# Improving the Kinect by Cross-Modal Stereo

Wei-Chen Chiu, Ulf Blanke, Mario Fritz

## Motivation

### Kinect: from gaming interface to robotic perception

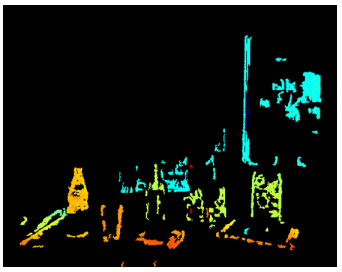

IR projector + IR camera: 3D depth sensor



## Problem

### No sensor is perfect: fuse Kinect depth sensing with cross-modal stereo

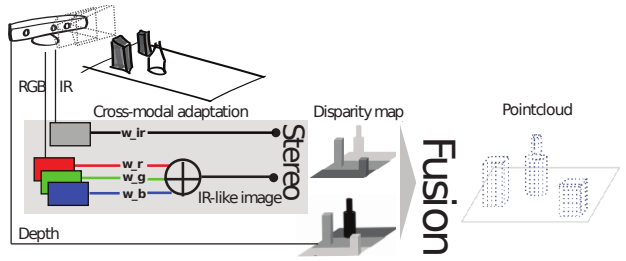
Kinect active sensing:
- good for homogenous region
- failed on some surfaces
  - specular
  - transparent
  - reflective



Passive stereo vision:
- hard for homogenous region
- enable to detect disparities at edges of transparent or reflective objects

## Method

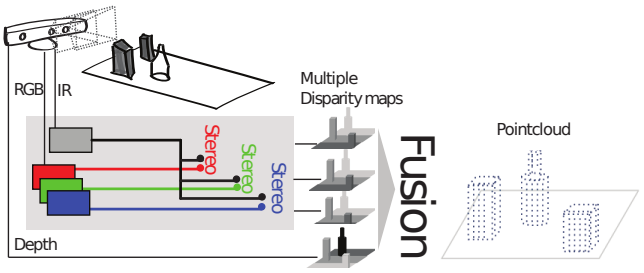
### Early fusion




Estimating optimal weighted combination of RGB channels to be IR-like for improved stereo matching.

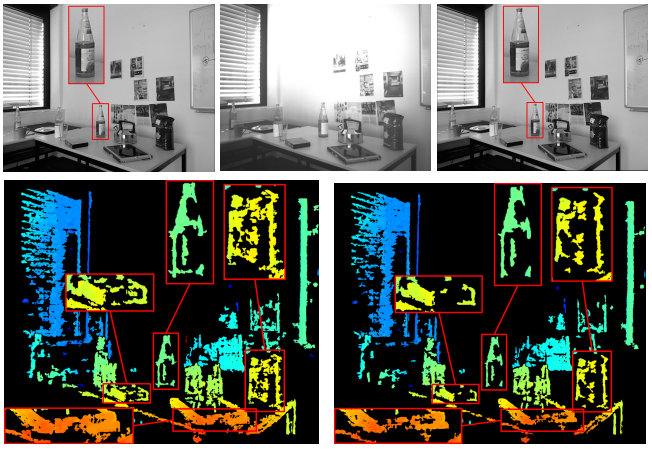
$$\max_{w_r,w_g,w_b} \quad num\_of\_stereo\_match(w_r * I_r^{rgb} + w_g * I_g^{rgb} + w_b * I_b^{rgb}, I^{ir})$$

### Late fusion



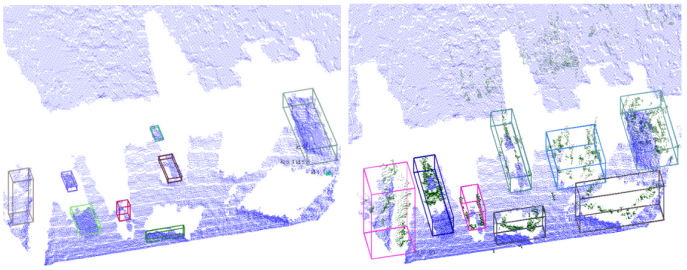

Delay combination of different color channels and compute stereo correspondences w.r.t. the IR image independently. Fuse resulting depth estimate with depth sensed by Kinect by union of point clouds.

## Result



(a)Converted RGB image by optimized weights. (b)IR image (covered projector). (c)RGB image converted to grayscale. (d)Disparity from (a) and (b). (e)Disparity from (c) and (b).

### Evalution on object segmentation task from 3D point cloud



Kinect only | fused Kinect and stereo

- Dataset of table top scenario: 106 objects in 19 images
- Best result of proposed fusion schemes achieves an average precision of 76.6%. Comparing to 48.8% of built-in Kinect depth estimate, we achieve a significant improvment of nearly 30%.

MAX-PLANCK-GESELLSCHAFT

max planck institut informatik