https://doi.org/10.2352/ISSN.2169-4451.2019.35.177

Detection, Imaging, and Quantification of DNA-Based Pathogen Based on Inkjet-Printed Test Strips

Min Zhao, Runzhe Zhang, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA
Susana Diaz-Amaya, Li-Kai Lin, School of Materials Engineering, Purdue University, West Lafayette, IN, USA
Amanda J. Deering, Department of Food Science, Purdue University, West Lafayette, IN, USA
Lia Stanciu, School of Materials Engineering, Purdue University, West Lafayette, IN, USA
George T.-C. Chiu, School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA
Jan P. Allebach, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA

Abstract

It is known that the foodborne pathogen E.coli O157:H7 from contaminated food can cause severe disease in our bodies and even death. Therefore, the detection of foodborne pathogens in our daily diet is crucial for global public health. We have previously reported an affordable, rapid, and simple method for detecting E.coli O157:H7 that uses inkjet patterning to create functionalized biosensing test strips detecting the target organism down to 10^2 *CFU/ml. In this paper, we focus on optimizing the response variations of our biosensors for detecting the same concentration of the target in two aspects: the comparison of various image segmentation methods and the optimization of the number of print layers.*

Introduction

The foodborne pathogen *E.coli* O157:H7 that produces toxins that damage the lining of the intestine, is a worldwide threat to public health. *E.coli* O157:H7 can be easily found in contaminated water and contaminated food, especially undercooked ground beef, milk and juice, raw fruits and vegetables. And, *E.coli* O157:H7 infections from contaminated food continue to occur regularly, and result in sever disease and even loss of life. Therefore, a fast, reliable, and affordable biosensor is more and more in demand. Many traditional lateral flow biosensors have been developed for foodborne pathogen detection. However, most of these are laboratory-based and cannot be manufactured economically. To address this problem, we incorporate printing and biomaterials technologies to design inkjet-printed test strips (PTSs) that guarantee rapid, affordable, reliable, and reproducible detection of the pathogens.

The detection result of biosensors can be determined by visual inspection or by a color measuring device. To have a quantitative and objective color analysis, color measuring instruments are more favored. As digital technologies continue to develop, cheap and compact image sensors are widely used in common electronics, like cell phones. The phone-based imaging system is promising for signal detection due to the above features, and also is emerging in different fields, such as high-resolution microscopy [1] and fluorescence imaging [2]. In our work, we propose a phone-based image processing algorithm to analyze images of the exposed test strip that have been captured with a mobile phone camera.

Overall, we develop the inkjet printing process for producing the pattern for the capture of *E.coli* O157:H7. The readouts of test results can be interpreted either by our image analysis system or by the naked human eye. We prove that our PTSs can successfully detect the presence of *E.coli* O157:H7 down to 10^2 CFU/ml [3].

One of the conditions that a successful biosensor must be meet is that the detection results show a low variation between assays. To measure response variations of our PTSs, ten independent experiments were performed for the same concentrations of *E.coli* O157:H7 test solution to obtain statistically reliable data by dropping the same amount of test solution onto the PTSs using a pipette. Figure 1 shows the color change in test zones for *E.coli* O157:H7 with the concentration 10^2 CFU/ml. The visible responses in the test zone and control zone indicate that the PTSs are able to successfully capture *E.coli* O157:H7 with the low concentration. Visually, we notice that the response of the test zones varies. The reasons are the following: (i) the random flow path of the reagents in the porous substrate (natural cellulose paper: Hi-flow HFC075, EMD Millipore); (ii) the extra amount of the bio-inks on the test zones increases the diffusion area; and (iii) the response intensity can be varied by using different segmentation methods. We optimize this problem in two aspects: the comparison of various image segmentation methods, and the optimization of the printing process in print layers that can reduce the sample-to-sample variation.

In this work, we propose two methods that reduce the response variation between assays. First, we review the various well-used image segmentation methods and apply them to detect the responses in the test lines of test strips. The usefulness of these segmentation methods is evaluated by comparing the response variance of the corresponding segmentation results. Second, we propose the optimization of the printing process, and obtain less response variation.

Image segmentation techniques

In order to quantify the response of detection results, a given phone-captured image needs to be appropriately segmented, so that the test zone can be separated from the background in the image. An example of phone-captured images is shown in Figure 1. The target region will consist of a contiguous region of pixel locations indicating low brightness. And the background consists of pixels that correspond to brighter or higher intensity pixels surrounding the target. However, there are some noisy black areas, such as scratch or black dots in the background, which increases

the difficulty of segmentation. Preprocessing is required before segmenting the test zones from the background. Based on the printing location of biomaterials, we crop the test zone containing detection information and background images and scale them to 60×130 and 60×50 pixels, respectively, using Adobe Photoshop, and save them in TIFF format without compression, as illustrated in Figure 2. Then, we transform the gamma-corrected sRGB values of the digital images to *CIE* $L^*a^*b^*$ color space which has a visually uniform distribution of colors and is closer to human perception of color differences than is sRGB. After transformation, an image difference matrix ΔE of each test image is obtained by subtracting the test zone images from the corresponding background image matrix using equation (1).

$$\Delta E = \sqrt{(L^* - L_{avg})^2 + (a^* - a_{avg})^2 + (b^* - b_{avg})^2} \qquad (1)$$

where $(L_{avg}, a_{avg}, b_{avg})$ are the average values from the local background. Then, we normalize the ΔE values of the images to the range [0, 255]. Image segmentation is one of the most important steps of image processing, in which an image is subdivided into several regions with the aid of pixel information, such as color, intensity, and texture. A number of automatic segmentation methods have been developed [4][5][6]. It is observed that the segmentation result can directly affect the subsequent analysis, and there is not a best image segmentation method because different images have different characteristics. Here, we evaluate how well several existing segmentation techniques can determine the response signal in our PTSs when the boundary between the test zone and the background region is unclear. In particular, we use Otsu's method [7], the valley-emphasis method [8], the twice Otsu method, the histogram thresholding using hierarchical cluster analysis [9], multi-level thresholding [10][11], and a Support Vector Machine (SVM) method [13]. The usefulness of these segmentation methods is also evaluated in this paper.

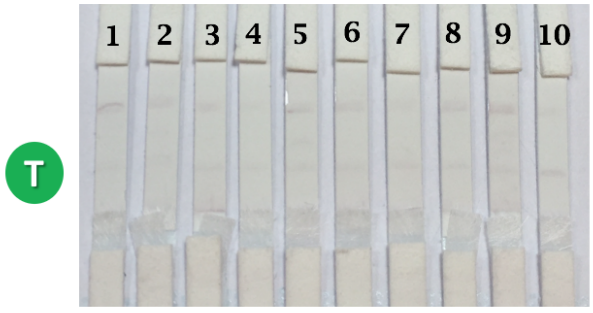

Figure 1. *An image of test strips detecting the target at the concentration of* 10^2 *CFU/ml.*

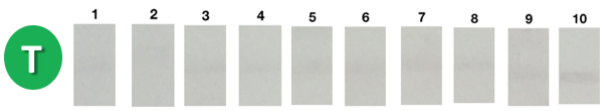

Figure 2. *The crop of just the test region for each test strip.*

Otsu's method

Otsu's thresholding method is one of the best-known methods for automatic image segmentation [7]. Based on the histogram of a grayscale image, Otsu's method finds the optimal threshold t^* that maximizes the between-class variance $\sigma_B^2(t)$. The optimal threshold is expressed as follows:

$$t^* = \underset{0 \le t < L}{\text{Arg Max}} \{\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t)\} \qquad (2)$$

where L is the number of distinct gray values ranging from 0 to L - 1, $\omega_1(t)$ and $\omega_2(t)$ are the probabilities of the two classes, and $\mu_1(t)$ $\mu_2(t)$ are the average gray values of the two classes. This method works well when the histogram has a strongly bimodal distribution. However, the segmentation result is poor when the histogram is close to an unimodal distribution, or when the background variance is large.

The valley-emphasis method

The valley-emphasis segmentation is a weighted Otsu's method. The idea of this method is to select the optimal threshold value corresponding to a grayscale value that has a small probability of occurrence and also maximizes the between-class variance, as in Otsu's method [8]. The objective function of this method is expressed as:

$$t^* = \underset{0 \le t < L}{\text{Arg Max}} \{(1 - p_t)(\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t))\}, \qquad (3)$$

where p_t is probability of occurrence of the grayscale value corresponding to the threshold t.

The twice Otsu method

We can observe that the background variance is large for each image, so single threshold segmentation algorithm is hard to effectively segment the target test zones from such kinds of background, as illustrated in Figure 3 O-7. It causes some of the background pixels to be classified as foreground pixels. To address this problem, we apply Otsu's method again to the segmented image after first application of single Otsu's threshold method. One of the examples that the segmented image after second application of single Otsu segmentation is shown in Figure 3 T-7, which is expected to remove extra background regions. This method is referred to as the twice Otsu method. In our application, the criterion for an image applied twice Otsu method is that the ratio between the height and the width of the foreground should be higher than an empirical chosen threshold. We use 0.4 as the threshold in our application.

The histogram thresholding method

The histogram thresholding method using cluster analysis is adopted to segment images with overlapping intensity distributions. Initially, every non-empty gray level is regarded as a separate mode contained in a cluster. Then, the smallest distance pair is merged based on the computation of distance between adjacent clusters. The distance between the clusters C_{k1} and C_{k2} is defined as

$$Dist(C_{k1}, C_{k2}) = \sigma_I^2(C_{k1} \cup C_{k2})\sigma_A^2(C_{k1} \cup C_{k2}), \qquad (4)$$

where $\sigma_I^2(C_{k1} \cup C_{k2})$ and $\sigma_A^2(C_{k1} \cup C_{k2})$ are inter-class variance and intra-class variance, respectively. $\sigma_I^2(C_{k1} \cup C_{k2})$ is the sum

of the square distances between the means of the two clusters and the total mean of both clusters. And $\sigma_A^2(C_{k1} \cup C_{k2})$ is the variance of all pixel values in the merged cluster. [9]

The optimal threshold for the m-level thresholding is obtained by iterating the previous process until m groups of gray levels are obtained. For the two-level thresholding, we obtain two clusters, C_1 and C_2, and the estimated threshold T_1, which is the highest gray level of the background.

The multilevel thresholding method

There is another multilevel segmentation method which is an extension of Otsu's method, called the multilevel thresholding method. [12] Assume there are m thresholds, $(t_1, \ldots, t_m)$ dividing the image into m classes. Then, the optimal thresholds can be determined by the following equations:

$$\begin{aligned} \text{Maximize } & J(t) = \sigma_0 + \sigma_1 + \sigma_2 + \ldots + \sigma_m \\ \text{where } & \sigma_0 = \omega_0(\mu_0 - \mu_T)^2 , \\ & \sigma_1 = \omega_1(\mu_1 - \mu_T)^2 , \\ & \sigma_2 = \omega_2(\mu_2 - \mu_T)^2 , \quad \ldots \\ & \sigma_m = \omega_m(\mu_m - \mu_T)^2 . \end{aligned} \tag{5}$$

where σ_0, ..., σ_m and μ_0, ..., μ_m, μ_T are the variance, the mean intensity of each class, and the mean intensity for the whole image, respectively. Each of the grayscale images in Figure 3 is divided into three classes with three optimal thresholds t_1, t_2, and t_3. And t_2 is the highest gray level of the background for each image.

Support vector machine classifier

The support vector machine method [13] is a supervised machine learning algorithm that has been applied to numerous classification problems. The key idea is to find the optimal hyperplane to separate two classes. The optimal hyperplane has a maximum margin between it and the nearest data points of both classes, termed Support Vectors. From our image data, we can observe that the grayscale values of the brightest pixels from the target are much closer to or a little larger than the grayscale values of the darkest pixels from the background. Therefore, due to overlapping intensity distributions and unclear boundaries, we choose the soft margin SVM algorithm [14] to apply to non-separable data. The training dataset is defined as (x_i, y_i), $i = 1, \ldots, n$ where x_i is the training sample and $y_i \in \{-1, +1\}$ is the class label. The objective function is the dual representation of the maximum margin problem:

$$\begin{aligned} & L_D(\alpha) = -\sum_{ij} \alpha_i \alpha_j t_i t_j x_i x_j \\ \text{subject to } & \quad 0 \le \alpha_i \le C \\ & \quad \sum_i \alpha_i t_i = 0 \end{aligned} \tag{6}$$

where the parameter C controls the trade-off between the training error and the margin, and the α_i's are the Lagrange multipliers. The optimal hyperplane is given by:

$$\begin{aligned} & f(x) = w^T x + b \\ \textit{where } & \quad w = \sum_i \alpha_i t_i x_i \end{aligned} \tag{7}$$

Here, w is the surface normal to the hyperplane, and $|b|/||w||$ represents the perpendicular distance between the hyperplane and the origin. The features adopted for classification in this work consist of normalized values of the grayscale images, and the $L*, a*, b*$ values at each pixel position. Drawing on the training experience of medical image processing [14][15], we employ a subject-specific training scenario. The golden standard segmentation results of the training images are obtained by manual segmentation based on the results from the multilevel thresholding method and the morphological operations. The ten images used as the training set consist of 10,000 test zones data samples and 20,000 background data samples.

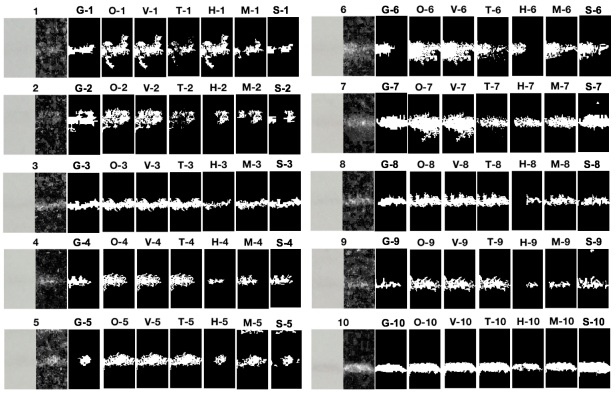

Figure 3. *The original images, grayscale images, and segmentation results for test images by the selected methods. (G, O, V, T, H, M, and S represent the abbreviation of the ground truth segmentation result, Otsu's method, valley-emphasis method, twice-Otsu method, histogram thresholding method, multi threshholding, and SVM, respectively.)*

Data analysis

Figure 3 shows the final versions of binary results using the above six segmentation methods combined with a 4-point connected component method. It can be seen that the segmentation results of Otsu's and the valley-emphasis methods are very close, and they can detect the presence of test zones. However, when the images have poor signal-to-noise ratio, neither method can precisely segment the test zone, as illustrated in Figure 3 O-7 and V-7. The twice Otsu method can extract the refined test zone from the backgrounds. For most images, the histogram thresholding method mainly segments a significant part of the test zones with larger grayscale values, but this can cause information loss, as illustrated in Figure 3 H-5. Compared with the SVM method, the detection results of the multilevel thresholding method are more refined. The segmentation results of the SVM method is mostly determined by the training data. For example, a training set with a low brightness contrast between the target origin and the background leads to a classifier that has a greater probability of confusing overlapping intensities of foregrounds and backgrounds, as illustrated in Figure 3 S-10. The segmentation results indicate that the multilevel thresholding method will extract the relatively bright region in the grayscale image.

To assess the variation among the responses in the test lines, we use the sum of ΔE from the detected region as the metric to quantify the visual response. Figure 4 shows graphs of the varia-

tions of the grayscale values for each test strip based on the metric and the selected segmentation methods. When analyzing the variation in the response results of the Otsu's method, the valley-emphasis method, and the multilevel thresholding method, we observe that the data of these three plots follow the same trend since they all related to Otsu's method. For the first two methods, 70% of the data is located in the area comprising the mean $\pm$ 1 SD (Standard Deviation), while the response of the multilevel thresholding result varies greatly. 70% of the data values from the twice Otsu method are also located in the area comprising the mean $\pm$ 1 SD, while they are more concentrated closely near the mean. The variation result of the histogram thresholding method shows that 60% of the data is distributed in the mean $\pm$ 1 SD. More data values spread out from the mean in the histogram thresholding method because of the incomplete extracted pattern. Although the binary results of the multilevel threshold method are better, its response variation is large. The plot of the SVM method shows that 80% of the data is located in the mean $\pm$ 1 SD. Overall, we conclude that the SVM method can segment test zones quite accurately and presents less variation of color response in the test zones.

The optimization of the printing process

The extra amounts of bio-inks printed on substrates increases the diffusion areas of the test lines. To address this problem, we optimize the print layers for test lines and control lines in the test strips. To find the optimal print number of layers of drop deposition, it is advantageous to observe the color change in test strips based on the different numbers of print layers. Here we use the checkboard print mask [3] which has been previously proved to provide better control of the printed DNA pattern. Figure 5 shows the image of the response signal in the test strips with varying numbers of print layers captured by a mobile phone camera. Note that for the target concentration at 10^3 CFU/ml, there is a less sample-to-sample variation of visual signals in the test strips with nine print layers compared with other numbers of print layers.

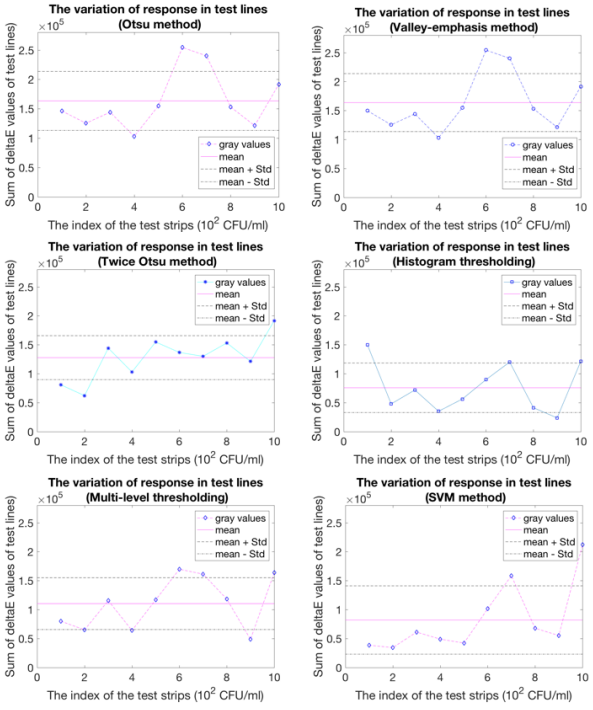

Figure 4. *The plots of the variation of the grayscale values as a function of the index of test strips.*

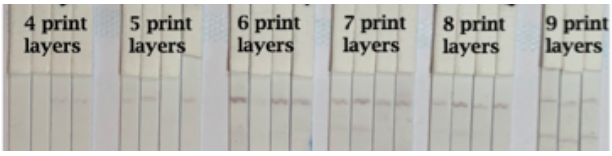

Figure 5. *An image of the response signal in the test strips with varying numbers of print layers captured by a mobile phone camera.*

Conclusion

The response variation of the optical properties of the test lines detecting the target at the same concentration is investigated. To reduce the sample-to-sample variation in response to *E.coli* O157:H7, as well as an effort to reduce material cost and printing time, we optimize the numbers of print layers. The analysis data of response variation using the various image segmentation methods has also been given. We evaluate selected thresholding methods for test zone images with the extracted pattern. And one example with ten samples has been provided to show the comparison of the response variations using these methods. We observe that the proposed SVM method has the least response variance. Also, we show that there is a less sample-to-sample variation of response signals in the test strips with the optimal set of printing layers.

Acknowledgments

This manuscript is based upon work supported by the U.S. Department of Agriculture, Agricultural Research Service, under Agreement No. 59-8072-6-001. Any opinions, findings, conclusion, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

References

[1] D.N. Breslauer, R.N. Maamari, N.A. Switz, W.A. Lam, and D.A. Fletcher, Mobile phone based clinical microscopy for global health applications, PloS one, 4(7), pg. e6320. (2009).

[2] H. Zhu, O. Yaglidere, T.W. Su, D. Tseng, and A. Ozcan, Cost-effective and compact wide-field fluorescent imaging on a cell-phone, Lab on a Chip, pg. 315. (2011).

[3] M. Zhao, S. Diaz Amaya, S.A. Jin, L.K. Lin, A.J Deering, L. Stanciu, G.T.C Chiu, and J.P. Allebach, Inkjet Printing platforms for DNA-based pathogen detection, NIP & Digital Fabrication Conference, pg. 107. (2018).

[4] N.R. Pal, and S.K. Pal, A review on image segmentation techniques, Pattern Recognition, pg. 1277. (1993).

[5] L. Lucchese, and S.K. Mitra, Colour image segmentation: a state-of-the-art survey, Proceedings-Indian National Science Academy Part A, pg. 207. (2001).

[6] W. Khan, Image segmentation techniques: A survey. Journal of Image and Graphics, pg.166. (2013).

[7] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics, pg. 62. (1979).
[8] H. F. Ng, Automatic thresholding for defect detection, Pattern Recognition Letters, pg. 1644. (2006).
[9] A. Z. Arifin, A. Asano, Image segmentation by histogram thresholding using hierarchical cluster analysis, Pattern Recognition Letters, pg. 1515. (2006).
[10] P.S. Liao, T.S. Chen, and P.C. Chung, A fast algorithm for multilevel thresholding, Journal of Information Science and Engineering, pg. 713. (2001).
[11] D.Y. Huang, and C.H. Wang, Optimal multi-level thresholding using a two-stage Otsu optimization approach, Pattern Recognition Letters, pg. 275. (2009).
[12] P.D. Sathya, and R. Kayalvizhi, Modified bacterial foraging algorithm based multilevel thresholding for image segmentation, Engineering Applications of Artificial Intelligence, pg. 595. (2011).
[13] C. Cortes, and V. Vapnik, Support-vector networks, Machine Learning, pg. 273. (1995).
[14] C.H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, Segmenting brain tumor with conditional random fields and support vector machines, in Proc. Workshop Comp, pg. 469. (2005).
[15] K. Zhang, J. Deng, and W. Lu, Segmenting human knee cartilage automatically from multi-contrast MR images using support vector machines and discriminative random fields, IEEE International Conference on Image Processing, pg. 721. (2011).

Author Biography

Min Zhao *is a graduate student in Electrical and Computer Engineering Department at Purdue University under the supervision of Professor Jan P. Allebach. She received her B.S. degree in Mechanical Engineering in 2010 from Qingdao Technology University, Shandong, China. Her current research interests include bio-printing and image analysis.*

Runzhe Zhang *is a Ph.D. student in Electrical and Computer Engineering at Purdue University under the supervision of Professor Jan P. Allebach. He received the master's degree in mechanical engineering from Boston University in 2016. His research interests are image processing, and printer defect detection using machine learning algorithms.*

Susana Diaz-Amaya*, PhD(c) in materials engineering at Purdue University with a bachelor's degree in microbiology from the Pontifical Javeriana University, Colombia. Susana joined the University of Tolima as an assistant professor in 2011, and started her industrial career, serving as leader of R&D for the agroindustry and as an independent consultant for 5 years. Her current research interest is focused on the design and fabrication of low-cost, novel nanomaterials for high-throughput manufacturing of biosensing platforms.*

Li-kai Lin *is currently a Ph.D. student in Materials Engineering at Purdue University. He earned a Master degree in Materials Engineering from the University of Southern California in 2015 and a Master of Science degree in Nanotech and Microsystems Engineering in 2009 plus bachelor degrees in Materials Engineering and Civil Engineering from National Cheng Kung University in 2007. His current research interests focus on fabricating biosensors for rapid diagnosis as well as semiconductor manufacturing.*

Amanda Deering *is a clinical assistant professor of Food Science at Purdue University. Her research focuses on examining the internalization of human pathogenic bacteria in plants, and routes of contamination that can contribute to plants harboring pathogenic bacteria. Amanda also works closely with industry to develop and test novel sanitization treatments that can be used for fresh produce, and works directly with Indiana fruit and vegetable growers to address food safety issues on the farm.*

Lia Stanciu *is a professor of Materials Engineering at Purdue University. She earned her Ph.D. in Materials Science from U.C. Davis in 2003. In 2005, she joined Purdue's School of Materials Engineering as a faculty member. Her main research areas include advanced biomaterials for implantation, materials for biosensing devices and nanotechnology.*

George T. Chiu *is a Professor of Mechanical Engineering with courtesy appointments in Electrical and Computer Engineering and Psychological Sciences at Purdue University. He received the B.S. degree from National Taiwan University and the M.S. and Ph.D. degrees from University of California at Berkeley. His research interests are mechatronics and control with applications to digital printing and imaging systems, digital fabrications and functional printing. He is a Fellow of ASME and IS&T.*

Jan P. Allebach *is Hewlett-Packard Distinguished Professor of Electrical and Computer Engineering at Purdue University. Allebach was named Electronic Imaging Scientist of the Year by IS&T and SPIE, and was named Honorary Member of IS&T, the highest award that IS&T bestows. He has received the IEEE Daniel E. Noble Award, the IS&T/OSA Edwin Land Medal, is a Fellow of the National Academy of Inventors, and is a member of the National Academy of Engineering.*