YONSEI UNIVERSITY

Microsoft Research

Unified Depth Prediction and Intrinsic Image Decomposition from a Single Image via Joint Convolutional Neural Fields

Seungryong Kim[1], Kihong Park[1], Kwanghoon Sohn[1], and Stephen Lin[2]

[1]Yonsei University, [2]Microsoft Research

Introduction

❖ Goal

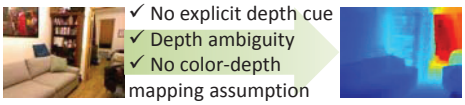
- To jointly predict depth and intrinsic images from a single-image

Color Depth Albedo Shading

- Two tasks are formulated in a synergistic manner though CRF+CNN

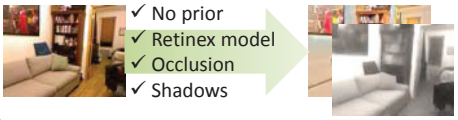
Challenging

❖ Depth Prediction



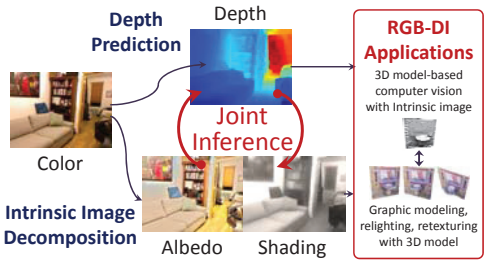
✓ No explicit depth cue
✓ Depth ambiguity
✓ No color-depth mapping assumption

❖ Intrinsic Image Decomposition



✓ No prior
✓ Retinex model
✓ Occlusion
✓ Shadows

Our Solution



Formulation

❖ Key-Insights

- Correlations are stronger among gradient domain

❖ Joint Conditional Random Field (CRF)

- Energy Function

$$\mathbf{E}(D,A,S|I) = \mathbf{E}_u(D|I) + \mathbf{E}_u(A,S|I) + \lambda_D \mathbf{E}_s(D|I,A,S) + \lambda_A \mathbf{E}_s(A|I,D,S) + \lambda_S \mathbf{E}_s(S|I,D,A)$$

- Unary Potentials

$$\mathbf{E}_u(D|I) = \sum_p (D_p - \mathcal{F}(I_{\mathcal{P}};\mathbf{w}_{\mathcal{F}}^D))^2 \rightarrow \text{Depth Value}$$

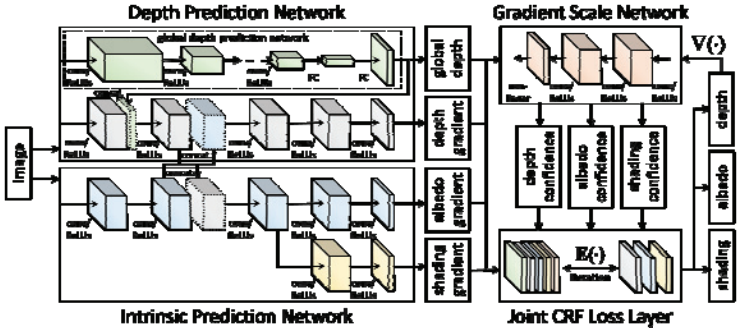
$$\mathbf{E}_u(A,S|I) = \sum_p (L_p(I_p - A_p - S_p))^2$$

- Pairwise Potentials

$$\mathbf{E}_s(D|I,A,S) = \sum_p \|\nabla D_p - \mathcal{G}(\nabla I_{\mathcal{P}}, \nabla A_{\mathcal{P}}, \nabla S_{\mathcal{P}}; \mathbf{w}_{\mathcal{G}}^{\nabla D}) \circ \mathcal{F}(I_{\mathcal{P}}; \mathbf{w}_{\mathcal{F}}^{\nabla D})\|^2$$

Confidence of Gradient — Depth Gradient

❖ Network Configuration



1) Joint Depth and Intrinsic Prediction Network

- Depth Prediction Network: Global Depth + Depth Gradient
- Intrinsic Prediction Network: Albedo & Shading Gradient

2) Gradient Scale Network

- Confidence of Estimated Gradients
- Color/Intrinsic Gradient → Depth Gradient Confidence

	global depth net.							gradient scale net.		
	conv1	conv2	conv3	conv4	conv5	FC1	FC2	conv1	conv2	conv3
kernel	11×11	5×5	3×3	3×3	3×3	1×1	1×1	3×3	3×3	1×1
channel	96	256	384	384	256	4096	-/16	64	64	2 or 6
stride	4	1	1	1	1	1	1	1	1	1
pad	4	2	1	1	0	0	0	0	0	0
pooling	max	max	-	-	max	-	-	-	-	-
non-lin.	ReLUs	ReLUs	ReLUs	ReLUs	ReLUs	ReLUs	ReLUs	ReLUs	ReLUs	$f(\cdot)$

	depth gradient net.					intrinsic gradient net.				
	conv1	conv2	conv3	conv4	conv5	conv1	conv2	conv3	conv4	conv5
kernel	11×11	3×3	3×3	3×3	3×3	11×11	3×3	3×3	3×3	3×3
channel	96+1	64+64	64	64	2	96	64+64	64	64	6
stride	1	1	1	1	1	1	1	1	1	1
pad	0	0	0	0	0	0	0	0	0	0
pooling	-	-	-	-	-	-	-	-	-	-
non-lin.	ReLUs	ReLUs	ReLUs	ReLUs	-	ReLUs	ReLUs	ReLUs	ReLUs	-

Unified Depth and Intrinsic Image Prediction

❖ Training Procedure

- Global Depth Training

$$\mathcal{L}(\mathbf{w}_{\mathcal{F}}^D) = \sum_{\{i,p\}} (D_p^i - \mathcal{F}(I_{\mathcal{P}}^i; \mathbf{w}_{\mathcal{F}}^D))^2$$

- Gradient Training + Gradient Scale Training

$$\mathcal{L}(\mathbf{w}_{\mathcal{G}}^{\nabla D}, \mathbf{w}_{\mathcal{F}}^{\nabla D}) = \sum_{\{i,p\}} \|\nabla D_p^i - \mathcal{G}(\nabla I_{\mathcal{P}}^i, \nabla A_{\mathcal{P}}^i, \nabla S_{\mathcal{P}}^i; \mathbf{w}_{\mathcal{G}}^{\nabla D}) \circ \mathcal{F}(I_{\mathcal{P}}^i; \mathbf{w}_{\mathcal{F}}^{\nabla D})\|^2$$

Iterative Training

❖ Testing Procedure

- Iterative Joint Prediction

$$\mathbf{E}(D|I) = \sum_p (D_p - D_p^*)^2 + \lambda_D \sum_p \|\nabla D_p - C(\nabla D_p^*) \circ \nabla D_p^*\|^2$$

$$\mathbf{E}(A,S|I) = \sum_p (L_p(I_p - A_p - S_p))^2 + \sum_p \lambda_A \|\nabla A_p - C(\nabla A_p^*) \circ \nabla A_p^*\|^2 + \lambda_S \|\nabla S_p - C(\nabla S_p^*) \circ \nabla S_p^*\|^2$$

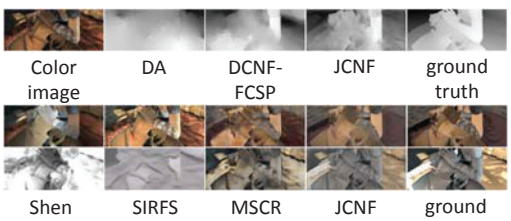
Joint Inference

Experimental Results and Discussion

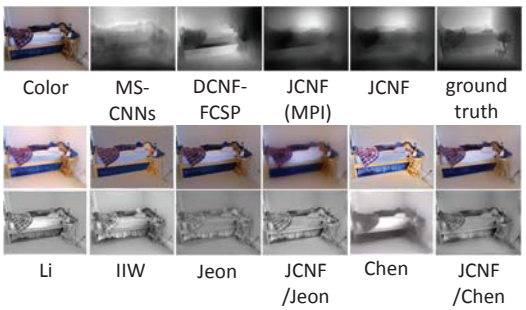
❖ Implementation Details

- VLFeat MatConvNet Toolbox

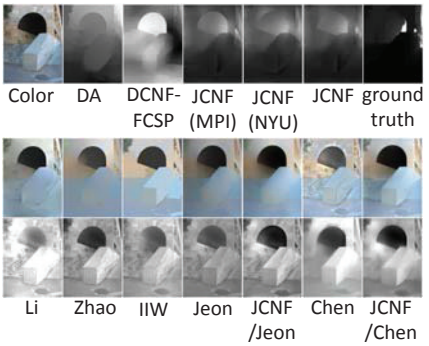
❖ MPI SINTEL Benchmark



❖ NYU v2 RGB-D Benchmark



❖ Make3D RGB-D Benchmark



Conclusion

❖ Joint Convolutional Neural Field (JCNF) Model

- To jointly predicting a depth map and intrinsic images from single-image input

Contact

- E-mail: srkim89@yonsei.ac.kr
- Homepage: http://diml.yonsei.ac.kr/~srkim/

European Conference on Computer Vision (ECCV) 2016