1A-44

Robust Stereo with Flash and No-flash Image Pairs

Changyin Zhou
Columbia University

Alejandro Troccoli
Nvidia Research

Kari Pulli
Nvidia Research

Abstract:

We propose a new stereo technique using a pair of flash and no-flash stereo images that is both efficient and robust in handling occlusion boundaries. Our work is motivated by the observation that the brightness variations introduced by the flash can provide a robust cue for establishing stereo matches at occlusion boundaries. This photometric cue is computed per pixel, and though on its own is not robust to reliably resolve depth, it can provide a new discriminant to support patch-based stereo matching algorithms.

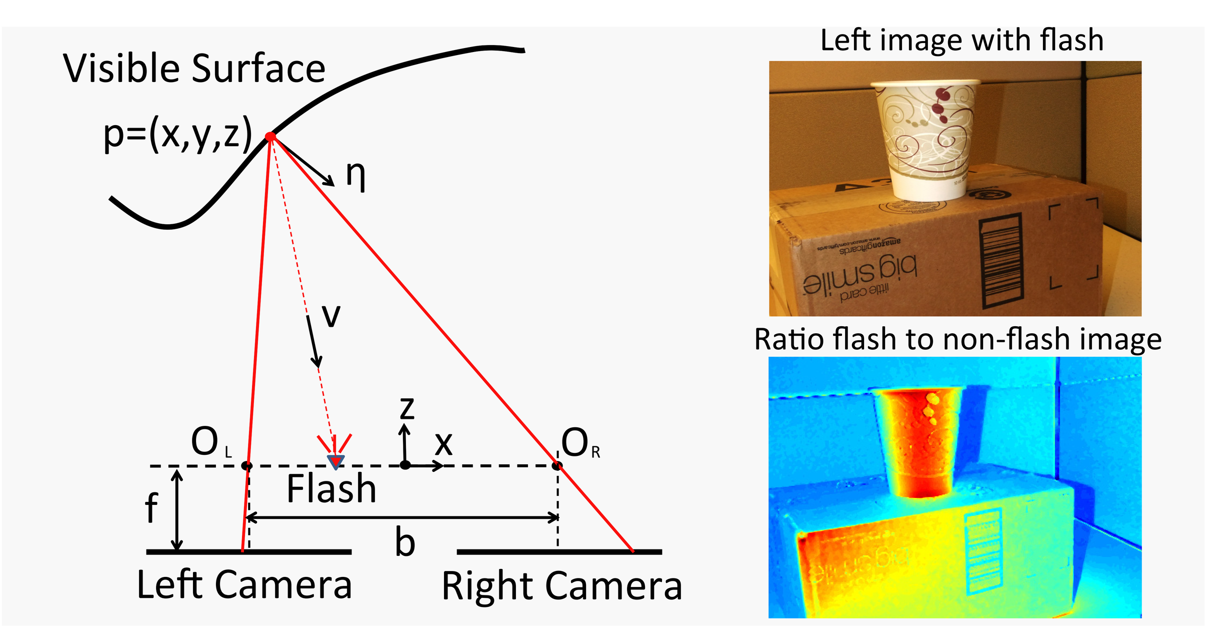
Consumer stereo cameras with flash:

Stereo cameras have found its ways to consumer products (e.g., digital cameras, mobile phones, and tablets.)

Fujifilm FinePix 3D Camera　Android Tablet with Stereo Camera

Motivation:

Stereo camera with a flash captures various depth cues. We use them for ***robust and efficient*** depth recovery on ***mobile devices***.



Raito map: $R(p) = \log\frac{F(p)}{G(p)} = \log\left(1+\frac{I_f}{I_a}\cdot\frac{\langle \hat{n},\hat{v}\rangle}{r^2}\right).$

where I_f and I_a are the intensity of flash and ambient light, respectively.

Stereo [1]:	texture, patch-based	=>	depth
Shading [2]:	near-field, no-texture, pixel-based	=>	normal
Shadow [3,4,5]:	near-field, depth boundary	=>	normal & layer
Light fall-off [6,7]:	*really near* field	=>	relative depth

Flash/no-flash for stereo matching:

Among many possible solutions, we choose to stay simple and local, which often means efficiency and robustness in practice.

Stereo camera captures a pair of stereo images with flash, F_l and F_r, and without flash, G_l and G_r. Denote the ratio between flash and no-flash images as: $R_l = \log(F_l/G_l)$ and $R_r = \log(F_r/G_r)$

*** R map is determined by distance and surface normal ***

We use this particular property of R to solve the well-known difficulty of depth discontinuity in stereo. For Pixel x in F_l, the matching cost of disparity D is computed as:

$$C(x,D) = \sum_{|\Delta|<r} \underbrace{N_{\sigma_\Delta}(\Delta)\cdot N_{\sigma_R}(d_R)}_{\text{pixel weight due to spatial distance and R difference}} \cdot \underbrace{|d_F|^2}_{\text{pixel matching cost}}$$

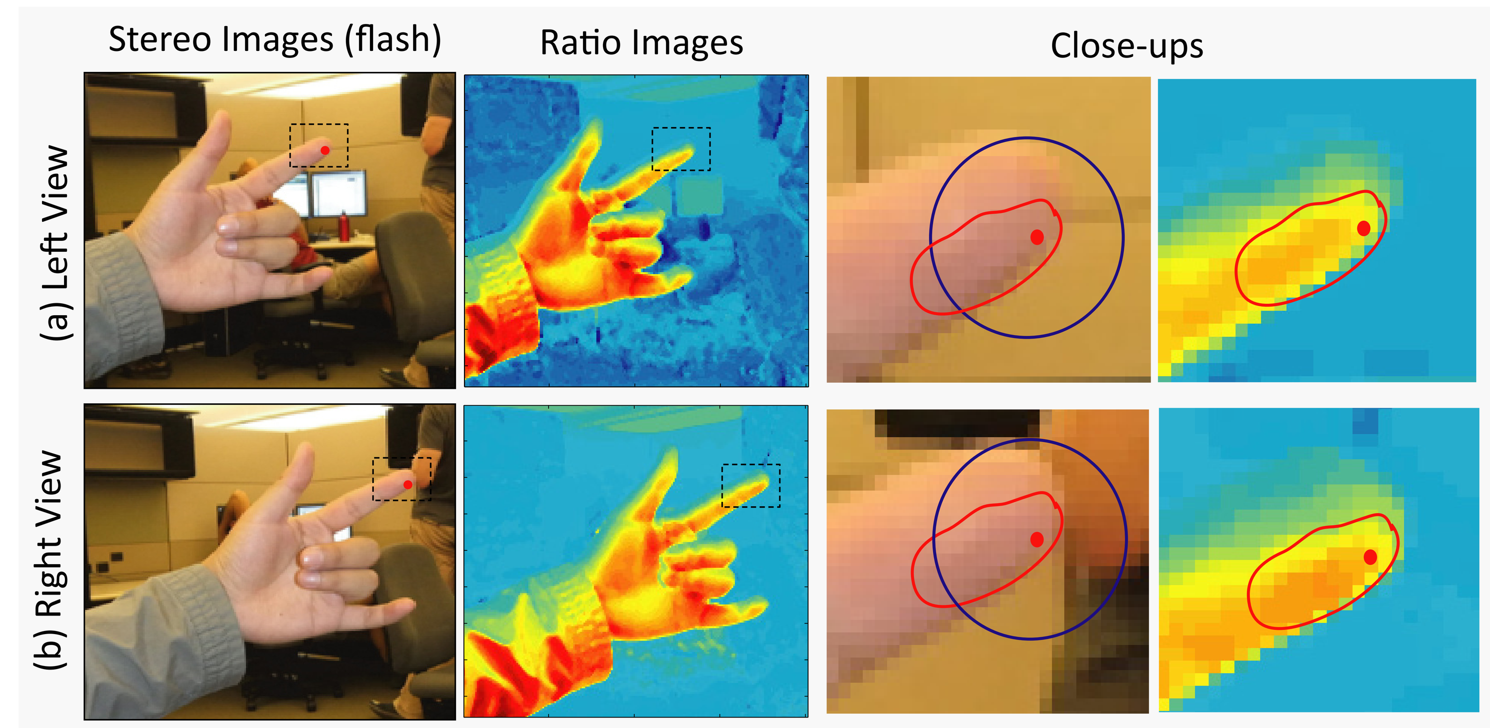
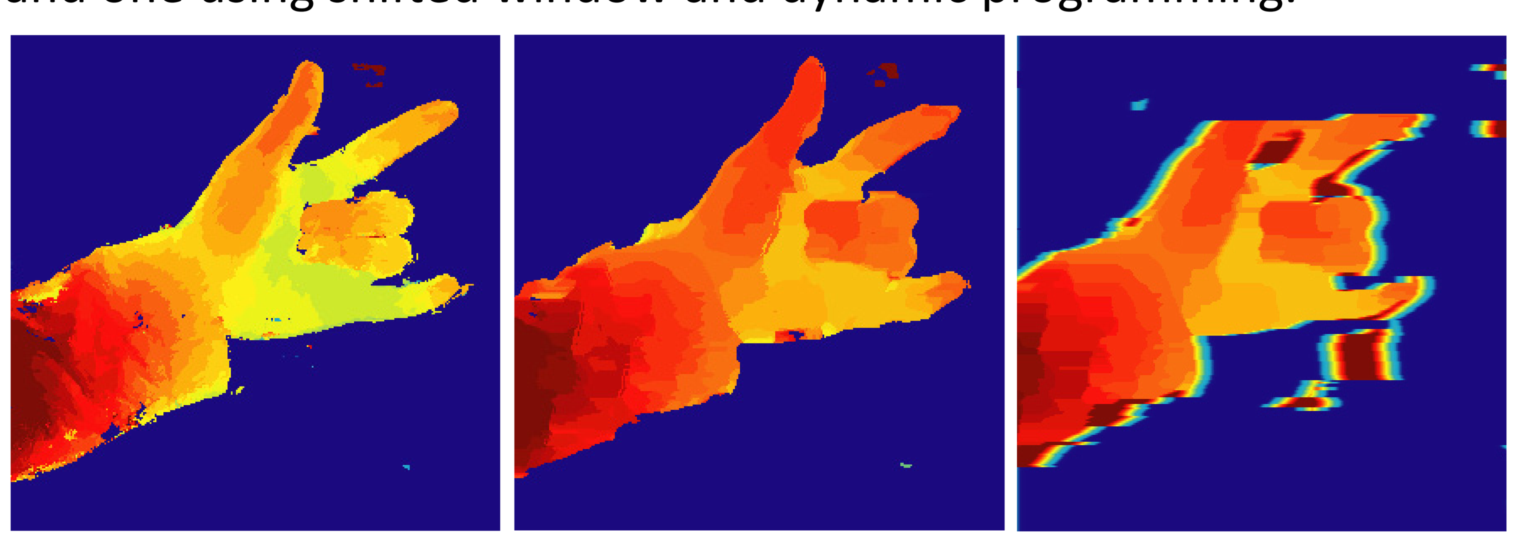


Figure: At occlusion boundary, traditional fixed neighborhoods have different backgrounds in the left (a) and right (b) images. Our cross-bilateral weight that multiplies distance term (blue) and ratio term (red) compares mostly only foreground pixels.

The proposed matching cost can be used with various stereo algorithms. We choose the simplest winner-takes-all (WTA) strategy for efficiency and then refine depth map using Left-Right-Consistency (LRC) [8].

Below is an estimated depth map in comparison to two traditional stereo techniques, one using shifted window and WTA, and one using shifted window and dynamic programming.

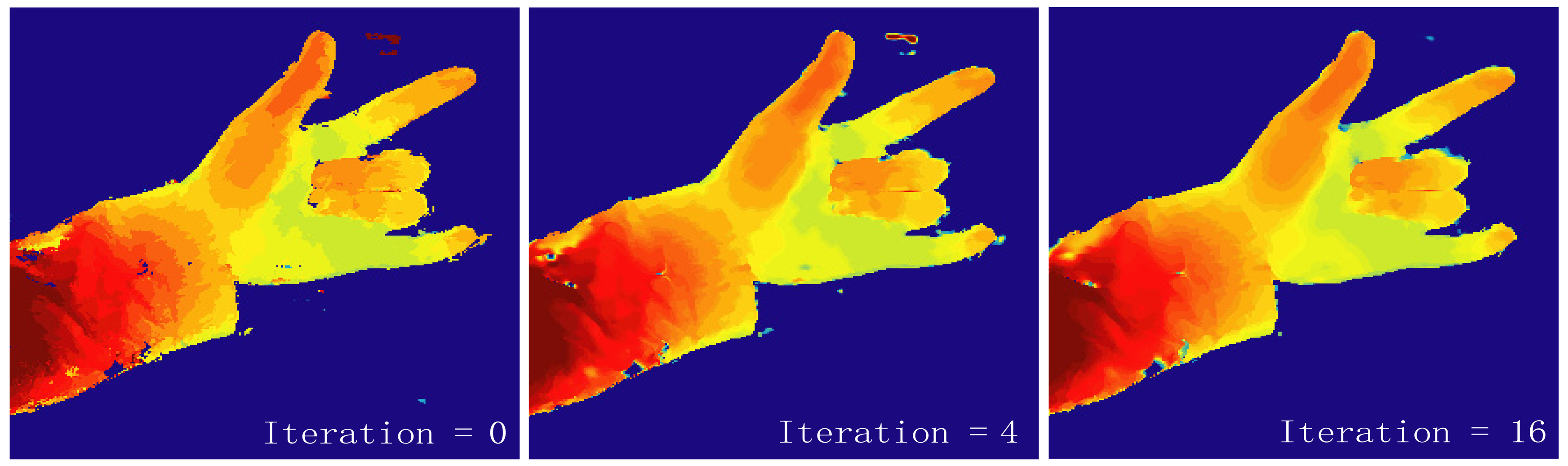


(a) Flash Stereo (WTA)　(b) Traditional Stereo (WTA)　(c) Traditional Stereo (DP)

Filtering to reduce depth quantization error

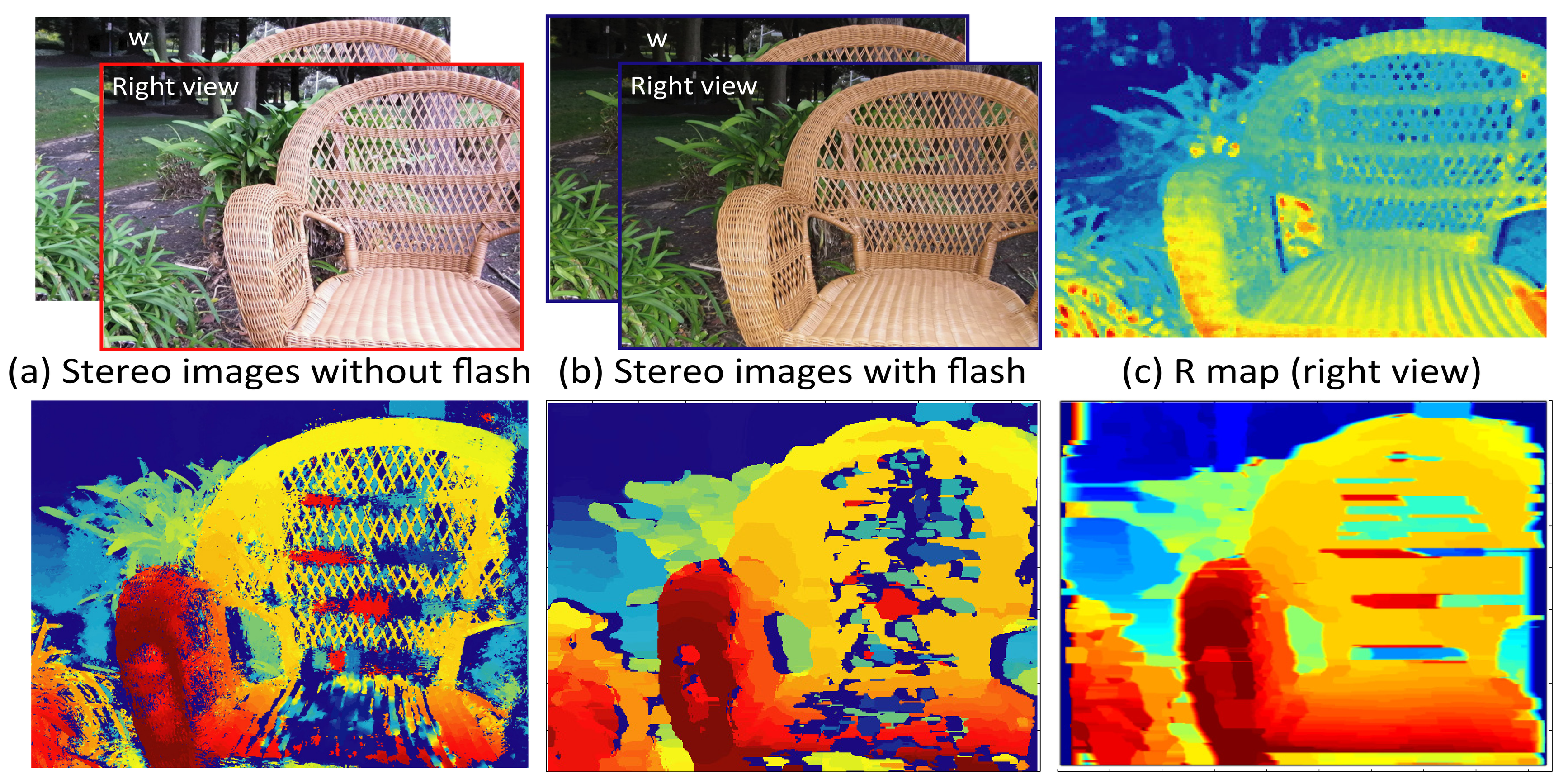
Disparities are often computed in a unit of pixel because sub-pixel computation is expensive and less reliable, which leads to low precision depth map. We propose a simple yet efficient algorithm to increase the disparity precision.

Observations: both the disparity and R are locally linear for any planar surface.
Basic idea: Smooth depth along the dimension when R varies little.



Experiments

Experiments were done using a Fujifilm FinePix Real 3D camera. Two stereo pairs were captured in succession with a shutter lag of 0.x sec with and without flash. We left most settings (e.g., focus, exposure, f#, ISO, and WB) in automatic mode.



(a) Stereo images without flash　(b) Stereo images with flash　(c) R map (right view)

(d) Flash stereo (WTA, LRC)　(e) Traditional stereo (shift window, WTA, LRC)　(f) Traditional sterero (shift window, DP)

References (incomplete):

[1] Scharstein&Szeliski, 2002, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.
[2] Horn, 1989, Shape from shading.
[3] Shafer&Kanade, 1983, Using shadows in finding surface orientations.
[4] Kriegman&Belhumeur, 2001, what shadows reveals about object structure.
[5] Raskar et al., 2004, Nonphotorealistic camera: depth edge detection and stylized rendering using multi-flash imaging.
[6] Martin, 2000, Depth from Flash.
[7] Mulligan&Brolly, 2004, Surface determination by photometric ranging.
[8] Weng et al., 1988, Two-view matching.