Revisiting Dead Leaves Model: Training with Synthetic Data

Pavan C Madhusudana, Seok-Jun Lee and Hamid R. Sheikh

Samsung Research America, USA

ICIP 2022, Bordeaux, France

Outline of the Talk

Outline of the Talk

1. Introduction
   - Challenges
   - Motivation
2. Proposed Method
3. Experiments and Results
4. Conclusion and Future Work

Introduction

- Supervised learning - deep neural networks
  - Excellent performance on many computer vision tasks - image classification, object detection etc.
  - Availability of large datasets - Imagenet, COCO etc.
- Current work - stereo disparity estimation problem
  - The goal is to estimate depth (disparity) using from images captured using two cameras with known distance between them.

Challenges

- Large scale labeled data - expensive and time consuming
- Stereo data
    - Synchronized capture of images and 3D scene
    - Careful registration of acquired stereo data
- Alternative - synthetic data
    - Inexpensive to generate - ideally infinite data can be generated.
    - Ground truth - easy to obtain
- Synthetic data - shortcomings
    - Lack sufficient realism - domain gap between real and synthetic data
    - Scenes - careful design of background, objects, shapes, color etc.

Dead Leaves Model - Motivation

- Objective - simplistic generation model with sufficient realism
- Synthetic data - similar statistics as seen in natural images
- Dead leaves model
  - Originally proposed by Matheron [1] to model occlusion of objects
  - Synthetic image - adding independent shapes in layered manner was shown to have similar statistics as natural images[2]
  - Recent work[3] - successfully used generated data to train model for image restoration tasks
- Current work - extend the idea to disparity estimation problem

[1] G. Matheron, "Mod'ele s´equentiel de partition al´eatoire," Centre de Morphologie Math´ematique, 1968.

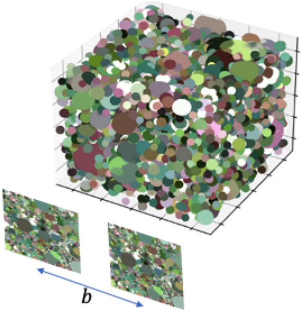
[2] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," International Journal of Computer Vision, vol. 41, no. 1, pp. 35–59, 2001.

[3] R. Achddou, Y. Gousseau, and S. Ladjal, "Synthetic images as a regularity prior for image restoration neural networks," in Eighth International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)., 2021.

Outline of the Talk

3D Dead Leaves Model

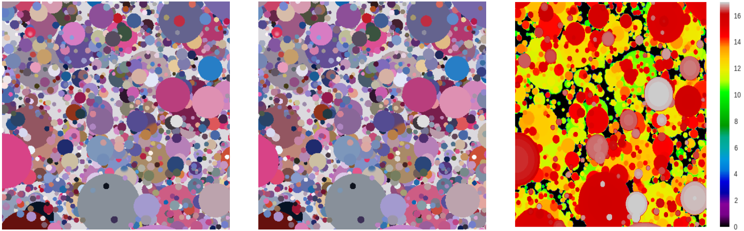
- 3D dead leaves space - opaque spheres with random radius
- The radii r of the spheres - sampled from distribution $f(r) = Kr^{-3}$
- Stereo images - projected onto parallel camera planes



(a) Setup

3D Dead Leaves Model

- Ground truth disparity - $d(x,y) = \frac{fb}{D(x,y)}$, f is camera's focal length, b is the baseline width between camera centres and $D(x,y)$ is the depth value at pixel (x,y)

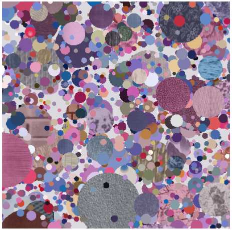


(b) left image (c) right image (d) disparity map

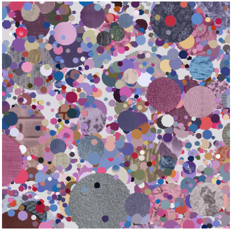
Figure: Sample dead leaves stereo data

Textured Dead leaves Model

- Adding textures
  - Statistics more closer to natural images
  - More efficient to determine corresponding points in stereo images
  - Improves image gradients in smooth regions
- Texture addition - blending with textures sampled from Brodatz[4] texture database
- Disparity values - not affected by texture addition



(a) left image

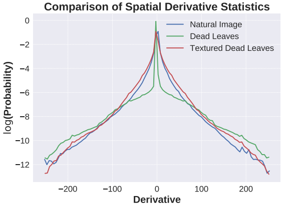


(b) right image

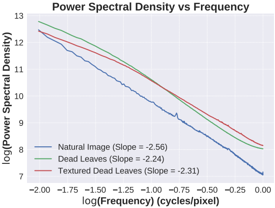
Figure: Sample textured dead leaves data

Significance of Dead Leaves Model

- Marginal and joint distributions of spatial derivatives - similar to natural images
- Power spectrum - exhibit $\frac{1}{f^2}$ observed in natural images



(a) Marginal Distribution



(b) Power Spectrum

Figure: Comparison of statistics of dead leaves with natural images

Outline of the Talk

Experiment Details

- Dataset generation
  - Each scene - 20,000 spheres projected on 1024×1024 resolution camera planes
  - Every scene - 3 focal length values and 9 baseline widths, total of 27 stereo pairs
  - 480 scenes - $27 \times 480 = 12,960$ size of generated dataset
- Model training
  - Architecture - Pyramid Stereo Matching network (PSMNet)[5]
  - 3250 iterations with batch size of 12
  - Training images - random crops of size 256×512
- Evaluation datasets - KITTI 2012, KITTI 2015 and Scene Flow.

[5] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5410–5418.

Performance Comparison

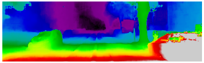
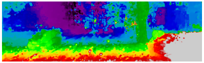
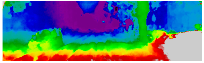
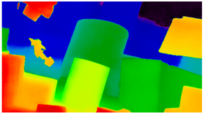
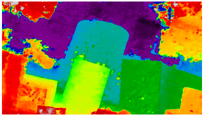
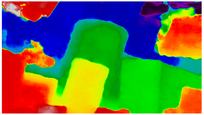
- Evaluation metric - End Point Error

$$EPE = \frac{1}{M}\sum_{i=1}^{M} |disp^i_{pred} - disp^i_{GT}|$$

- Model trained only on synthetic data - no additional fine-tuning on evaluation datasets
- Dead leaves trained models - comparable performance

Training Dataset	KITTI 2012	KITTI 2015	Scene Flow	
			Train	Test
Scene Flow	1.35	1.83	-	1.09
Dead Leaves	3.01	3.14	13.26	11.52
Textured Dead Leaves	3.38	2.29	9.97	8.3

Visual Comparison

RGB Image	Scene Flow	Dead Leaves	Textured Dead Leaves
			
	EPE $= 1.623$	EPE $= 2.452$	EPE $= 1.955$
			
	EPE $= 0.760$	EPE $= 4.681$	EPE $= 3.512$

- Texture addition - smoother and noise-free estimates

Outline of the Talk

1. Introduction
   - Challenges
   - Motivation
2. Proposed Method
3. Experiments and Results
4. Conclusion and Future Work

Conclusion and Future Work

- Dead leaves - simplistic and computationally inexpensive model, easy to generate large scale data
- Trained model - good generalizability on synthetic and real world data with no additional fine-tuning
- Future work
  - Synthetic images - pin-hole camera assumption
  - Inclusion real world effects - lens blur, shot noise etc. lead to statistics closer to natural images

References

[1] P. C. Madhusudana, Seok-Jun Lee and Hamid R. Sheikh, "Revisiting Dead Leaves Model: Training with Synthetic Data", in IEEE Signal Processing Letters, vol. 29, pp. 209-213, 2022.

[2] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5410–5418.

[3] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," International Journal of Computer Vision, vol. 41, no. 1, pp. 35–59, 2001.

Thank You!