Statistics of RGBD Images

Yair Weiss and Dan Rosenbaum
School of Computer Science and Engineering
The Hebrew University of Jerusalem

ICRI-CI Retreat
Outline

▶ The 3D camera: revolution and challenges.
▶ What have we learned from ordinary image restoration.
▶ Statistics of RGBD.
The 3D camera revolution

- High quality depth cameras for under $25.
- Considered impossible only 10 years ago (G. Medioni, personal communication).
- Dramatic change in how we do robotics and vision.
Challenges in RGBD imaging

Compared to RGB the D channel has:

- Missing data.
- Low resolution.
- Noise.
Current Approaches (single frame)

Assume that depth edges correlate with color edges.

(Silverman et al. 2012). Similar results with joint bilateral filters.
Colorization Using Optimization (Levin et al. 2004)

\[ J(d) = \sum_{ij} w_{ij}(d_i - d_j)^2 \]

\[ w_{ij} = e^{-\|l_i - l_j\|^2} \]
Current Approaches

Assume that depth edges correlate with color edges.

Use the colorization cost to fill in holes in the D channel. This is state-of-the-art. Can we do better with learning?
Isn’t the answer obvious?

Train deep network to predict D from RGB?
What did we learn from RGB restoration?

- Learning based approaches can give the best performance. This requires rich models with many free parameters and a lot of training data.

- Both deep learning approaches and generative (GMM) approaches give excellent performance given sufficient training data.

- Generative approaches can be used for multiple image degradation problems. Deep learning is very specific for a particular problem.
Noisy
Gaussian mixtures for image patches

\[
\Pr(x) = \sum_h \pi_h \mathcal{N}(x; \mu_h, \Sigma_h)
\]
Gaussian mixtures for image patches

\[ \Pr(x) = \sum_h \pi_h \mathcal{N}(x; \mu_h, \Sigma_h) \]

(Zoran and Weiss 2011)
A Convolutional Neural Network approach to Denoising
(Jain and Seung 2007)

Figure 1: Architecture of convolutional network used for denoising. The network has 4 hidden layers and 24 feature maps in each hidden layer. In layers 2, 3, and 4, each feature map is connected to 8 randomly chosen feature maps in the previous layer. Each arrow represents a single convolution associated with a $5 \times 5$ filter, and hence this network has 15,697 free parameters and requires 624 convolutions to process its forward pass.

Results better than FOE. Not as good as GMM trained on same data.
By increasing the number of training images and free parameters gives similar performance to GMM (slightly better) with much smaller number of training images.
Noisy
Denoised with Deep Architecture
Denoised with GMM
In GMM, you learn once \( \Pr(x) \) and then use Bayes rule to apply to any image degradation. Backprop is done separately for each image degradation.
Denoising with learned GMM
Deblurring with the same learned GMM
What did we learn from ordinary image restoration?

- Learning based approaches can give the best performance. This requires rich models with many free parameters and a lot of training data.
- Both deep learning approaches and generative (GMM) approaches give excellent performance given sufficient training data.
- Generative approaches can be used for multiple image degradation problems. Deep learning is very specific for a particular problem.
Sintel synthetic dataset

Generate millions of patches of RGBD and use them to train generative models of patches.
Generative models of depth patches

Ground truth  GMM

Patches are either flat or edges (duh!). Much stronger assumption than L1 or L2 smoothness.
Generative models of disparity patches given intensity patches

Disparity edges tend to co-occur with intensity edges (basis of all existing algorithms).
Learned models outperform handcrafted ones.
Very little improvement when conditioning on the data.
Optimal patch restoration shows same trend.
What’s going on?

Conditioning on intensity only gives boost in performance in “large holes”. These are rare in the image but visually salient.
Our Approach

- Use learned models plus Bayes optimal restoration on all patches except large holes.
- For large holes (where Bayesian optimality is intractable) revert to colorization.
- Room for improvement.
Results

noisy intensity

GT disparity

noisy disparity

LRC (41.38)  

DL2|int (37.89)  

Our approach (42.22)
Results

noisy intensity | GT disparity | noisy disparity

LRC (36.44) | DL2|int (35.88) | Our approach (37.96)

Outperforms state-of-the-art on standard benchmark.
Conclusion

- RGBD images present great opportunity for image restoration.
- Statistical models trained on synthetic data learns that (1) patches are either flat or edges and (2) depth edges tend to correlate with color edges.
- Correlation between color edges and depth edges (basis of all handcrafted methods) is a relatively weak cue.
- Learned statistical model gives best existing method for denoising RGBD on standard benchmark even when trained on synthetic data.