

A PROBABILISTIC FRAMEWORK FOR SIMULTANEOUS SEGMENTATION AND CLASSIFICATION OF MULTIPLE CELLS IN MULTI-MARKER MICROSCOPY IMAGES

Renuka Shenoy, Min-Chi Shih, Kenneth Rose

Department of Electrical and Computer Engineering, University of California, Santa Barbara

ABSTRACT

Segmentation and classification of cells in biological data are important problems in bio-medical image analysis. This paper outlines a novel probabilistic approach to simultaneously classify and segment multiple cells of different classes in a multi-variate setting. Superpixels are extracted from the input vector-valued image, and a 2D hidden Markov model (HMM) is set up on the superpixel graph. HMM emission probabilities are used to ensure high confidence in local class selection based on superpixel feature vectors. Spatial consistency of labels is enforced by proper choice of transition probabilities, which are conditioned on the feature vectors of neighboring superpixels at each location. Optimal superpixel-level class labels are inferred using the HMM, and are aggregated to obtain global multiple object segmentation. The performance is demonstrated on a challenging microscopy dataset. Experiments show, both quantitatively and qualitatively, that the proposed approach effectively segments and classifies cells, outperforming related techniques.

Index Terms— Cell segmentation, cell classification, molecular marker, microscopy, multi-variate

1. INTRODUCTION

Multi-variate analysis is an active area of research in bioinformatics, with applications in sub-fields as diverse as genetic studies [1], hyperspectral imaging [2] and analysis of microscopy data [3]. Multi-variate methods offer many advantages over traditional methods. Jointly analyzing data provided by several markers can provide insight into correlation between phenotypes. Further, while traditional univariate methods require specialized markers for each class, multivariate methods can target multiple classes via different combinations of a small number of probes. It is also possible to discover new classes without explicitly probing for them. Finally, as multivariate methods take into account the response of multiple markers, they are more robust to variations that may be encountered over large volumes of data.

An example of a multi-marker microscopy modality is computational molecular phenotyping or CMP [3]. Consecutive tissue sections are probed for unique micromolecular markers, and each cell class can be distinguished by its multidimensional micromolecular signature. The RC1 connectome dataset [4] contains 6 capstone sections of CMP data followed by 341 sections of electron micrographs. Segmentation and classification of CMP is a critical step in the analysis of RC1 since cell types, their molecular phenotypes and their response to stimuli form an important source of supplementary information to neuronal connectivity data. Fig. 2 highlights some challenging aspects of CMP data. These include sudden spurious presence (Fig. 2 (a)) or absence (Fig. 2 (b)) of a marker within a cell,

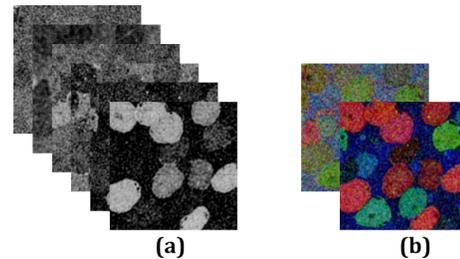


Fig. 1: (Best in color)(a) CMP data (from the RC1 connectome) consisting of 6 channels, each corresponding to a unique marker (b) RGB visualization CMP data channels taken three at a time.

cell classes that are difficult to detect due to low background contrast (Fig. 2 (c)) and large changes in intra-class feature variance across classes (Fig. 2 (d-e)).

A popular multi-variate segmentation algorithm is mean shift [5]. In this method, spatial coordinates are used in conjunction with feature vectors to determine local clusters, using user-defined spatial and feature bandwidths and minimum segment size parameters. However, there is often no set of parameters that results in accurate segmentation over the large range of cell sizes and class variances present in CMP data. There have been several papers that use random field formulations for classification and segmentation of objects. A superpixel-based technique was proposed in [6], in which a classifier is constructed on the histogram of local features and a conditional random field (CRF) is used to refine classification. In [7], a random forest classifier was used with hierarchical CRF to segment and classify images at multiple scales. The authors of [8] use a global bag of features model to combine top-down and bottom-up potentials to solve to segment multiple classes of objects. The drawback of the described methods is that they are not equipped to handle problems frequently occurring in microscopy data, such as varying contrasts and imaging artifacts.

The proposed segmentation algorithm addresses these issues in its formulation. Rather than pixels, we use superpixels as our atomic unit. This offers two major advantages over a sliding window approach: (i) we exploit redundancies in neighboring pixels to achieve a significant reduction in the number of computations (ii) local region statistics are calculated on a meaningful neighborhood rather than a window of fixed size. The major drawback of superpixel approaches is that the final segmentation result relies on the preservation of true boundaries in the initial oversegmentation. We mitigate this issue by using a state-of-the-art superpixel generation algorithm [9] which has high boundary recall. The parameters implementing spatial label consistency and local classification confidence are embedded in a two dimensional hidden Markov model (2D HMM) framework, and learned in a principled manner from training data. Cells are segmented by grouping contiguous regions with the same label.

This work was supported by NSF OIA 0941717 and NSF III 0808772.

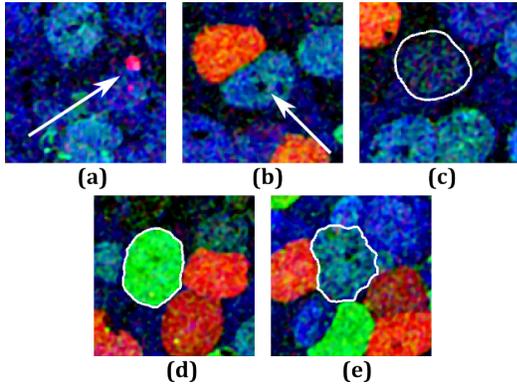


Fig. 2: (Best in color) Challenges in segmenting CMP images (a) Molecular marker artifact (b) “Hole” inside a cell (c) Cell having low contrast with background (d-e) Change in intra-cell feature variance - low variance in (d), high variance in (e)

We note that though experimental results in this paper are provided on CMP data from the RC1 connectome, the method can easily be applied to a wide variety of multivariate datasets as the formulation is general and system parameters are automatically learned from data.

2. PROPOSED METHOD

For data consisting of N cell types, we consider an M -class classification problem, where $M = N + 1$ (the additional “non-cell” class provides for background points in the image which lie between cells). We operate in the D -dimensional vector space, where D is the number of channels in the multi-marker image. Directly predicting the class label of each superpixel from its feature vector often results in incorrect labeling due to imaging artifacts or noise, or when even the label with the highest posterior probability has low confidence. Utilizing information from adjacent superpixels can help overcome this problem as there is typically a high probability of label agreement between neighboring regions. The trade-off between maintaining spatial label consistency and ensuring local selection of classes having high likelihood can be naturally modeled by embedding the system in a 2D HMM. The confidence of local classification in each superpixel is quantified by the emission probabilities of the HMM, while label coherence across neighboring superpixels is maintained by its transition probabilities.

Parameters of the HMM are learned, in a principled manner, from training data and the optimal label sequence is inferred using an efficient decoding algorithm. Finally, groups of superpixels bearing the same label are aggregated to obtain cell segmentation.

2.1. Superpixel Extraction

Superpixels are extracted using the SLIC (Simple Linear Iterative Clustering) algorithm [9], with the implementation provided in [10]. SLIC superpixels are computed on the vector-valued input image with a spatial regularization of 1 to ensure a regular lattice structure in the extracted oversegmentation.

2.2. Constructing the HMM

We construct a 2D HMM over a first order Markov mesh random field of size $X \times Y$, where X and Y are, respectively, the number of superpixels per row and column of the oversegmentation. Each superpixel \mathcal{S} corresponds to a node at location (x, y) in the 2D HMM, and is denoted by $\mathcal{S}_{x,y}$. Each state q of the HMM corresponds to one of M classes $\{\omega_m, m = 1, 2, \dots, M\}$. Our aim is to find the optimal state sequence, $Q^* = \{q_{x,y}^*, x = 1, 2, \dots, X, y = 1, 2, \dots, Y\}$, and hence, segment the image.

Direct 2D extension of standard learning and inference algorithms for 1D HMMs is intractable in most practical applications due to exponential increase in complexity. Consequently, many approximations have been suggested. The “turbo-HMM” [11, 12] consists of horizontal and vertical 1D HMMs that are decoded separately but “communicate” by iteratively inducing priors on each other. The turbo-HMM (T-HMM) framework is used in this work since it offers efficient approximations for learning and inference while outperforming alternative approaches.

2.3. Local Class Probabilities

Some classes are tightly packed in feature space (Fig. 2 (d)) while other classes have feature vectors that are more spread out (Fig. 2 (e)). In order to account for the difference in intra-class variance across classes while maintaining low complexity, we employ a quadratic discriminant analysis (QDA) classifier. Class likelihood functions at pixel p_k area calculated from its D -dimensional feature vector \mathbf{f}_k :

$$p(\mathbf{f}_k | \omega_m) = \frac{\exp\{-\frac{1}{2}(\mathbf{f}_k - \mu_m)^T \Sigma_m^{-1} (\mathbf{f}_k - \mu_m)\}}{(2\pi)^{\frac{D}{2}} |\Sigma_m|^{\frac{1}{2}}} \quad (1)$$

The parameters of the Gaussian likelihood functions, $\{\mu_m\}$ and $\{\Sigma_m\}$, are estimated from data labeled at pixel resolution. The emission probability $b_{x,y}^m$ represents the probability of superpixel $\mathcal{S}_{x,y}$ being emitted by class ω_m , without taking neighborhood information into consideration. The emission probability at each superpixel is obtained by combining contributions from its constituent pixels, by taking the geometric mean of all pixel-level likelihoods. In practice, $b_{x,y}^m$ is computed by taking the exponential of average log-likelihood within the superpixel, to improve numerical stability.

$$b_{x,y}^m = \left[\prod_{p_k \in \mathcal{S}_{x,y}} p(\mathbf{f}_k | \omega_m) \right]^{\frac{1}{|\mathcal{S}_{x,y}|}} \quad (2)$$

2.4. Neighborhood Label Consistency

Each element of the transition probability matrix $a_{m',m}$, is the probability of moving to state m' from state m . We define a spatially varying transition probability matrix, conditioned on local feature characteristics at each node.

$$a_{m',m}^H(x, y) = \begin{cases} C_m^S & m' = m \\ C_m^B & m' = M \\ \frac{1}{\beta_{m',m}} \exp\left(-\frac{R_{x,y}^H}{\beta_{m',m}}\right) & \text{otherwise} \end{cases} \quad (3)$$

where $R_{x,y}^H = \frac{1}{1 + K_{x,y}^H}$ and $K_{x,y}^H$ is the symmetric Kullback-Leibler (KL) divergence between the D -dimensional histogram of superpixel $\mathcal{S}_{x,y}$, given by $h_{x,y}$, and that of its horizontally neighboring superpixel, $\mathcal{S}_{x+1,y}$, given by $h_{x+1,y}$.

$$K_{x,y}^H = \sum_i h_{x,y}(i) \log \frac{h_{x,y}(i)}{h_{x+1,y}(i)} + \sum_i h_{x+1,y}(i) \log \frac{h_{x+1,y}(i)}{h_{x,y}(i)} \quad (4)$$

C_m^S represents the probability of self transition for each class ω_m . A high value of C_m^S indicates a higher probability of neighboring pixels taking on the same label. C_m^B represents the probability of transitioning from class ω_m to the background. This model favors state transitions (i.e., cell label changes) across neighboring superpixels with abrupt changes in their histograms, as quantified by the

exponential distribution with parameter $\beta_{m',m}$. We define a unique parameter $\beta_{m',m}$ for each pair of classes $\{\omega_m, \omega'_m\}$ to account for the variation in contrast between different pairs of classes.

We use a model similar to (3) for the transition probability matrix in the vertical direction. Since cellular microscopy images typically do not exhibit directionality along the coordinate axes, we make the simplifying (but removable) assumption that the parameters of the transition probability matrix, $\{C_m^S\}$, $\{C_m^B\}$ and $\{\beta_{m',m}\}$, are identical in the vertical and horizontal 1D HMMs.

2.5. Estimation of the Transition Probability Matrix

The parameters of the transition probability matrix are learned using the Baum-Welch algorithm [13], an implementation of the expectation-maximization (EM) algorithm. Re-estimation formulas for the parameters are derived by maximizing Baum's auxiliary function, given by:

$$\mathcal{Q}(\lambda, \lambda') = \sum_Q P(Q, I, S_{SP} | \lambda) \log P(Q, I, S_{SP} | \lambda') \quad (5)$$

with respect to λ' , where λ denotes the current estimate of HMM parameters, λ' the model re-estimate and Q , a sequence of states $Q = \{q_{x,y}\}$. I denotes the input image and S_{SP} , its superpixel oversegmentation.

During the E-step, the modified forward-backward algorithm of the T-HMM is used to estimate the occupancy probabilities (6) and ancillary training variables (7) in the horizontal and vertical directions.

$$\begin{aligned} \gamma_{x,y}^{H,\tau} &= P(q_{x,y}^H = \tau | I, S_{SP}, \lambda) \\ \gamma_{x,y}^{V,\tau} &= P(q_{x,y}^V = \tau | I, S_{SP}, \lambda) \end{aligned} \quad (6)$$

$$\begin{aligned} \xi_{x,y}^H(m', m) &= P(q_{x+1,y} = m', q_{x,y} = m | \lambda) \\ \xi_{x,y}^V(m', m) &= P(q_{x,y+1} = m', q_{x,y} = m | \lambda) \end{aligned} \quad (7)$$

During the M-step, $\mathcal{Q}(\lambda, \lambda')$ is maximized with respect to each parameter to obtain the following re-estimation formulas:

$$\hat{C}_m^S = \frac{\sum_{x,y} [\xi_{x,y}^H(m, m) + \xi_{x,y}^V(m, m)]}{\sum_{x,y} [\gamma_{x,y}^H(m, m) + \gamma_{x,y}^V(m, m)]} \quad (8)$$

$$\hat{C}_m^B = \frac{\sum_{x,y} [\xi_{x,y}^H(M, m) + \xi_{x,y}^V(M, m)]}{\sum_{x,y} [\gamma_{x,y}^H(M, m) + \gamma_{x,y}^V(M, m)]} \quad (9)$$

$$\hat{\beta}_{m',m} = \frac{\sum_{x,y} [\xi_{x,y}^H(m', m) R_{x,y}^H + \xi_{x,y}^V(m', m) R_{x,y}^V]}{\sum_{x,y} [\xi_{x,y}^H(m', m) + \xi_{x,y}^V(m', m)]} \quad (10)$$

Transition probability matrices are uniformly initialized, and parameters are re-estimated using iterative passes of the Baum-Welch algorithm until the likelihood of the training set converges.

2.6. Inference

The optimal label sequence is inferred using the Viterbi decoding procedure with modified forward-backward iterations [11].

3. EXPERIMENTAL RESULTS

We present the performance of the proposed approach on the CMP data in the capstone region of the RC1 connectome [4]. This region consists of 6 channels of data, each an intensity image obtained by probing for a unique molecular marker and captured at a resolution of 70 nm. The resulting data is a vector-valued image constituted

Method	Mean \pm Std of F-measure	Classificn. Accuracy	Running Time (s)
Mean shift [5]	0.7424 \pm 0.1793	-	30.51
FVS [6]	0.8093 \pm 0.1515	77.04%	21.30
Proposed	0.8372 \pm 0.1305	85.97%	23.04

Table 1: Comparison of results on CMP data from the RC1 connectome. Classification accuracy is not reported for mean shift segmentation as the algorithm does not directly provide classification output.

of six channels. The data consists of 581 total cells belonging to 7 major types. Manual segmentation of all cells, along with expert annotated cell labels, is used as ground truth in evaluating quality of both segmentation and classification.

We compare our method with two well known algorithms - mean shift segmentation [5] and the algorithm proposed by Fulkerson, Vedaldi and Soatto (FVS) in [6]. The optimal parameters for mean shift were found to be $(h_s, h_r, M) = (20, 8, 5000)$. We use 4-fold cross validation to train and test FVS as well as the proposed method. While running FVS, we set $K = 25$ as increasing beyond this value resulted in overfitting the data. SLIC was used to produce superpixels with an average size of 250 pixels for the proposed method. Segmentation results were obtained by aggregating contiguous superpixels having the same label. For all three methods, clumped cells of the same class were separated using the method described in [14].

The accuracy of segmentation is measured by comparing with ground truth. Using magnitude of pixel overlap, each ground truth cell is associated with at most one cell in the segmentation output. F-measure (F) is used as a measure of similarity between each ground truth cell (S_{gt}) and its corresponding segmented cell (S_{seg}).

$$F = \frac{2|S_{seg} \cap S_{gt}|}{|S_{seg}| + |S_{gt}|} \quad (11)$$

where $|\cdot|$ denotes number of pixels. The area (in pixels) of each cell in the ground truth is used to weight the corresponding F-measure in calculation of F-measure statistics. The weighted mean and standard deviation of the F-measure across all cells is used to compare the accuracy of segmentation results obtained from the three methods.

Classification accuracy is evaluated at the pixel level. The label of each superpixel is assigned to all the pixels within it. The resulting pixel-level labeling is compared with ground truth labeling. Accuracy is calculated as the percentage of pixels that are correctly classified.

Numerical results comparing the performance of the proposed approach to related methods are provided in Table 1, along with average run time for a 1024×1024 pixel region for each method. We observe significant improvement over the competing methods in both segmentation and classification accuracy.

Visual results on some example cells, along with the corresponding F-measure of each segmentation, are presented in Fig. 3. The proposed approach demonstrates the ability to handle challenging scenarios such as the presence of "holes" within cells (see Fig. 3 (a)), having large intra-cell variance (see Fig. 3 (b)) and cells having low contrast with background and surrounding cells (see Fig. 3 (e)), whereas competing methods are unable to accurately capture cell boundaries in these cases.

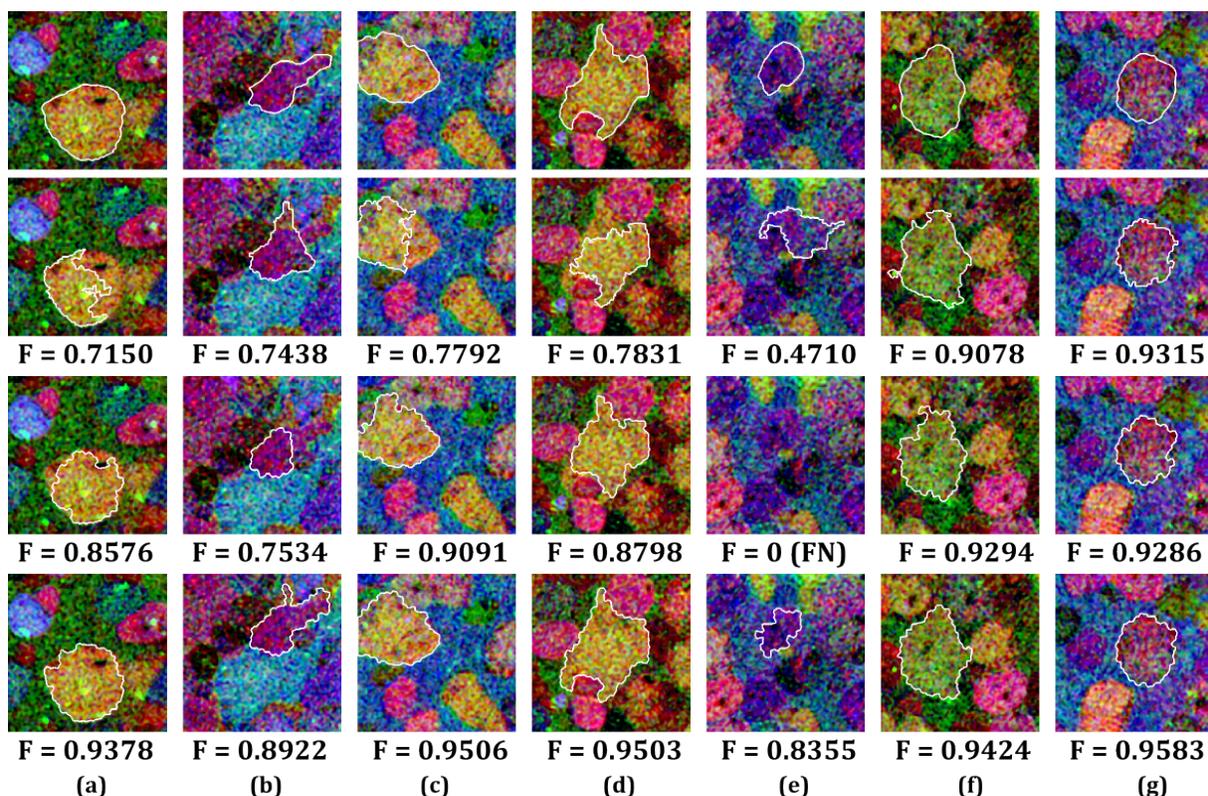


Fig. 3: (Best in color) Examples of segmentation results showing challenging as well as easy cases. Each column (a) - (g) shows results on a specific cell. The first row shows the ground truth segmentation, with the boundary outlined in white. The second, third and fourth rows show results obtained from mean shift segmentation [5], FVS [6] and the proposed method respectively, along with the F-measure F of the resulting segmentation. (An F-measure of 0 indicates a missed detection.) To visualize each result, we choose the three channels exhibiting the highest contrast between the given cell and background.

4. CONCLUSION

This paper presents a novel probabilistic algorithm for simultaneous segmentation and classification of cells in multi-marker images. Costs associated with neighborhood label coherence and local class membership probabilities are embedded in a 2D HMM framework. The T-HMM approximation is used to learn parameters of the HMM and to infer the optimal solution. We provide experimental validation on cellular microscopy data. The proposed method overcomes some of the main pitfalls of segmentation of such challenging data. As a result, we observe significant gains over competing methods in terms of both segmentation performance and classification accuracy.

5. REFERENCES

- [1] A. Maity, P.F. S., and J-i Tzeng, "Multivariate phenotype association analysis by marker-set kernel machine regression," *Genetic epidemiology*, vol. 36, no. 7, pp. 686–695, 2012.
- [2] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics*, vol. 19, no. 1, pp. 010901–010901, 2014.
- [3] R.E. Marc, R.F. Murry, and S.F. Basinger, "Pattern recognition of amino acid signatures in retinal neurons," *The Journal of neuroscience*, vol. 15, no. 7, pp. 5106–5129, 1995.
- [4] J. R. Anderson et al., "Exploring the retinal connectome," *Molecular vision*, vol. 17, pp. 355, 2011.
- [5] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [6] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *IEEE 12th International Conference on Computer Vision, 2009*, 2009, pp. 670–677.
- [7] M.Y. Yang and W. Forstner, "A hierarchical conditional random field model for labeling and classifying images of man-made scenes," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, pp. 196–203.
- [8] D. Singaraju and R. Vidal, "Using global bag of features models in random fields for joint categorization and segmentation of objects," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2313–2319.
- [9] R. Achanta et al., "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [10] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," <http://www.vlfeat.org/>, 2008.
- [11] F. Perronnin, J-L Dugelay, and K. Rose, "Iterative decoding of two-dimensional hidden markov models," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.*, 2003, vol. 3, pp. III–329.
- [12] F. Perronnin, J-L Dugelay, and K. Rose, "A probabilistic model of face mapping with local transformations and its application to person recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1157–1171, 2005.
- [13] L.E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.
- [14] J. Cheng and J.C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 741–748, 2009.