Multi-view stereo estimation with pixel-wise view selection network

Xucheng Wan, Zhengyang Lou, Xiao Wang

5/2/2018

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

OUTLINE

- Introduction
  - Disparity
- Binocular Stereo
- Multi-view Stereo
  - View selection
  - Proposed method
- Experiments
- Reference

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Introduction

This project is aiming at:

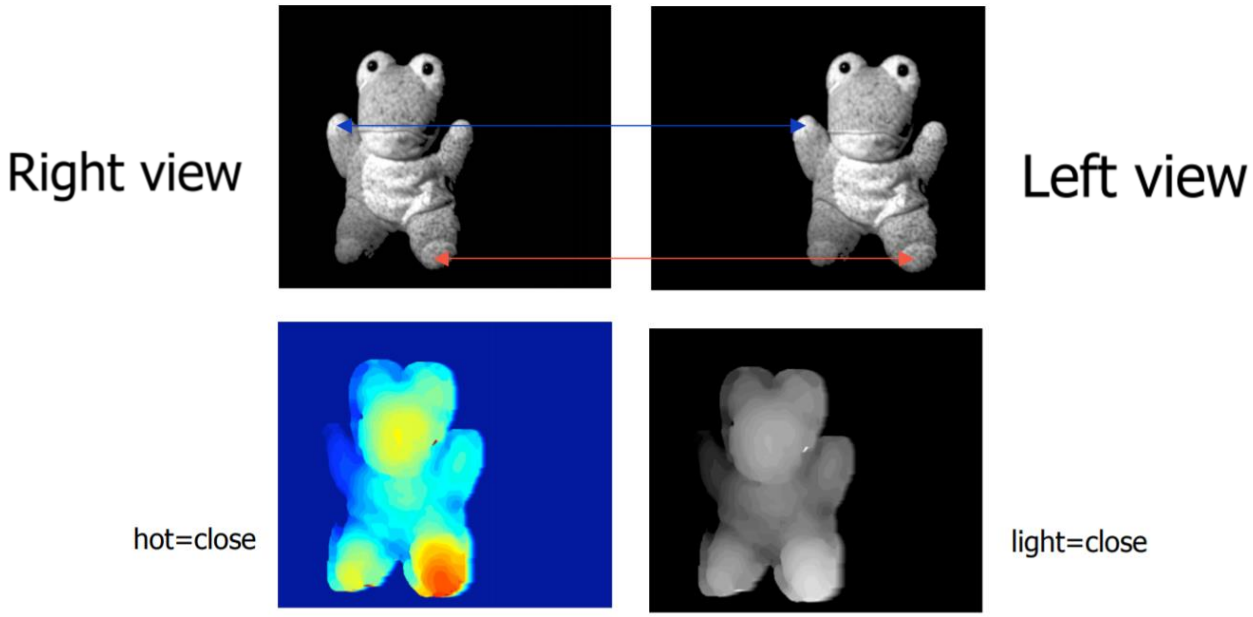
- Integrate state-of-art binocular disparity estimation algorithms to multiview disparity estimation.
- Implement neural networks to determine the views to use for every pixel.

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Why disparity estimation is important

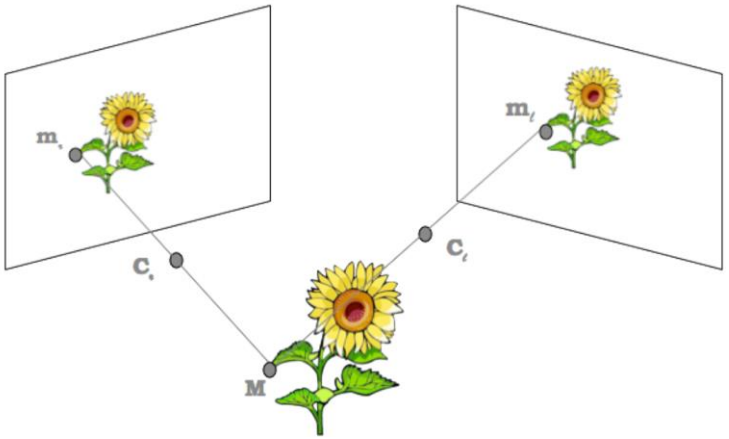
Stereo estimation is a fundamental computer vision:

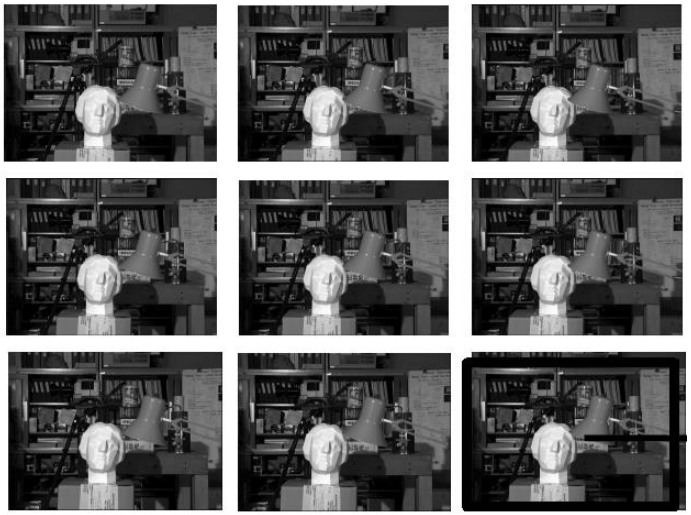
Given two images for the same scene from different views, compute the disparity for each pixel and then generate depth map.



Introduction of disparity

Binoculor v.s. Multi-view

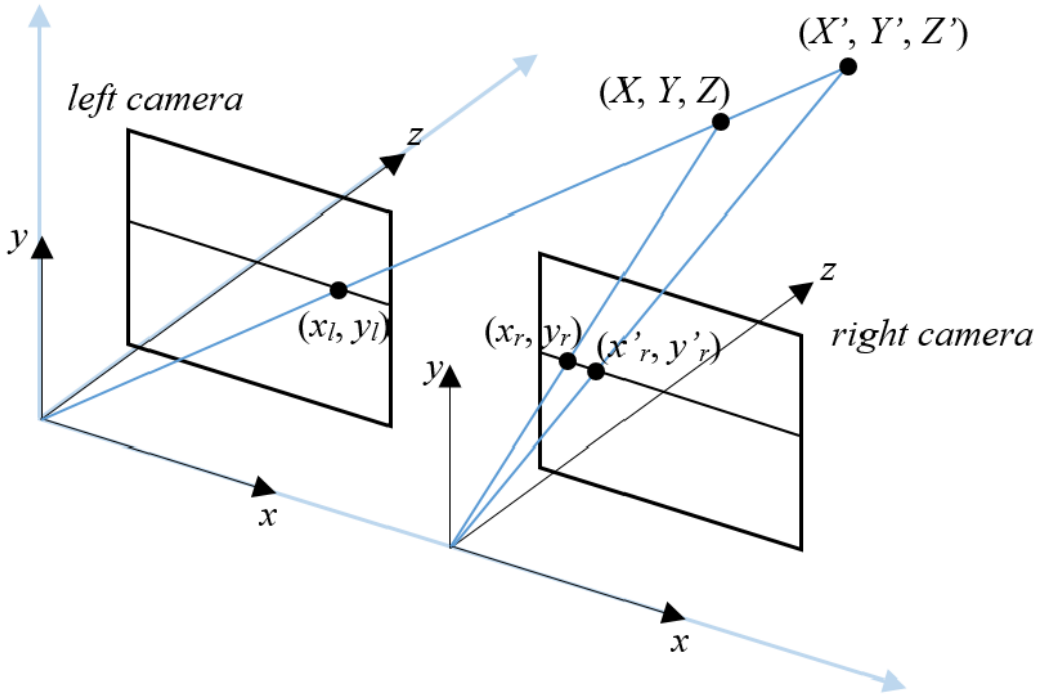




Binocular disparity is just 1-D estimation which may ignore some vertical information.

Multi-view disparity is the extension of binocular method at 2-D estimation, which uses more than two images.

Disparity in binocular

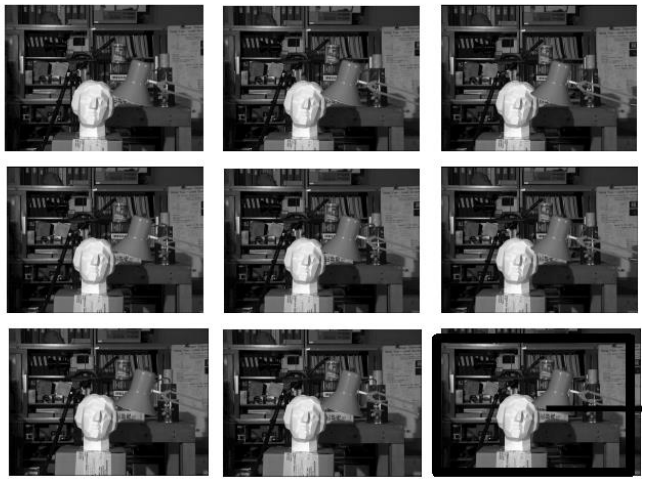


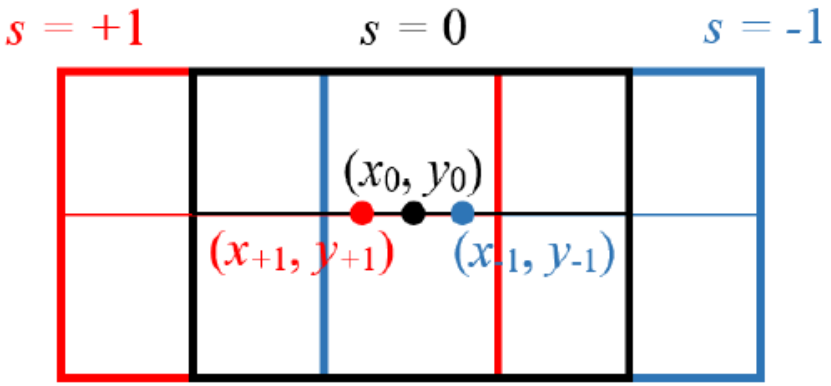
- The horizontal shift of the corresponding points in two images is called disparity.

$$x' = x + s\,d(x,y), \quad y' = y,$$

- S is a sign (±1) which ensures the disparity would always be positive.

Disparity in Multi-view





- s is used to denote the relative position between an image and the reference image(usually the center image)
- For each disparity d, the corresponding pixel intensity would be:

$$I_{s,t}^{d}(x,y) = I_{s,t}(x+(s-s_0)d, y+(t-t_0)d)$$

Basic Steps of disparity estimation

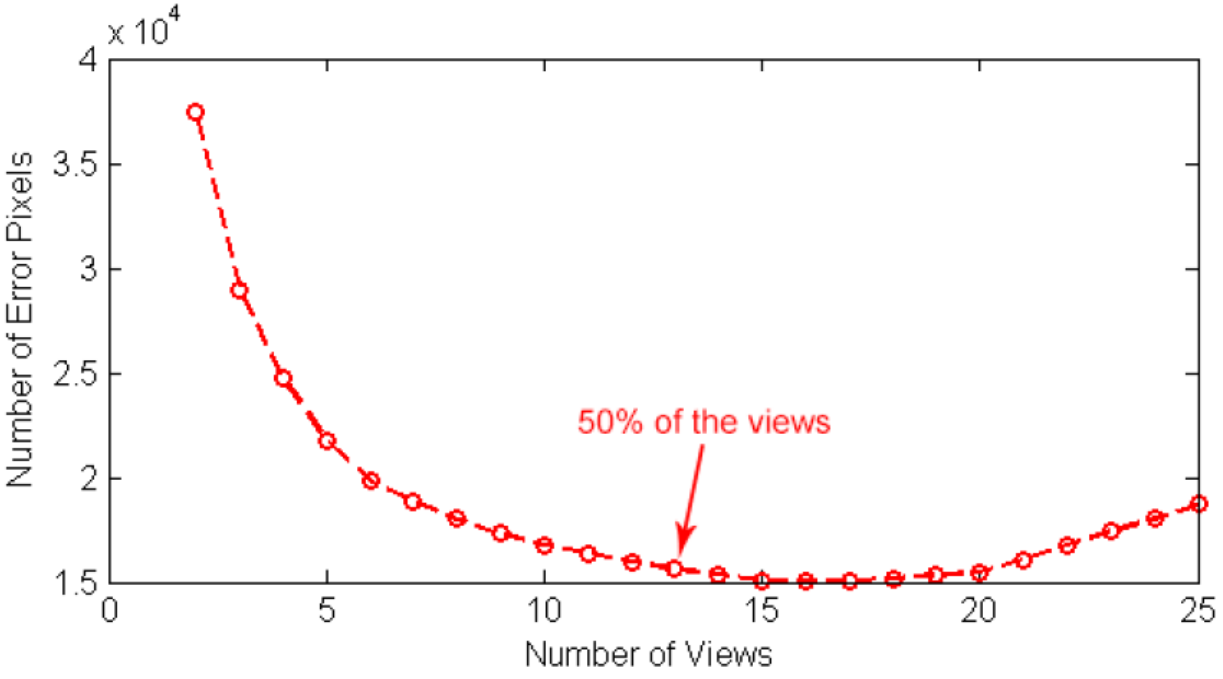
Basically, a stereo algorithm generally performs the following 4 steps[1]:

1. matching cost computation;
   - Relationship like distance between corresponding points
2. cost aggregation;
   - Smooth cost map
3. disparity computation / optimization;
   - To compute or predict the disparity for each pixel
4. disparity refinement.
   - Encourage discontinuity at edges of the an object
   - Encourage continuity at surface of an object

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

View Selection

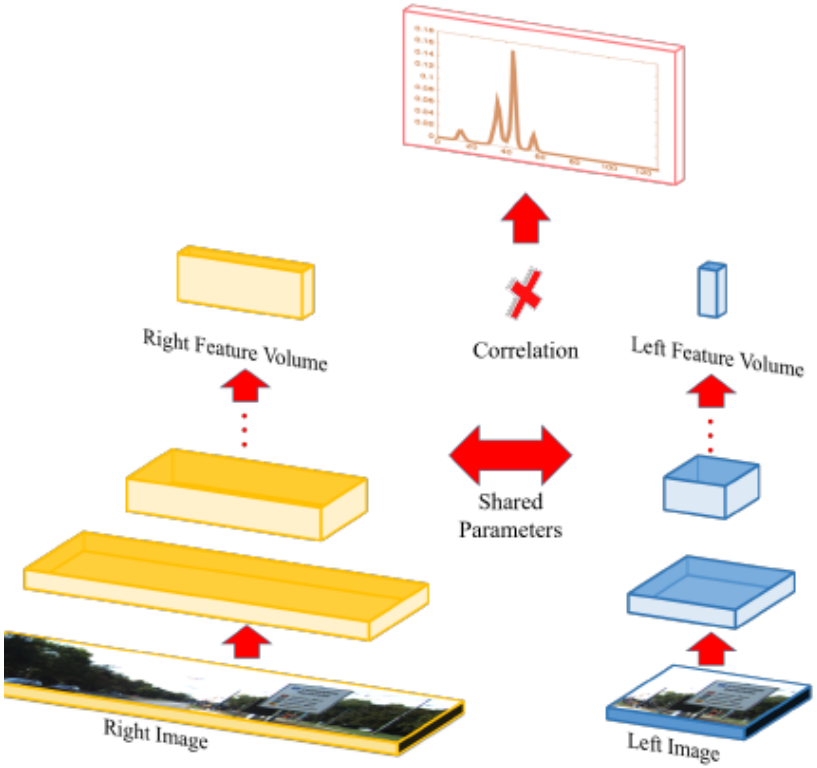
However, multi-view disparity estimating method often yields noisy, spurious disparity maps due to occlusions, scene discontinuity, imperfect light balance and other disturbance.
Only need to select some of the views to reach best performance.
**This is the problem we are going to solve.



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Efficient deep learning for stereo matching

[2] Luo, W., et al. (2016). Efficient deep learning for stereo matching. *In international conference of CVPR.*



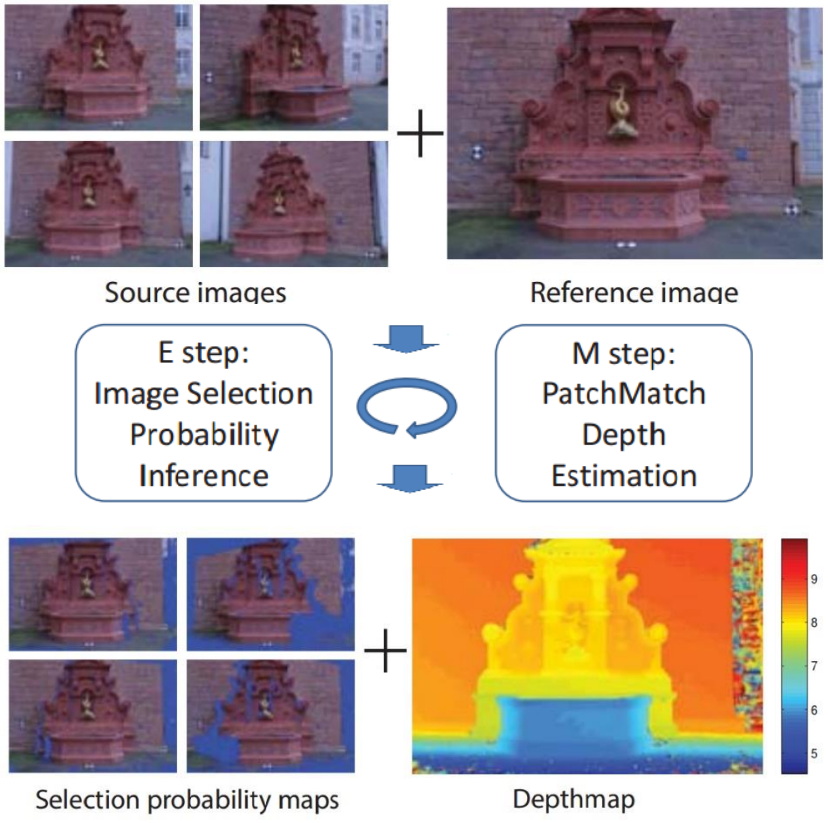
Contribution:

- Build a Siamese network to process two images from an image pair simultaneously.
- Each unary structure extracts features from the input image using the same parameters.
- Exploit a correlation layer which computes inner product between two representations of a siamese architecture.

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Joint view selection and depth map estimation

[3] Zheng, E., et al.(2014). PatchMatch Based Joint View Selection and Depth map Estimation. *In international conference of CVPR.*



Contribution:

- Posing the problem within a probabilistic framework that jointly models pixel-level view selection and depthmap estimation given the local pairwise image photoconsistency.
- The corresponding graphical model is solved by EM-based view selection probability inference and PatchMatch-like depth sampling and propagation.
- EM-propagation process, which is different from former work, is in a single direction so every pixel is determined by one of its neighbor.

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Proposed method 1

Fundation of this project:
[4]. Shiwei Zhou, et, al.(2018) Improving disparity map estimation for multi-view noisy images. *ICASSP 2018 conference.*

A disparity estimation method for multi-view images with noise is investigated by constructing multi-focus image and view selection

Assumption: disparity values are integers.

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

3D focus image stack (3DFIS)

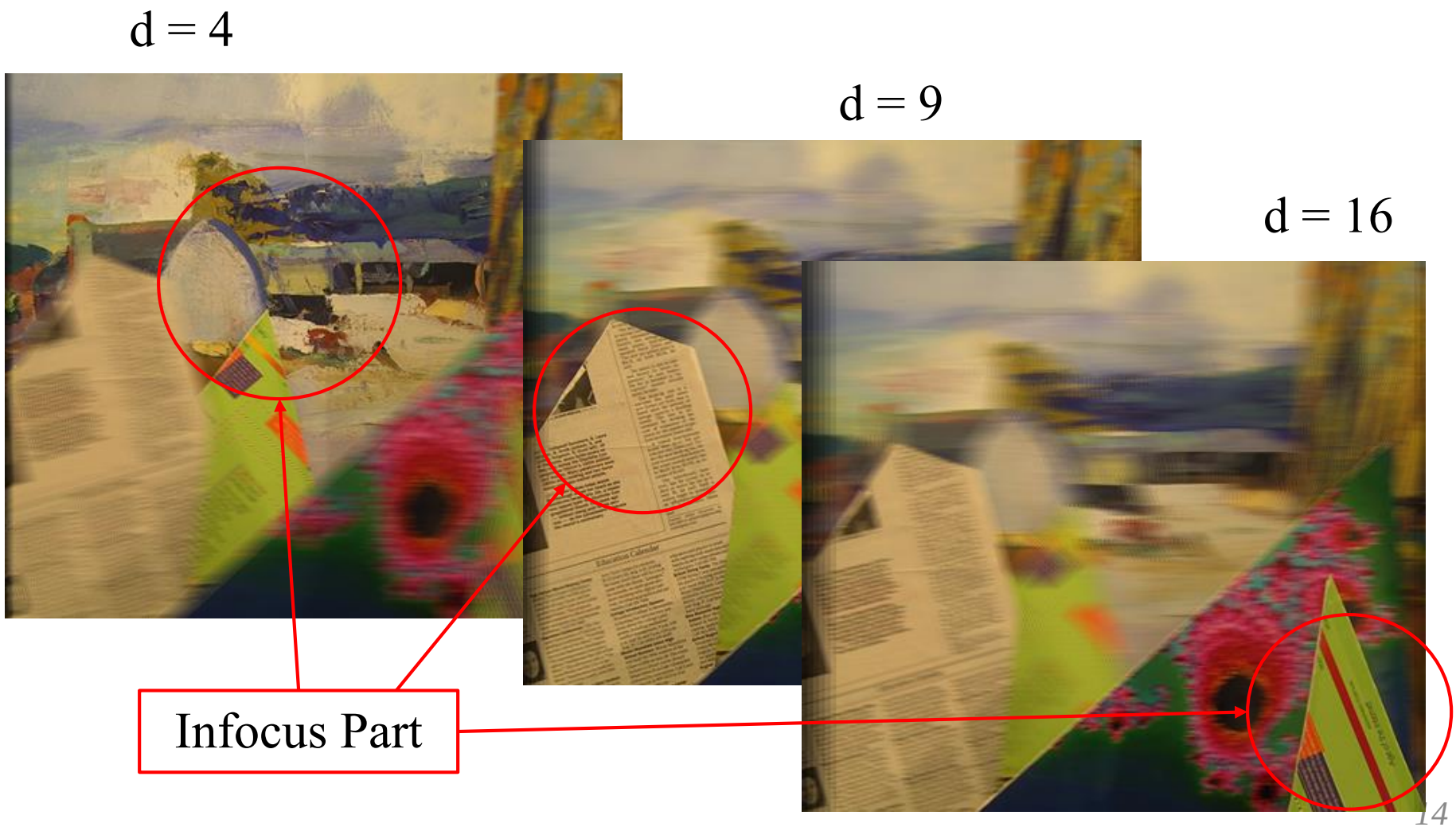
Implementation:
Given m multi-view images, use one as the reference image and given d_max possible disparity values.

for d = 1 to d_max:
    for i = 1 to m:
        Move the other images towards a certain direction
            for a certain distance $L_i \propto$ (d, relative distance);
    end
end
Then we obtain 3d focus image stack as:

$$F^d(x, y, k) = I_{s,t}(x - sd, y - td)$$

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

3D focus image stack



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Proposed method 1

To compute matching cost, we use windows (regarded as ***v***) on each view to find correspondence.

$$C^*(x,y,d) = \frac{1}{n(h-1)} \sum_{k'=2}^{h} \left\| \tilde{\mathbf{v}}_{k'}^{d} \right\|_1$$

Where $\tilde{\mathbf{v}}_{k'}^{d}$ denotes the vector difference between the reference image and the ***k*** th image. And ***h*** is the number of images that will make best performance.

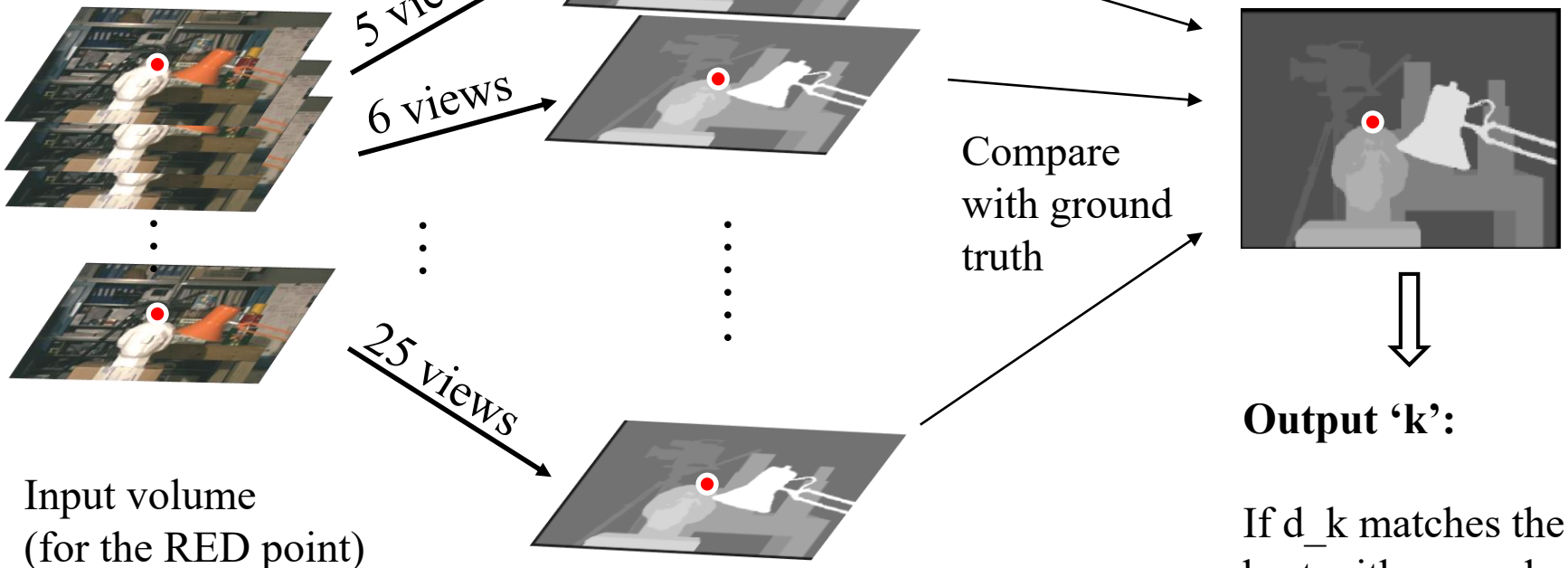
Finally, the disparity value can be computed as:

$$\hat{d}^*(x,y) = \arg\min_{d} C^*(x,y,d)$$

Proposed method 1

Intuition: (view selection system)

- We designed and trained a neural network to adaptively learn how to select the propriate views for disparity estimation.
- The input of the neural network is the reference image and multi-view images; the output of the neural network is an image where the value of each pixel is the number of views that can achieve the best disparity estimation.
- Then we use the number of views to continue doing multi-view disparity estimation.



Input volume
(for the RED point)
: 15 pix *15 pix
(patch size)
*25 views

21 disparity maps
from d_5 to d_25.
(use 5 to 25 views)

Output 'k':

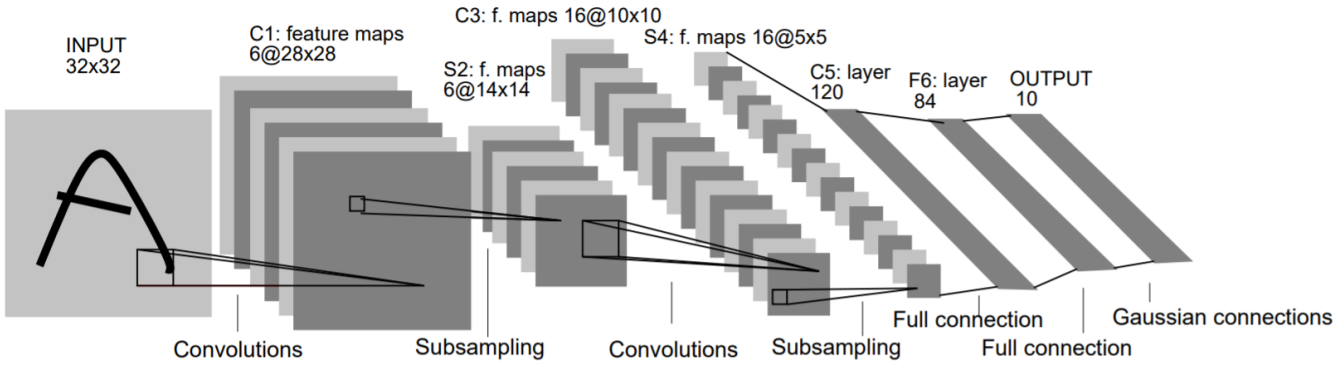
If d_k matches the best with ground truth, then it means using 'k' views give best performance

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Proposed method 1

Intuition: (view selection system)

- The basic structure is LeNet with Cross-Entropy as the loss function。
- Basic structure is as:

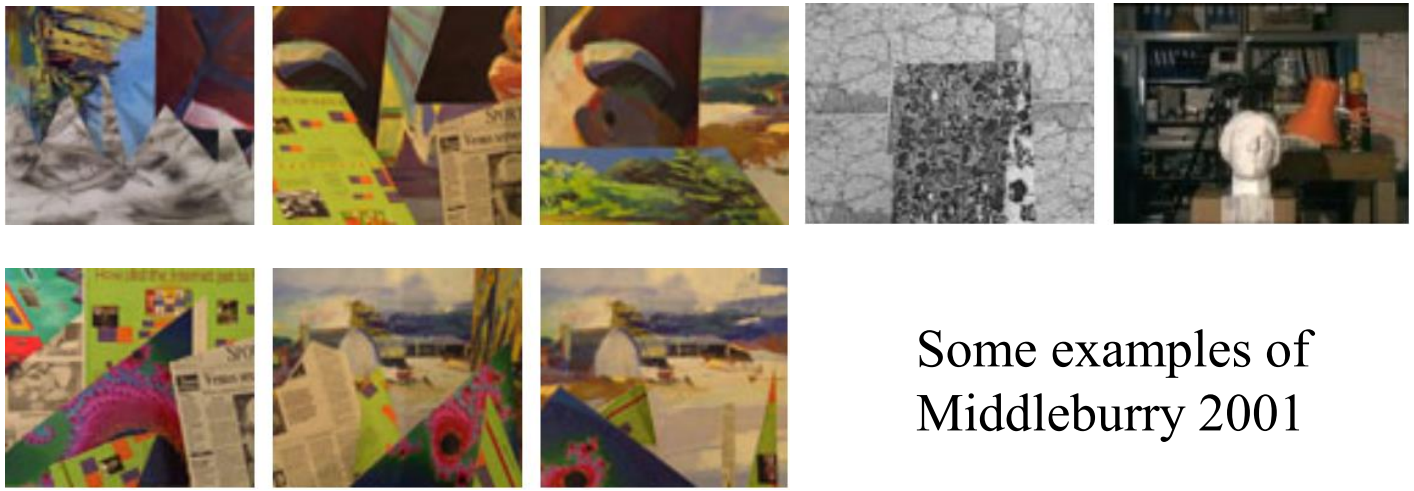


- Then we use the number of views to continue doing multi-view disparity estimation.

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Dataset – Middleburry

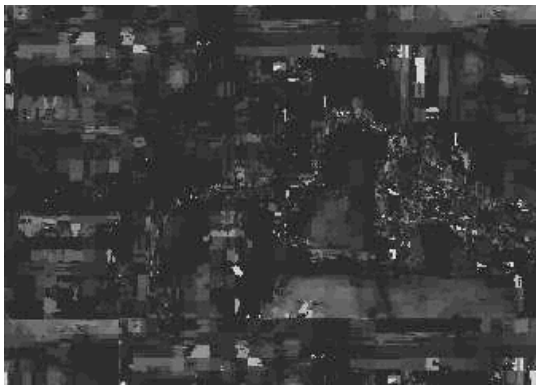
"**2001 Stereo datasets with ground truth**"
These datasets of piecewise planar scenes were created by Daniel Scharstein, Padma Ugbabe, and Rick Szeliski. Each set contains 9 images (im0.ppm - im8.ppm) and ground-truth disparity maps for images 2 and 6.



Some examples of Middleburry 2001

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

View size maps



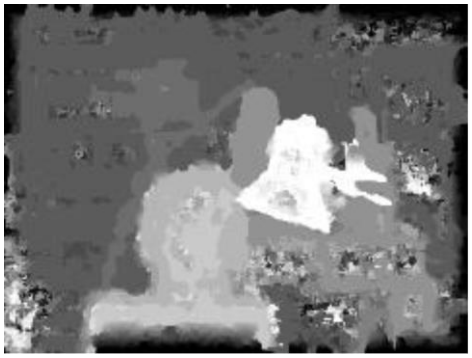
predicted

True

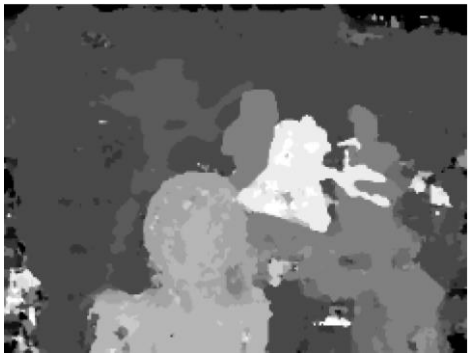
WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Proposed method 1

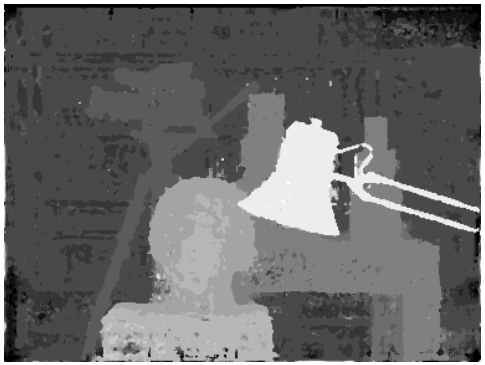
Results:



Miyata et al.[5]



Zhou et al.[4]



Ours (clean)

Ground truth

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Results and comparison

	Miyata et al.[5]	Zhou et al.[4]	Ours
Tsukula	53.02	37.26	29.07

Table 1: Error percentage(%) comparison

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Future work
a.k.a: proposed methond 2

Future work:

- Instead of using current method, we will use state-of-the-art method to compute the matching cost.
- Our method was currently limited to use discrete integer disparity, so we will try to conduct the method with continuous disparities.

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Reference

[1]. D. Scharstein R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms" International Joural of Computer Vision 2002

[2]. Luo,W., Schwing, A., and Urtasun, R. (2016). Efficient deep learning for stereo matching. *In International Conference on Computer Vision and Pattern Recognition*.

[3]. Zheng, E., et al.(2014). PatchMatch Based Joint View Selection and Depth map Estimation. *In international conference of CVPR.*

[4]. Shiwei Zhou, et, al.(2018) Improving disparity map estimation for multi-view noisy images. *ICASSP 2018 conference.*

[5]. M. Miyata, K. Kodama, and T. Hamamoto, “Fast multiple-view denoising based on image reconstruction by plane sweeping,” *in IEEE Conf. Visual Commun. Image Process.*, pp. 462-465, Dec. 2014.

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Thank you !

Xucheng Wan, Zhengyang Lou, Xiao Wang

5/2/2018