiments are presented, varying the order d of the model $\mathcal{M}_\text{V}^d$. With $d = 0$, only the mean shape is considered in the shape prior (no deformations), with $d = 1$ the vertical deformations are considered. All the parameters are the same in both experiments. Figures 2(b) and 2(c) show some steps in the minimization and the evolution of the shape dissimilarity measure, for both experiments, respectively. Steps ① and ⑤ show an intermediate curve in the evolution with $\lambda = 0$, and the projections, $\mathcal{P}_\text{H}^1\phi$ and $\mathcal{P}_\text{V}^d\phi$, to both models, dashed colored lines. The initial curve (in yellow) is also shown. Note that $\mathcal{P}_\text{V}^0\phi$ has no vertical deformations. The following steps (②,③ and ⑥,⑦) show the evolution after adding the E_SM term ($\lambda = 1.1$), and the obtained segmentation (④ and ⑧).

In the first experiment, the projections to both models end in the same shape (the mean shape), reflected in the graph of dissimilarity measure by the overlapping of the green and red curves. In the second experiment, $\mathcal{M}_\text{V}^1$ captures the variation of the input shape, as reflected in the obtained segmentation. In this case there is also a model switching around step 200 (step ⑥), where the $\mathcal{M}_\text{H}^1$ is selected while the occlusions are being filled. After this point, the vertical deformation determines the selection of $\mathcal{M}_\text{V}^1$ for the rest of the evolution, ending with an accurate segmentation.

For the next experiments, five models of a walking person cycle, $\mathcal{M}_{\text{w}k}^{21}$ ($k \in [1, 5]$, $d = 21$), shown in Figure 1, are used. This set of models is particularly challenging since they are different deformations of the "same object." Five new oc-
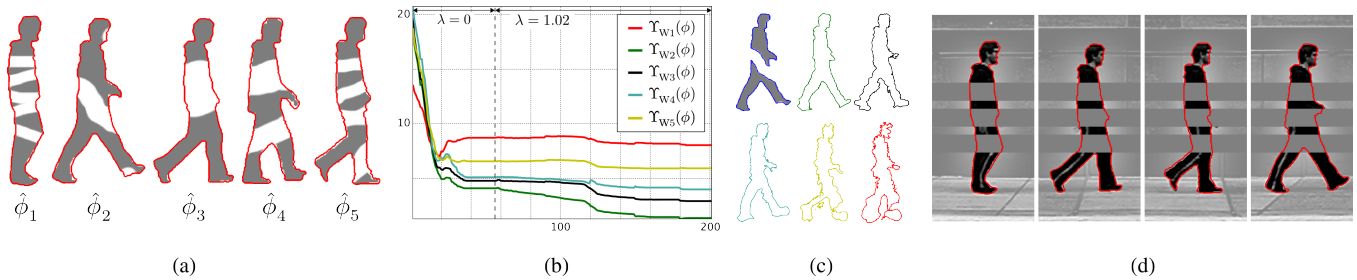
Fig. 3. Segmentation with walking model. See text for details.

cluded shapes $\hat{\phi}_k$, each one belonging to a different model and not in Φ_k, are segmented, Figure 3. In each case the correct model is selected and the segmented shape is adjusted to the gray levels, when present, and correctly completes the occlusions when the image information is missing, Figure 3(a) (in all the cases, the segmentations were validated with the proposed measure). Figure 3(b) plots the evolution of the shape dissimilarity measure for $\hat{\phi}_2$. Note that when the shape prior is added ($\lambda = 1.02$), $\Upsilon_{w2}(\phi)$ decreases faster than the others. The abrupt decay around step 120 corresponds to the filling of the main occlusion. Figure 3(c) shows $\hat{\phi}_2$ with the curve at the step when the shape prior is added, and the projections of ϕ (blue curve) to the five models, coded with colors. Finally, Figure 3(d) shows the segmentation obtained with our framework for gray level images. In this case, the automatically selected models, as well as the obtained segmentations, are correct and accurate.

Additional results, including the automatic classification and segmentation of multiple objects, generated from a standard dataset of fish, in noisy images, are presented at the conference.

5. CONCLUDING REMARKS

A framework for simultaneous and automatic model selection and object segmentation was introduced in this paper. The proposed technique is based on a new energy that combines region based segmentation with on-line selection of the best model for the object present in the image.

Possible directions for further improvements include incorporating high-order modes in the validation step and considering going beyond PCA, as well as including class-dependent model orders (d_k). Results in these directions will be reported elsewhere.

6. REFERENCES

[1] M. E. Leventon, W. E. L. Grimson, and O. D. Faugeras, "Statistical shape influence in geodesic active contours," in *CVPR*, 2000, pp. 1316–1323.

[2] D. Cremers, S. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation," *IJCV*, vol. 69, no. 3, pp. 335–351, 2006.

[3] A. Tsai, A. J. Yezzi, W. M. Wells III, C. Tempany, D. Tucker, A. C. Fan, W. E. L. Grimson, and A. S. Willsky, "Model-based curve evolution technique for image segmentation," in *CVPR (1)*, 2001, pp. 463–468.

[4] G. Charpiat, O. D. Faugeras, and R. Keriven, "Shape statistics for image segmentation with prior," in *CVPR*, 2007.

[5] D. Cremers, N. A. Sochen, and C. Schnörr, "A multiphase dynamic labeling model for variational recognition-driven image segmentation," *IJCV*, vol. 66, no. 1, pp. 67–81, 2006.

[6] N. Vu and B. S. Manjunath, "Shape prior segmentation of multiple objects with graph cuts," in *CVPR*, 2008, pp. 1–8.

[7] D. Cremers, T. Kohlberger, and C. Schnorr, "Shape statistics in kernel space for variational image segmentation," *Pattern Recognition*, vol. 36, no. 9, pp. 1929–1943, sep 2003.

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[9] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5 – 30, 2006, Diffusion Maps and Wavelets.

[10] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[11] T. Funkhouser, M. Kazhdan, P. Shilane, P. Min, W.Kiefer, A. Tal, S. Rusinkiewicz, and D. Dobkin, "Modeling by example," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 652–663, Aug. 2004.