COMPRESSIVE IMAGE RECOVERY USING RECURRENT GENERATIVE MODEL

Akshat Dave, Anil Kumar Vadathya, Kaushik Mitra

Department of Electrical Engineering
Indian Institute of Technology Madras, Chennai, India

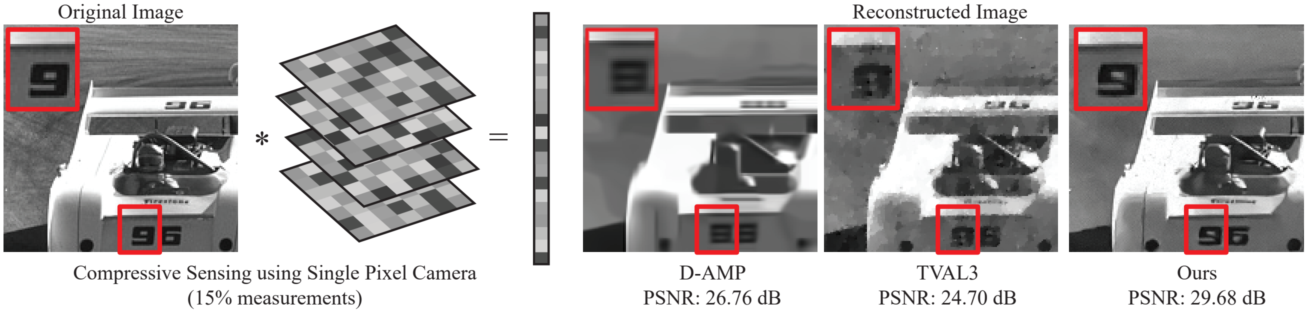


Fig. 1: We propose to use a deep generative model, RIDE [1] , as an image prior for compressive signal recovery. Since RIDE models long-range dependency in images using spatial LSTM, image recovery is better than other competing methods.

ABSTRACT

Reconstruction of signals from compressively sensed measurements is an ill-posed problem. In this paper, we leverage the recurrent generative model, RIDE, as an image prior for compressive image reconstruction. Recurrent networks can model long-range dependencies in images and hence can handle global multiplexing in compressive imaging. We perform MAP inference with RIDE using back-propagation to the inputs and projected gradient method. We propose an entropy thresholding based approach for preserving texture in images well. Our approach shows superior reconstructions compared to recent global reconstruction approaches like D-AMP and TVAL3 on both simulated and real data.

***Index Terms*—** Compressive imaging, generative models, deep learning, LSTMs, MAP inference

1. INTRODUCTION

Imaging in the non-visible region of the spectrum has a plethora of applications owing to its unique properties [2]. For example, improved penetration of infrared waves through fog and smog enables imaging through scattering media. However, prohibitive sensing costs in the non-visible range have limited its widespread use[1]. Many works have proposed Compressive Sensing (CS) [3, 4] as a viable solution for high-resolution imaging beyond the visible range of spectrum [5, 6, 7]. Compressive sensing theory states that signals exhibiting sparsity in some transform domain can be reconstructed from much lower measurements than sampling at Nyquist rate [4]. Lesser the number of measurements lesser is the cost of sensing. The single-pixel camera (SPC) is a classical example of CS framework [5]. In SPC, a single photo diode is used to sequentially capture compressive measurements and then reconstruct back the whole scene.

A challenge faced by CS reconstruction algorithms is to recover a high dimensional signal from a small number of measurements. This ill-posed nature of the reconstruction makes data priors essential. Often, visual signals exhibit sparse structure in some transform domain (wavelets, DCT coefficients or gradients). Initially, reconstruction methods exploited this prior knowledge about the signal structure thereby restricting the solution set to desired signal [8, 6, 7, 9]. However, using these simple priors at very low measurement rates results in low-quality reconstructions (see TVAL3 reconstruction in fig. 1). This is due to their inability to capture the complexity of natural image statistics. On the other hand, data-driven approaches have been proposed recently to handle the complexity [10, 11, 12]. They led to successful results in terms of reconstruction. But all of these approaches handle only local multiplexing i.e measurements are taken from image patches and recovery is also done patch wise. This is not appealing for classical SPC framework as such since measurements are acquired through global multiplexing.

To address these problems, in this work we propose to use a data-driven global image prior, RIDE, proposed by Theis et al. [1] for CS image recovery. RIDE uses recurrent networks

[1] Megapixel sensors in short-wave infrared, typically constructed using InGaAs, cost more than USD 100k.

with Long Short Term Memory (LSTM) units and is shown to model the long-term dependencies in images very well. Also, being recurrent it is not limited to patch size, hence can handle the global multiplexing in SPC. Our contributions are as follows:

- We propose to use RIDE as an image prior to model long-term dependencies for reconstructing compressively sensed images.
- We hypothesize that the model's uncertainty in prediction can be related to the entropy of component posterior probabilities. By thresholding the entropy, we enhance texture preserving ability of the model.

2. RELATED WORK

Single Pixel Camera: Signal reconstruction from CS measurements is an ill-posed problem and hence we need to use signal priors. Initially, algorithms were proposed minimizing the l_1 norm of signal assuming sparsity in the domain of wavelet coefficients [6], DCT coefficients or gradients. Most widely used minimum total-variation (TVAL3) [8] prior is based on piece-wise linear variation in natural images [7, 9]. Later class of algorithms known as approximate message passing (AMP) algorithms [13, 14] use off-the-shelf denoiser to iteratively refine their solution. ReconNet is another recent method using deep neural nets [11]. But as mentioned earlier, it can only handle local multiplexing since it is a patch based approach which is not suitable for SPC reconstruction.

Deep Generative Models: Recently advances with deep neural nets have led to powerful deep generative models. These include Generative Adversarial Nets (GAN) [15], Variational Auto Encoders (VAE) [16], Pixel Recurrent Neural Networks (PixelRNN) [17] and Recurrent Image Density Estimator (RIDE) [1]. Among these contemporary models, we find RIDE particularly suitable as a low-level image prior for our tasks involving Bayesian inference. GANs don't model the data distribution and VAE doesn't provide the exact likelihood measure. PixelRNN although models the probability distribution, it discretizes the distribution of a pixel to 256 intensity values resulting in optimization difficulties. Unlike these approaches, RIDE models a continuous distribution and provides exact likelihood, thus facilitating gradient based techniques for Bayesian inference.

3. BACKGROUND: RIDE

Let $\mathbf{x}$ be a gray scale-image and x_{ij} be the pixel intensity at location ij then $\mathbf{x}_{<ij}$ describes the causal context around that pixel containing all x_{mn} such that $m \leq i$ and $j < n$. Now the joint distribution over the image can be factorized as,

$$p(\mathbf{x}) = \prod_{ij} p(x_{ij}|\mathbf{x}_{<ij}), \tag{1}$$

Natural images, in general, exhibit long range correlations. In order to model such dependencies Theis et al. [1] have proposed to use two dimensional Spatial LSTM units [18]. Spatial LSTMs summarize the entire causal context $\mathbf{x}_{<ij}$ through their hidden representation $\mathbf{h}_{ij}$, as $\mathbf{h}_{ij} = f(\mathbf{x}_{<ij}, \mathbf{h}_{i-1,j}, \mathbf{h}_{i,j-1})$, where f is a complex non linear function. f has memory elements analogous to physical read, write and erase operations thus enabling LSTMs to model long term dependencies in sequences. Now each factor in the above joint distribution is modeled using conditional Gaussian Scale Mixtures, thus the complete distribution is given by,

$$\begin{aligned} p(\mathbf{x}) &= \prod_{ij} p(x_{ij}|\mathbf{h}_{ij}, \boldsymbol{\theta}), & (2)\\ p(x_{ij}|\mathbf{h}_{ij}, \boldsymbol{\theta}) &= \sum_{c,s} p(c, s|\mathbf{h}_{ij}, \boldsymbol{\theta}) p(x_{ij}|\mathbf{h}_{ij}, c, s, \boldsymbol{\theta}). & (3)\end{aligned}$$

For more details, we recommend the reader to go through [1].

4. COMPRESSIVE IMAGE RECOVERY USING RIDE

Here we consider the problem of image restoration, $\mathbf{x} \in \mathbb{R}^N$, from linearly compressed measurements, $\mathbf{y} \in \mathbb{R}^M$ as $\mathbf{y} = \Phi\mathbf{x}+\mathbf{n}$, where $\Phi \in \mathbb{R}^{M\times N}$ is the sensing matrix with $M < N$ and $\mathbf{n} \in \mathbb{R}^M$ is noise in the observation with known statistics.

4.1. MAP Inference via Backpropagation

Compressive image recovery: Here, we use Maximum-A-Posteriori principle to find the desired image as $\hat{\mathbf{x}} = arg\max_{\mathbf{x}} p(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})$. For SPC, we formulated the MAP inference as:

$$\hat{\mathbf{x}} = arg\max_{\mathbf{x}} \log p(\mathbf{x}) \;\; s.t. \;\; \mathbf{y} = \Phi\mathbf{x}. \tag{4}$$

Here we do reconstruction for the noise less case. The log-likelihood and gradients are given by the model as in Eqns. 2, 3 and 7. To optimize the above we use projected gradients method, where after each gradient update solution is projected back on to the affine solution space for $\mathbf{y} = \Phi\mathbf{x}$. Every k-th iteration consists of the following two steps,

$$\begin{aligned} \hat{\mathbf{x}}_k &= \mathbf{x}_{k-1} + \eta\nabla_{\mathbf{x}_{k-1}} \log p(\mathbf{x}), & (5)\\ \mathbf{x}_k &= \hat{\mathbf{x}}_k - \Phi^T \left(\Phi\Phi^T\right)^{-1} (\Phi\hat{\mathbf{x}}_k - \mathbf{y}). & (6)\end{aligned}$$

The gradient with respect to the log prior is given by,

$$\frac{\partial \log p(\mathbf{x})}{\partial x_{ij}} = \sum_{k\geq i, l\geq j} \frac{\partial \log p(x_{kl}|\mathbf{h}_{kl}, \boldsymbol{\theta})}{\partial x_{ij}}, \tag{7}$$

due to the recurrent nature of the model, each pixel through its hidden representation contributes to the likelihood of all the pixels that come after it in forward pass. Hence, during backward pass the gradient from each pixel propagates to all the pixels prior to it in the sequence.

Image inpainting: In image inpainting our goal is to recover the missing pixels from a randomly masked image. Here, we consider this as the special case of compressive imaging, where Φ can be written as row orthogonal binary matrix. The structure of Φ depends on the mask. The iterative updates (5) and (6) here simplify to gradient ascent of the prior over missing pixels while keeping the observed pixels constant. We have included the proof for this in our supplementary material (Appendix A).

In all of our experiments, we consider row orthonormalized Φ and the term $\left(\Phi\Phi^T\right)^{-1}$ reduces to identity matrix.

4.2. Tricks used for inference

4.2.1. Four directions

Joint distribution in eqn. (1) can be factorized in multiple ways, for example along each of the four diagonal directions of an image, i.e., top-right, top-left, bottom-right and bottom-left. Gradients from different factorizations are considered at each iteration of the inference, by flipping the image in the corresponding direction. This leads to twice faster convergence as compared to just considering one direction.

4.2.2. Entropy-based Thresholding

While solving the MAP optimization, we observed that we can recover the edges quite well but texture regions are blurred. This happens because the RIDE model may not have the right mixture component (see Eqn. (3)) to explain the latent texture at ij. In such cases, all the mixture components can be chosen with almost uniform probability, resulting in blurred texture. Aditionally, this results in high entropy over component probabilities $p(c, s|\mathbf{x}_{<ij}, x_{ij})$. On the contrary, if the point ij lies on an edge, then the entropy is low as there are only certain selected components which can explain that edge. Thus entropy, $H(i, j)$, serves as a surrogacy for model's confidence at that point. Hence, to prevent blurring of texture we set the gradients at ij to zero if $H(i, j)$ exceeds a threshold value. Figure 2 shows the effect of entropy constraint on the texture reconstruction.

$$H(i,j) = -\sum_{c,s} p(c, s|\mathbf{x}_{<ij}, x_{ij}) \log(p(c, s|\mathbf{x}_{<ij}, x_{ij})).$$

5. EXPERIMENTS

For training the RIDE model we have used publicly available Berkeley Segmentation dataset (BSDS300). Following the instincts from [1], we trained the model with increasing patch size in each epoch. Starting with 8x8 patch we go till 22x22 in steps of 2 for 8 epochs. We used the code provided by authors of RIDE in caffe. We start with a learning rate of 0.0001 and decrease it to half the previous value after every epoch. We used Adam method [19] for training the model. We find RIDE with one layer SLSTM gives satisfactory results for all our inference tasks. Also, we have used entropy based gradient thresholding (sec. 4.2.2) with threshold 3.5, to avoid blurring the texture regions in all the experiments. In order to accommodate for boundary issues, we remove a two-pixel neighborhood around the image for PSNR and SSIM calculations in all the experiments. For a fair comparison, we do the same for the reconstructions of TVAL3 [8] and D-AMP [14].

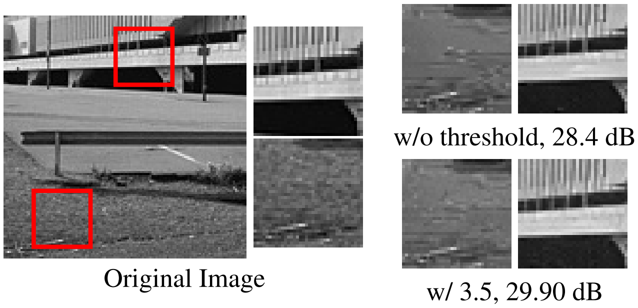


Fig. 2: Compressive image reconstructions from 30% measurements obtained with entropy thresholding. The texture of the grass in magnified patch is recovered better with the threshold 3.5. The PSNR values mentioned are for whole reconstructed image.

5.1. Image Inpainting

For image inpainting, we randomly removed 70% of pixels and estimated them using aforementioned inference method. We compared our approach with the multiscale adaptive dictionary learning approach [20], which is an improvement over the KSVD algorithm, see Figure 3. It is clear from the figure that our approach is able to recover the sharp edges better than the multiscale KSVD approach. This is because our method is based on the global image prior as compared to the patch-based multiscale KSVD approach.

5.2. Single-Pixel Camera

In general, the SPC framework involves global multiplexing of the scene. Our model, using Spatial LSTMs, can reason for long-term dependencies in image sequences and is preferable for such kind of tasks. Unlike the recent state-of-the-art methods, like ReconNet, which are designed for local spatial multiplexing and can't handle the global multiplexing case directly. We show SPC reconstruction results on some randomly chosen images from the BSDS300 test set which were cropped to 160 × 160 size for computational feasibility, see figures on top of table 1. We generate compressive measurements from them using random Gaussian measurement matrix with orthonormalized rows. We take measurements at three different rates 0.3, 0.25 and 0.15. Using the projected gradient method, we perform gradient ascent for 300 iterations for 0.3 and 0.25 measurement rates. For lower measurement rates, we run gradient ascent for 500 iterations. In all the cases, we start with a random image uniformly sampled from $(0, 1)$. Reconstruction results for five images are shown in Table 1 and Figure 1. We were able to show improvements both in terms of PSNR and SSIM values for different measurement rates. Even at low measurement rates, our method preserves the sharp and prominent structure in the image, qualitative results are included in the supplementary material. D-AMP has the tendency to over-smooth the image, whereas TVAL3 adds blotches to even the smooth parts of the image.

Real Image Reconstruction: Here we consider real measurements acquired from SPC using Fast Walsh Hadamard transform (FWHT) as Φ. Figure 4 shows the reconstructions at 15%. It can be observed that our method provides superior reconstructions similar to the simulated case. Since we dont have original image here, we take reconstruction from D-AMP at 100% measurements as the ground truth. Using this we evaluate PSNR and SSIM.

6. CONCLUSIONS AND FUTURE WORK

We demonstrate that recurrent generative image models such as RIDE can be used effectively for solving compressive image recovery problems. The main advantage of using such recurrent generative models is that they are global priors and hence can model long term image dependencies. In future work, we would like to improve the texture preserving ability of the model. Also, This approach can be extended to other image restoration tasks such as image deblurring, superresolution etc. and to computational photography problems.

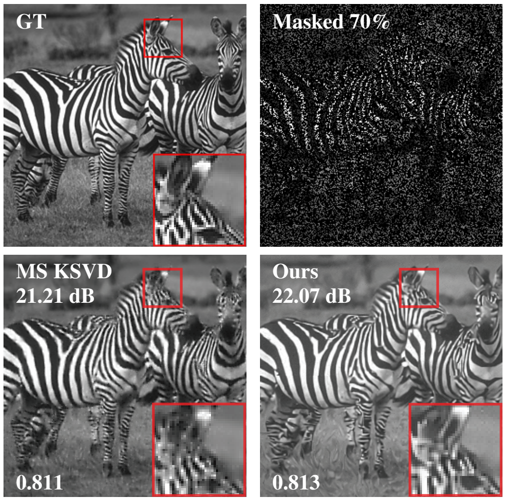
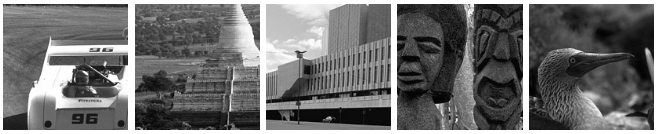


Fig. 3: Inpainting comparisons with multiscale(MS) KSVD [20]. (GT: Ground Truth)



S. No	Method	M.R.=30%		M.R.=25%		M.R.=15%	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	TVAL3	29.54	0.833	28.40	0.802	26.76	0.736
	D-AMP	31.59	0.867	29.70	0.836	24.70	0.716
	Ours	**34.24**	**0.901**	**32.91**	**0.868**	**29.58**	**0.776**
2	TVAL3	26.65	0.726	25.86	0.6810	24.51	**0.575**
	D-AMP	26.08	0.674	25.44	0.633	23.67	0.497
	Ours	**29.73**	**0.809**	**28.78**	**0.766**	**24.93**	0.543
3	TVAL3	26.20	0.789	25.08	0.745	22.63	0.638
	D-AMP	31.66	0.923	29.09	0.883	24.5	0.757
	Ours	**33.82**	**0.935**	**32.21**	**0.913**	**27.60**	**0.816**
4	TVAL3	**26.80**	0.717	26.08	0.675	24.13	0.563
	D-AMP	25.84	0.614	25.27	0.584	23.72	0.498
	Ours	26.59	**0.742**	**26.12**	**0.711**	**24.14**	**0.599**
5	TVAL3	30.20	0.849	29.10	0.820	25.44	0.719
	D-AMP	30.17	0.858	28.06	0.796	25.81	0.70
	Ours	**35.19**	**0.922**	**33.52**	**0.892**	**29.30**	**0.786**
Mean	TVAL3	27.858	0.783	26.90	0.745	24.69	0.646
	D-AMP	29.07	0.787	27.51	0.7464	24.48	0.633
	Ours	**31.91**	**0.862**	**30.71**	**0.830**	**27.11**	**0.704**

Table 1: Comparisons of compressive imaging reconstructions at different measurement rates for the images shown above (numbering of images are from left to right). Our method outperforms the existing global prior based methods in most of the cases.

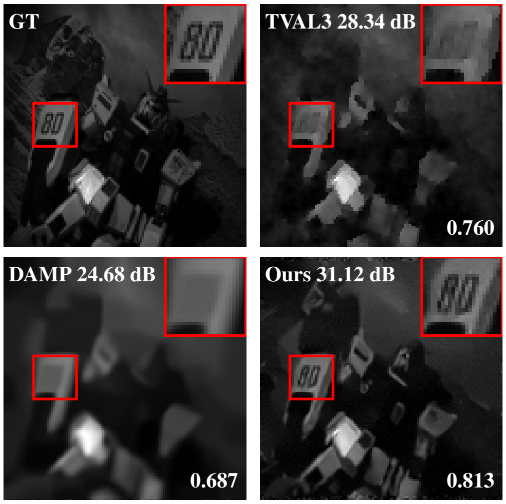


Fig. 4: Real SPC reconstructions at 15% compression, our approach recovers the details better than others in real case also. Ground Truth (GT) is obtained from 100% reconstruction.

Acknowledgements: This work is supported by Qualcomm Innovation Fellowship India 2016. The authors wish to thank ILDS lab of CS department at IIT Madras for sharing their computational resources.

7. REFERENCES

[1] Lucas Theis and Matthias Bethge, “Generative image modeling using spatial lstms,” in *Advances in Neural Information Processing Systems*, 2015, pp. 1927–1935.

[2] Marc P Hansen and Douglas S Malchow, “Overview of swir detectors, cameras, and applications,” in *SPIE Defense and Security Symposium*. International Society for Optics and Photonics, 2008, pp. 69390I–69390I.

[3] Richard G Baraniuk, “Compressive sensing,” *IEEE signal processing magazine*, vol. 24, no. 4, 2007.

[4] David L Donoho, “Compressed sensing,” *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin E Kelly, Richard G Baraniuk, et al., “Single-pixel imaging via compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83, 2008.

[6] Aswin C Sankaranarayanan, Christoph Studer, and Richard G Baraniuk, “Cs-muvi: Video compressive sensing for spatial-multiplexing cameras,” in *Computational Photography (ICCP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1–10.

[7] Huaijin Chen, M Salman Asif, Aswin C Sankaranarayanan, and Ashok Veeraraghavan, “Fpa-cs: Focal plane array-based compressive imaging in short-wave infrared,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2358–2366.

[8] Chengbo Li, Wotao Yin, Hong Jiang, and Yin Zhang, “An efficient augmented lagrangian method with applications to total variation minimization,” *Computational Optimization and Applications*, vol. 56, no. 3, pp. 507–530, 2013.

[9] Jian Wang, Mohit Gupta, and Aswin C Sankaranarayanan, “Lisens-a scalable architecture for video compressive sensing,” in *Computational Photography (ICCP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–9.

[10] Mohammad Aghagolzadeh and Hayder Radha, “Compressive dictionary learning for image recovery,” in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 661–664.

[11] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok, “Reconnet: Non-iterative reconstruction of images from compressively sensed measurements,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 449–458.

[12] Ali Mousavi, Ankit B Patel, and Richard G Baraniuk, “A deep learning approach to structured signal recovery,” in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 1336–1343.

[13] David L Donoho, Arian Maleki, and Andrea Montanari, “Message-passing algorithms for compressed sensing,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.

[14] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk, “From denoising to compressed sensing,” 2014.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[16] Diederik P Kingma and Max Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.

[17] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu, “Pixel recurrent neural networks,” *arXiv preprint arXiv:1601.06759*, 2016.

[18] Alex Graves, “Neural networks,” in *Supervised Sequence Labelling with Recurrent Neural Networks*, pp. 15–35. Springer, 2012.

[19] Diederik Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[20] Julien Mairal, Guillermo Sapiro, and Michael Elad, “Learning multiscale sparse representations for image and video restoration,” *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.