https://doi.org/10.2352/EI.2022.34.14.COIMG-229

Multiresolution DECOLOR for Camouflaged Moving Foreground Detection Using a Redundant Wavelet Transform

***Zoe M. Fowler*[1]*, James E. Fowler*[2]*, Agnieszka Miguel*[3]**
1 Georgia Institute of Technology, Atlanta, GA USA
2 Mississippi State University, Starkville, MS USA
3 Seattle University, Seattle, WA USA
Email: zfowler3@gatech.edu, fowler@ece.msstate.edu, amiguel@seattleu.edu

Abstract

Detection of moving foreground objects is essential to many image-sequence-analysis applications. However, preexisting methods tend to work best when the foreground is visually distinct from the background, suffering when objects are camouflaged. To address this shortcoming, a foreground-extraction algorithm resilient to camouflage is proposed by incorporating a redundant discrete wavelet transform into the well-known DECOLOR technique based on a sparse and low-rank model of the foreground-extraction problem. Detection of camouflaged moving objects is enhanced as a result of the combination of multiple background estimates in independent wavelet subbands into an overall estimate of the background, leveraging the known robustness of redundant wavelet transforms to additive noise. Experimental results demonstrate that the proposed method offers robustness to camouflage superior to that of other competing methods for image sequences containing snow leopards in the wild.

Introduction

Extraction of foreground objects is an essential task in image-sequence analysis. Foreground-object extraction detects and demarcates moving people, animals, or other objects—a critical undertaking in visual surveillance. While many foreground-extraction methods have been developed in prior literature, they tend to work best when the foreground is visually distinct from the background, which proves to be a challenge when camouflaged foreground objects blend into the image-sequence background.

This camouflage issue is of paramount importance in many applications in the field of conservation biology wherein cameras are placed in the wild to observe and analyze populations of endangered species. For example, [1–3] consider image sequences of highly elusive snow leopards. The snow leopard is a vulnerable species found throughout Central Asia, where poaching and ecological disruptions threaten its survival. Conservation biologists have set up surveillance across Central Asia to monitor the snow-leopard population; these "camera traps" take pictures whenever a source of heat passes in front of the camera. Currently, human analysts must sort through thousands of images manually in order to determine whether a snow leopard is present or not, as well as to identify individual cats—clearly, automated processing of this task would be of great benefit. However, since snow leopards exhibit a high degree of camouflage in their natural habitat, they pose a significant challenge to existing foreground-extraction methods.

While there have been numerous foreground-extraction efforts in prior literature, only a few specifically address the issue of camouflage. One of the more recent and effective techniques is the "Fusion in the Wavelet domain for Foreground detection in Camouflaged scenes" (FWFC) algorithm [4] which leverages the multiresolution analysis provided by a wavelet transform. Such wavelet-domain detection is premised on an observation that camouflaged objects may be more distinguishable from the background at certain resolutions or orientations; FWFC thereby leverages the multiresolution character of a wavelet transform in a cross-scale probabilistic model applied to transform coefficients. However, FWFC fails to exploit other relevant aspects of the foreground-extraction problem—most notably, that we expect foreground objects to be spatially contiguous. On the other hand, the well-known "Detection of Contiguous Outliers in the Low-Rank Representation" (DECOLOR) [5] does impose explicit spatial contiguity on the foreground-extraction process by incorporating a graph-driven spatial-adjacency regularization into a sparse and low-rank optimization. However, like many other foreground-extraction techniques, DECOLOR does not explicitly address camouflage.

Consequently, in this paper, we propose a new foreground-extraction algorithm that adopts a wavelet-based strategy for robustness to camouflage inspired by FWFC. Unlike FWFC, however, we deploy the wavelets within the sparse/low-rank DECOLOR framework. The resulting algorithm—which we call multiresolution DECOLOR (MR-DECOLOR)—couples the robustness to camouflage of the wavelet transform from FWFC with the spatial contiguity of the foreground as explicitly enforced by DECOLOR. Additionally, robustness to camouflage is further enhanced because MR-DECOLOR forms multiple background estimates in independent wavelet subbands and combines them into a single overall background estimate, leveraging the known robustness of redundant wavelet transforms to additive noise [6]. Below, we describe MR-DECOLOR in detail and evaluate its performance experimentally on several camera-trap sequences of snow leopards in the wild.

Background

A simple method for background estimation is to find the median temporally over the image sequence. That is, let an N-frame image sequence be represented as $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N] \in \Re^{P\times N}$, where $\mathbf{x}_n$ is frame n of the sequence vectorized as a column vector, and P is the number of pixels in each frame. The median provides a single-image background estimation as

$$\mathbf{b} = \operatorname{median}\{\mathbf{X}\}, \tag{1}$$

where the median is calculated temporally for each pixel location. From this background image $\mathbf{b}$, the foreground can be extracted—for example, a pixel in $\mathbf{X}$ can be classified as foreground if it lies more than one standard deviation away from the background (e.g., [1]). However, as a foreground-extraction method, this approach tends to generate relatively poor results due to is failure to exploit not only spatial information but also frame-to-frame motion.

A more sophisticated approach to background estimation was proposed as a practical use of robust principal component analysis (RPCA) [7]. In this, the image sequence is modeled as the combination of a low-rank component (the background $\mathbf{B}$) and a sparse component (the moving foreground $\mathbf{S}$) which are jointly estimated by solving

$$\min_{\mathbf{B},\mathbf{S}} \|\mathbf{B}\|_* + \lambda\,\|\mathbf{S}\|_1 \quad \text{s.t.} \quad \mathbf{B}+\mathbf{S}=\mathbf{X}, \tag{2}$$

where the nuclear norm is a convex proxy for $\operatorname{rank}(\mathbf{B})$, the ℓ_1 norm is a convex proxy for the sparsity of $\mathbf{S}$, and, again, the columns of $\mathbf{X}$ are frames of the original image sequence. While $\mathbf{B}$ provides a background estimation directly, for foreground extraction, an explicit foreground mask must be derived from $\mathbf{S}$. For example, [2] classifies a pixel as foreground if its corresponding value in $\mathbf{S}$ lies more than one standard deviation away from the mean value of $\mathbf{S}$, similar to what was done for the median approach above. Like the median method, RPCA also does not impose any spatial coherence on the foreground and tends to have difficulty discerning the motion of camouflaged objects.

Inspired by the low-rank/sparse formulation of RPCA, but imbued with a goal of spatially contiguous foreground objects, DECOLOR [5] features an iterative optimization that integrates object detection and background learning into one process. The DECOLOR optimization is

$$\min_{\mathbf{B},\mathbf{M}} \frac{1}{2}\left\|\overline{\mathbf{M}}\circ(\mathbf{X}-\mathbf{B})\right\|_F^2 + \alpha\left\|\mathbf{B}\right\|_* + \beta\left\|\mathbf{M}\right\|_1 + \gamma\left\|\mathbf{A}\operatorname{vec}(\mathbf{M})\right\|_1, \tag{3}$$

where $\mathbf{B}$ is a matrix of the desired background images; $\mathbf{M}$ is the desired foreground mask, a binary matrix indicating regions in $\mathbf{X}$ which are detected as foreground; $\mathbf{A}$ is a node-edge incidence matrix that represents a spatial neighborhood surrounding each pixel in $\mathbf{X}$; and $\circ$ is the Hadamard product. Like RPCA, DECOLOR aims for a low-rank background (via regularization with nuclear norm $\|\mathbf{B}\|_*$) as well as a sparse foreground (via regularization with $\|\mathbf{M}\|_1$); unlike RPCA, though, DECOLOR features an explicit binary mask ($\mathbf{M}$) of the foreground as well as regularization designed to impose spatial contiguity on the foreground (via $\mathbf{A}$). The DECOLOR algorithm is summarized in Alg. 1 and consists of an alternating optimization: finding low-rank background $\mathbf{B}$ by SOFT-IMPUTE [8] assuming a fixed mask $\mathbf{M}$ (Alg. 2), and finding $\mathbf{M}$ by graph cuts [9] assuming a fixed $\mathbf{B}$ (Alg. 3). However, although DECOLOR successfully maintains a high degree of contiguity in the foreground objects it extracts, it tends to work best when the foreground is clearly distinct from the background. Like with many other foreground-extraction techniques, camouflaged foreground objects present a formidable challenge to DECOLOR.

Algorithm 1 The DECOLOR Algorithm

1: **Input:** $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N] \in \Re^{P\times N}$
2: **Initialize:** $\mathbf{B} \leftarrow \mathbf{X}$, $\mathbf{M} \leftarrow \mathbf{0}$, α, β
3: **repeat**
4: $\quad \{\mathbf{B},\alpha\} = \operatorname{EstimateBackground}(\mathbf{X},\mathbf{B},\mathbf{M},\alpha)$
5: $\quad \{\mathbf{M},\beta\} = \operatorname{EstimateMask}(\mathbf{X},\mathbf{B},\mathbf{M},\beta)$
6: **until** convergence
7: **Output:** $\mathbf{B},\mathbf{M}$

Algorithm 2 Background Estimation via SOFT-IMPUTE

1: **function** $\operatorname{EstimateBackground}(\mathbf{X},\mathbf{B},\mathbf{M},\alpha)$
2: $K \leftarrow \lfloor\sqrt{N}\rfloor$
3: **repeat**
4: $\quad \mathbf{B} \leftarrow \Theta_\alpha\left(\overline{\mathbf{M}}\circ\mathbf{X} + \mathbf{M}\circ\mathbf{B}\right)$
5: **until** convergence
6: **if** $\operatorname{rank}(\mathbf{B}) \le K$ **then**
7: $\quad \alpha \leftarrow \alpha/\sqrt{2}$
8: $\quad$ **goto** Step 3
9: **end if**
10: **return** $\mathbf{B}$, α

Note: $\Theta(\cdot)$ is the singular-value thresholding operator

On the other hand, FWFC [4] is constructed by design to address the problem of camouflaged foreground objects. Specifically, FWFC attempts to identify foreground in the domain of a redundant discrete wavelet transform (RDWT) [6], in the hope that small differences in the image domain become more visible in one or more wavelet subbands. FWFC estimates each RDWT coefficient's likelihood of being foreground or background by developing cross-scale probabilistic models for both foreground and background for each wavelet band; i.e.,

$$\begin{aligned} p(x^\theta|\text{fg}) &= p(x_1^\theta|\text{fg})\,p(x_2^\theta|\text{fg})\cdots\theta\,p(x_L^\theta|\text{fg}) \\ p(x^\theta|\text{bg}) &= p(x_1^\theta|\text{bg})\,p(x_2^\theta|\text{bg})\cdots\theta\,p(x_L^\theta|\text{bg}), \end{aligned} \tag{4}$$

where x_l^θ is an RDWT coefficient of subband orientation θ at transform level l, $\theta \in \{\text{LL},\text{LH},\text{HL},\text{HH}\}$. This calculation is possible due to the fact that each RDWT subband is equal in size to that of the original image. These likelihoods are then fused across the wavelet orientations, resulting in a probabilistic estimate of whether a wavelet coefficient is foreground or background; the maximum likelihood then produces a binary foreground mask. Experimental results in [4] demonstrate that this mask is accurate even under camouflage conditions due to the multiresolution aspect of the RDWT.

The Proposed MR-DECOLOR Algorithm

Inspired by FWFC's use of the RDWT for camouflage, as well as the spatial contiguity in DECOLOR, our proposed MR-DECOLOR method for foreground extraction incorporates an RDWT into the DECOLOR framework (Alg. 1), resulting in the MR-DECOLOR algorithm as described by Alg. 4. Specifically, in MR-DECOLOR, the background-estimation component of DECOLOR (as implemented via SOFT-IMPUTE as Alg. 2) takes place within the RDWT domain, with an inverse RDWT casting the estimated background back into the spatial domain prior to

Algorithm 3 Mask Estimation via Graph Cuts

function EstimateMask$\big(\mathbf{X},\mathbf{B},\mathbf{M},\beta\big)$
$\mathbf{A} \leftarrow$ graph adjacency matrix for spatial neighbors
$\gamma \leftarrow 5\beta$
estimate σ from $\mathbf{X}-\mathbf{B}$
$\beta \leftarrow \max(\beta/2, 4.5\sigma^2)$

$$\mathbf{M} \leftarrow \arg\min_{\mathbf{M}} \sum_{p,n} \left(\beta - \frac{1}{2}(\mathbf{X}_{p,n} - \mathbf{B}_{p,n})^2\right) \mathbf{M}_{p,n} + \gamma \|\mathbf{A}\,\mathrm{vec}(\mathbf{M})\|_1 \quad \text{(solve via graph cuts)}$$

return $\mathbf{M}, \beta$

Algorithm 4 The MR-DECOLOR Algorithm

Input: $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N] \in \Re^{P\times N}$, num subbands S
Initialize: $\mathbf{B} \leftarrow \mathbf{X}$, $\mathbf{M} \leftarrow \mathbf{0}$, $\{\alpha_s\}_{s=1,\ldots,S}$, β
$\{\tilde{\mathbf{X}}_s\}_{s=1,\ldots,S} \leftarrow \mathrm{RDWT}(\mathbf{X})$
for $s = 1,\ldots,S$ **do**
  $\tilde{\mathbf{B}}_s \leftarrow \tilde{\mathbf{X}}_s$
end for
repeat
  for $s = 1,\ldots,S$ **do**
    $\big\{\tilde{\mathbf{B}}_s, \alpha_s\big\} \leftarrow$ EstimateBackground$\big(\tilde{\mathbf{X}}_s, \tilde{\mathbf{B}}_s, \mathbf{M}, \alpha_s\big)$
  end for
  $\mathbf{B} \leftarrow \mathrm{RDWT}^{-1}\big(\{\tilde{\mathbf{B}}_s\}_{s=1,\ldots,S}\big)$
  $\big\{\mathbf{M},\beta\big\} \leftarrow$ EstimateMask$\big(\mathbf{X},\mathbf{B},\mathbf{M},\beta\big)$
until convergence
Output: $\mathbf{B},\mathbf{M}$

the mask-estimation component (implemented via graph cuts as in Alg. 3).

Conducting the background estimation within the RDWT domain yields two advantages. First, following the same motivation behind FWFC [4], the addition of the RDWT to MR-DECOLOR provides robustness to camouflage. That is, FWFC employs an RDWT to yield a multiresolution analysis that separates the image into separate subbands in hopes that a camouflaged object would be detectable in at least one of the wavelet subbands. MR-DECOLOR employs the complementary motivation—namely, we hope that the background is more separable from the camouflaged foreground in at least one of the wavelet subbands. Therefore, we employ SOFT-IMPUTE independently in each RDWT subband, generating independent background estimates for each subband.

The second advantage of the RDWT in MR-DECOLOR is that it leverages the known robustness to noise of the RDWT [6] to improve the background estimate prior to subsequent mask estimation. Specifically, each SOFT-IMPUTE background estimate can be considered to be the true background subband corrupted with additive noise due to an inaccurate estimation. The inverse RDWT merges each background estimate in the subbands into a final background estimate in the spatial domain while simultaneously diminishing this distortion, thus ultimately generating a background estimate closer to the true background. Finally, the usual DECOLOR mask estimation via graph cuts is then applied using this generated background.

Table 1. CCR (in percent) for the various techniques

	Dataset				
Method	*#1*	*#2*	*#3*	*#4*	*#5*
Median	88.8	95.1	82.9	88.3	89.6
RPCA	90.0	95.0	83.3	85.9	90.6
DECOLOR	79.6	92.4	73.6	86.3	91.8
FWFC	**93.5**	95.4	85.3	87.8	92.4
MR-DECOLOR	89.9	**96.4**	**88.5**	**89.0**	**95.1**

Experimental Results

The five image sets selected for the testing of MR-DECOLOR were obtained through camera traps set throughout Central Asia. These camera trap images were provided by Panthera, an organization devoted to the conservation of wild cats. Datasets #1, #2, and #3 were images taken during the daytime, where each snow leopard blends in with the rocky background. Datasets #4 and #5 were taken at night and also exhibit a high amount of camouflage, blending in with both the grass and rocks in the background. By selecting image sets with a high amount of camouflage and at varying times of the day, the robustness of MR-DECOLOR to camouflage is effectively tested.

We now test the performance of our proposed MR-DECOLOR foreground extraction in comparison to other foreground-extraction techniques. Specifically, we compare against four other techniques: the median method as used in [1, 2], RPCA [7], DECOLOR [5], and FWFC [4]. To make a quantitative comparison, the foreground masks produced by each technique were compared to ground-truth foreground masks that were drawn by hand for each sequence of images, and the correct classification rate (CCR),

$$\mathrm{CCR} = \frac{\mathrm{TruePositives} + \mathrm{TrueNegatives}}{\mathrm{TotalPixels}}, \tag{5}$$

was calculated with respect to the ground-truth mask. These CCR results are reflected in Table 1. Performance is also illustrated visually in Fig. 1 for a single frame from each dataset.

In all cases except Dataset #1, MR-DECOLOR outperformed the rest of the methods, as evidenced by the high resulting CCRs reported in Table 1. We note that Dataset #1 had a high level of camouflage present across all image frames as well as a faint moving shadow cast by the foreground leopard; however, this shadow was not included in the ground-truth mask. MR-DECOLOR more successfully extracts this moving shadow along with the foreground leopard than do the other techniques under consideration, ultimately resulting in MR-DECOLOR yielding a lower CCR than FWFC for this dataset. We anticipate that incorporation of explicit shadow suppression (e.g., [10, 11]) might rectify this issue, but this is beyond the scope of the present work.

Conclusions

In this paper, we proposed a new foreground-extraction approach, MR-DECOLOR, by incorporating an RDWT into the well-known DECOLOR algorithm. MR-DECOLOR was inspired by the use of wavelet-based multiresolution analysis originating in the recent FWFC algorithm, which in turn was motivated by an observation that foreground objects may be more separable from the background at certain wavelet resolutions or orientations than

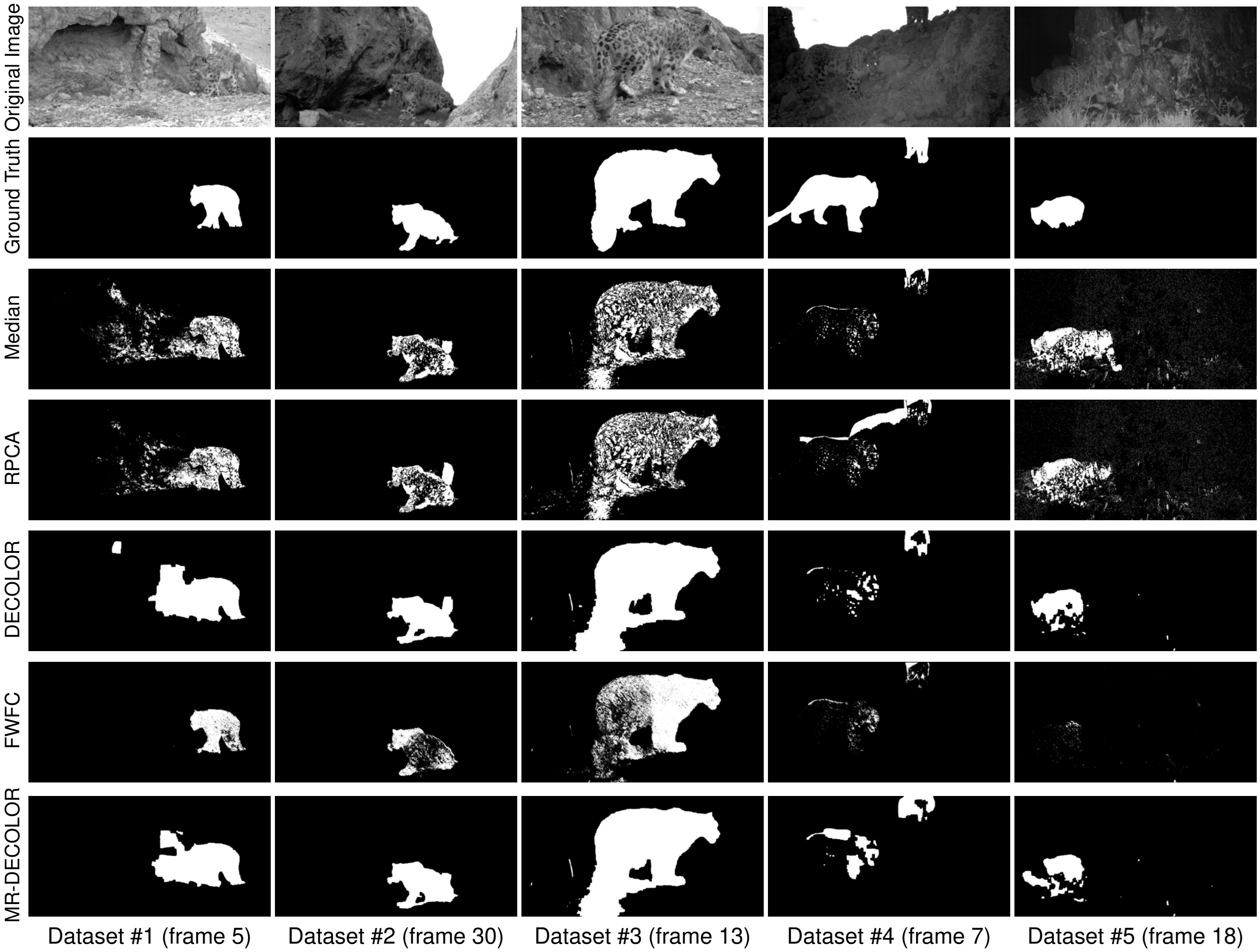

Figure 1. *Original images, corresponding ground-truth foreground masks, and the estimated foreground mask from the various algorithms under consideration.*

in the original image domain. Robustness to camouflage was also enhanced in MR-DECOLOR due to the fact that multiple background estimates in independent RDWT subbands were combined into a single overall background estimate, leveraging the known robustness of the RDWT to additive noise. Experimental results conducted on real camera-trap datasets featuring highly camouflaged snow leopards in the wild confirmed that MR-DECOLOR offered superior robustness against camouflage when compared to competing foreground-extraction methods for most of the scenarios considered.

Acknowledgment

The authors thank Rana Bayrakçismith and Panthera Corp. for access to the images used in this study.

References

[1] A. Miguel, S. Beery, E. Flores, L. Klemesrud, and R. Bayrakcismith, "Finding areas of motion in camera trap images," in *Proceedings of the IEEE International Conference on Image Processing*, Phoenix, AZ, September 2016, pp. 1334–1338.

[2] A. Miguel, J. S. Beard, C. Bales-Heisterkamp, and R. Bayrakcismith, "Sorting camera trap images," in *Proceedings of the IEEE Global Conference on Signal and Information Processing*, Montreal, Canada, November 2017, pp. 249–253.

[3] D. Grob, M. Vurkaç, A. Miguel, M. Mandich, and R. Bayrakçismith, "Pareto optimization of parameter selection speeds up and improves quality of motion computation: Applying evolutionary multi-objective optimization to randomized-subspace robust PCA," in *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops*, Singapore, November 2018, pp. 919–928.

[4] S. Li, D. Florencio, W. Li, Y. Zhao, and C. Cook, "A fusion framework for camouflaged moving foreground detection in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3918–3930, August 2018.

[5] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 597–610, March 2013.

[6] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 629–632, September 2005.

[7] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal compo-

nent analysis?" *Journal of the Association for Computing Machinery*, vol. 58, no. 3, pp. 1–37, May 2011.

[8] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *Journal of Machine Learning Research*, vol. 11, no. 80, pp. 2287–2322, 2010.

[9] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, November 2001.

[10] R. Guo, Q. Dai, and D. Hoiem, "Paired regions for shadow detection and removal," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2956–2967, December 2013.

[11] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, July 2003.

Author Biography

Zoe M. Fowler *received the B.S. degree in electrical engineering from Mississippi State University. She is currently pursuing a Ph.D. degree in electrical engineering at the Georgia Institute of Technology funded by a Graduate Research Fellowship from the National Science Foundation. Her current research interests include machine learning for biomedical applications.*

James E. Fowler *received the B.S. degree in computer and information science engineering and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University. He is currently a William L. Giles Distinguished Professor at Mississippi State University. He was previously Editor-in-Chief of IEEE Signal Processing Letters and is currently the Chair of the Computational Imaging Technical Committee of the IEEE Signal Processing Society. He is a Fellow of the IEEE.*

Agnieszka Miguel *received her Ph.D. in electrical engineering from the University of Washington, and MSEE and BSEE degrees from Florida Atlantic University. She is currently Associate Professor and Chair of the Department of Electrical and Computer Engineering at Seattle University. She is the American Society for Engineering Education Vice President for External Relations, a Vice Chair of the Electrical and Computer Engineering Department Heads Association, and the Conference Chair of the 2022 ECEDHA Annual Conference.*