



Local Background Enclosure for RGB-D Salient Object Detection

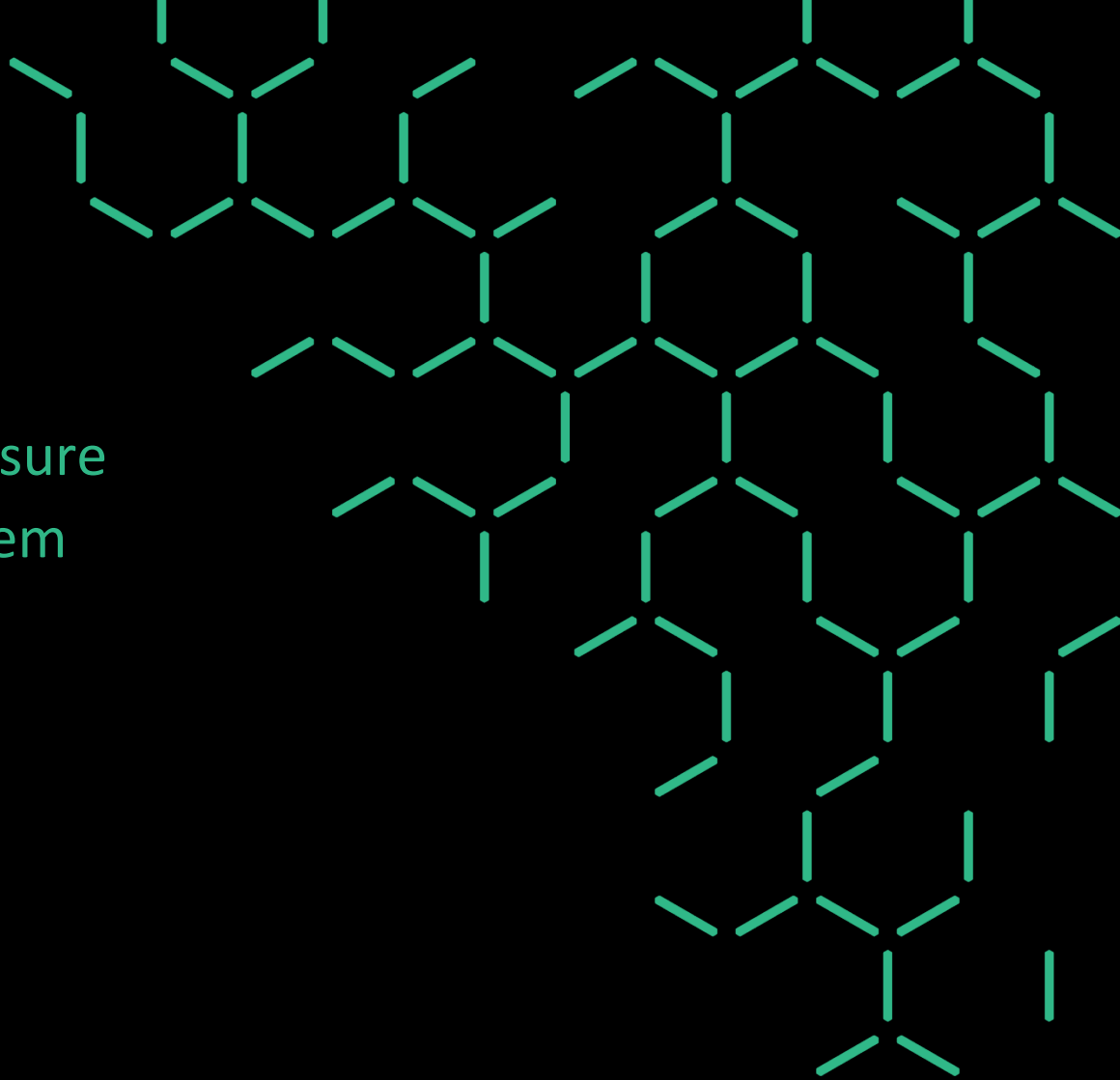
David Feng, Nick Barnes, Shaodi You, Chris McCarthy
CVPR Spotlight Presentation

www.data61.csiro.au



Overview

- **Overview of saliency**
- Previous work
- Local Background Enclosure
- Saliency Detection System
- Results



Saliency



- The **saliency** of an item is how much it stands out from its surroundings.

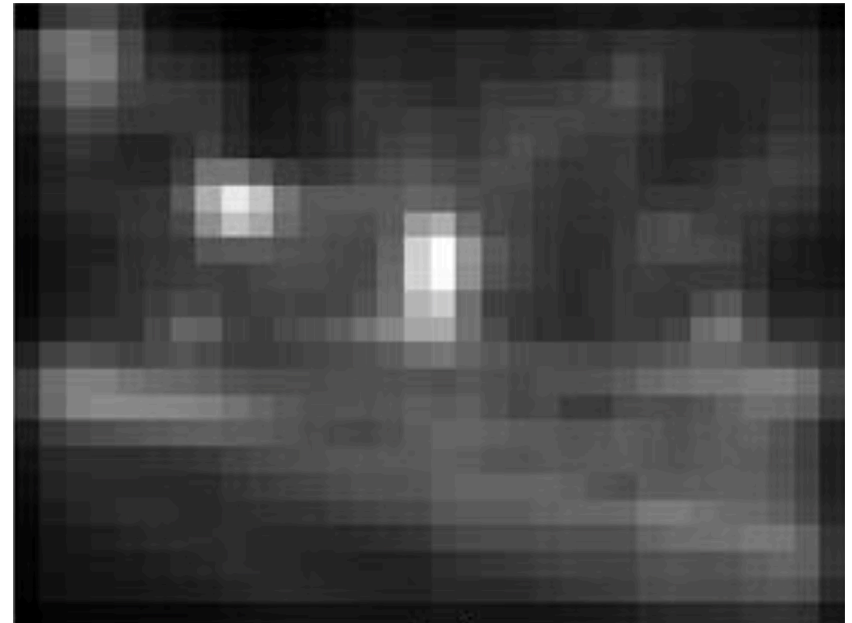


Saliency

- The **saliency** of an item is how much it stands out from its surroundings.



(a) RGB Image



(b) Saliency Map (Itti & Koch, 1998)

Saliency

- Fixation prediction:



(a) RGB Image

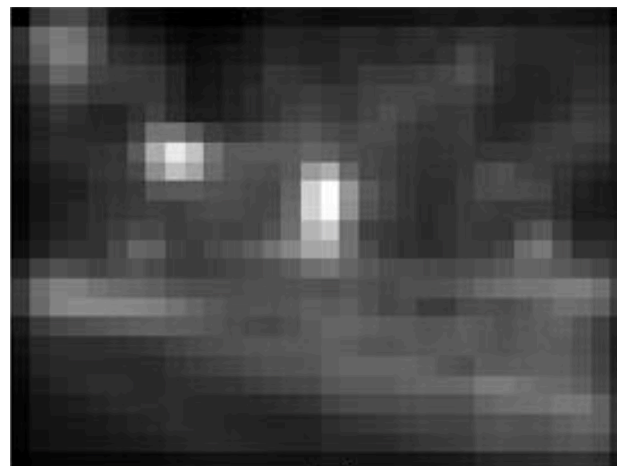


(b) Fixation Point Ground Truth

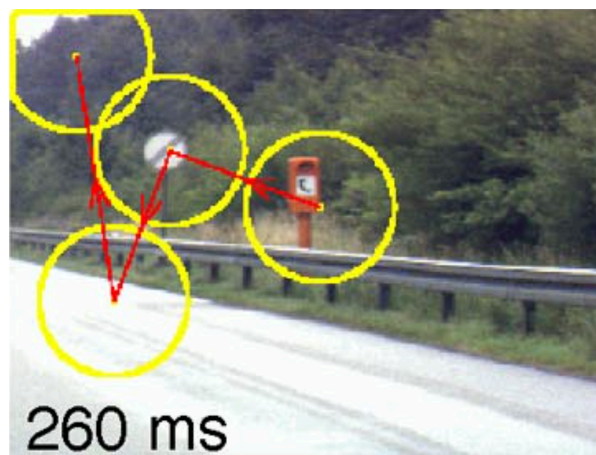
Saliency



(a) RGB Image



(b) Saliency Map (Itti & Koch, 1998)

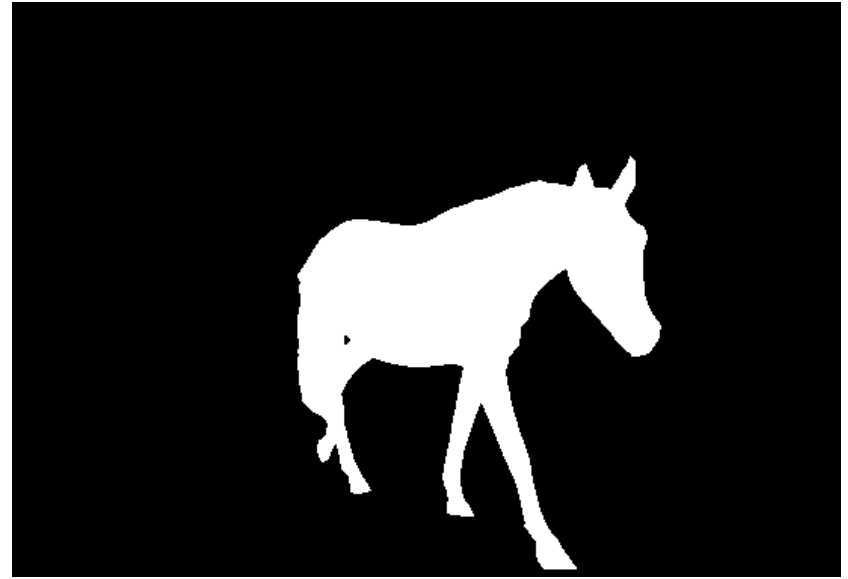


(c) Fixation Points

Salient Object Detection



(a) RGB Image

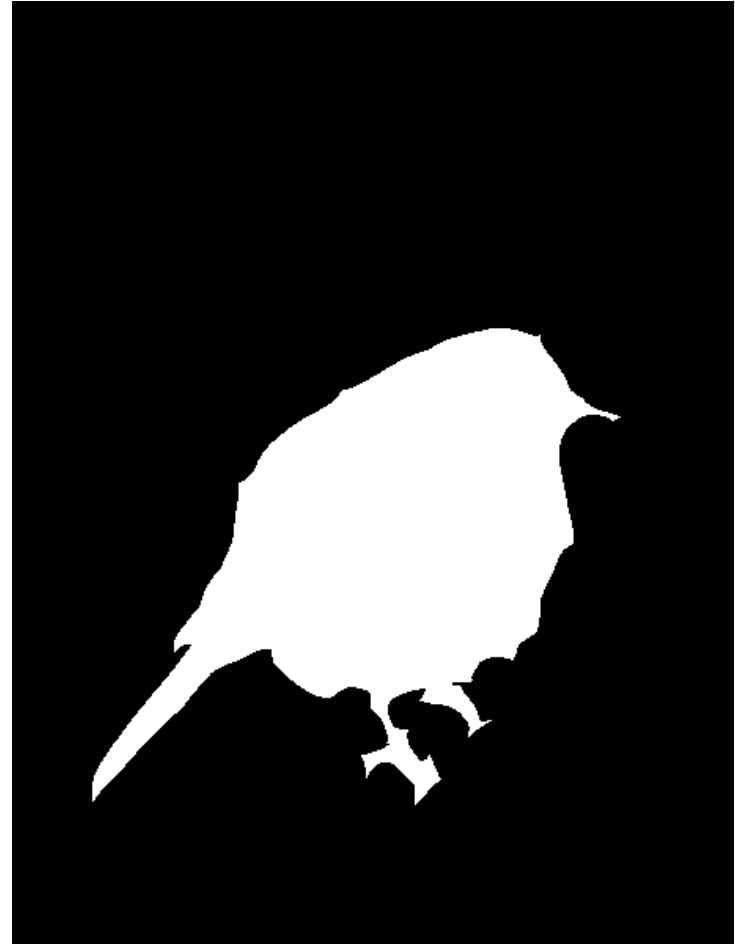


(b) Ground Truth

Salient Object Detection



(a) RGB Image



(b) Ground Truth

RGB-D Salient Object Detection

- Incorporating depth can improve performance



(a) RGB Image



(b) Depth Image



(c) Ground Truth

Application: Image Thumbnailing



(a) RGB



(b) Depth

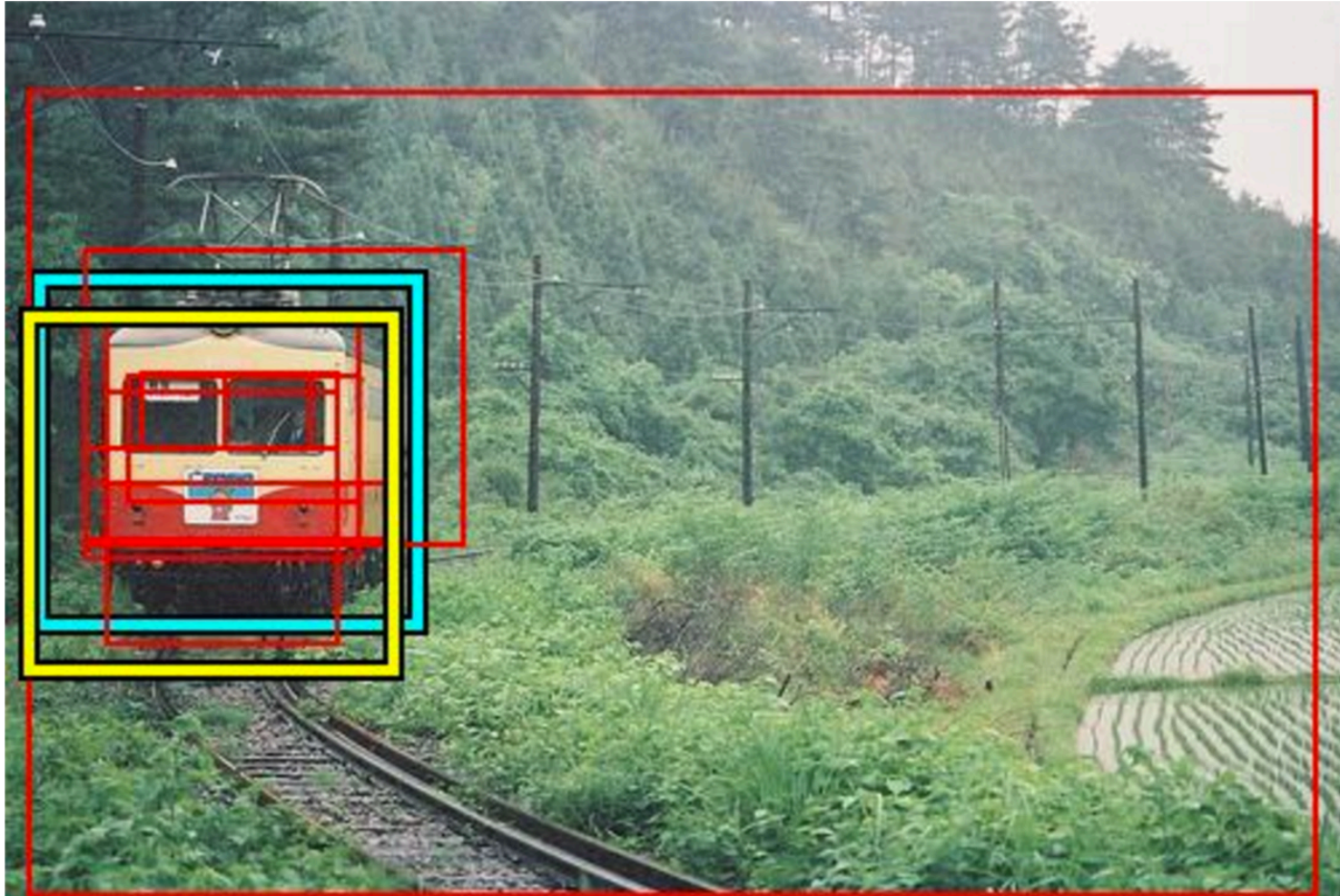


(c) Resized Thumbnail



(d) Saliency Thumbnail

Application: Object Detection



<http://groups.inf.ed.ac.uk/calvin/objectness/>

Application: Search Space Reduction



(a) Robot



(b) Focus Point in RGB



(c) Focus Point in Depth

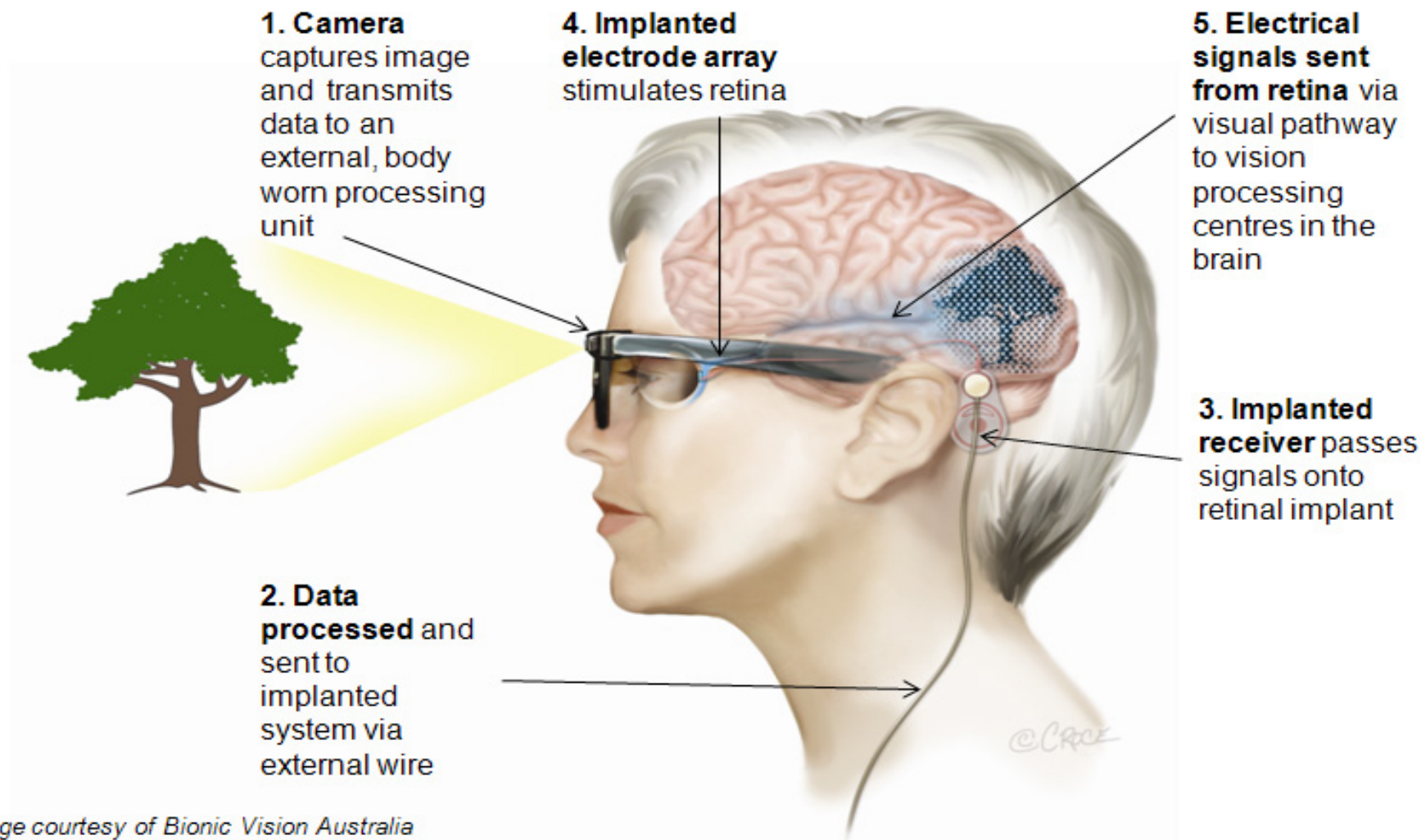
Frintrop *et al.* (2005)

The bionic eye - how it works

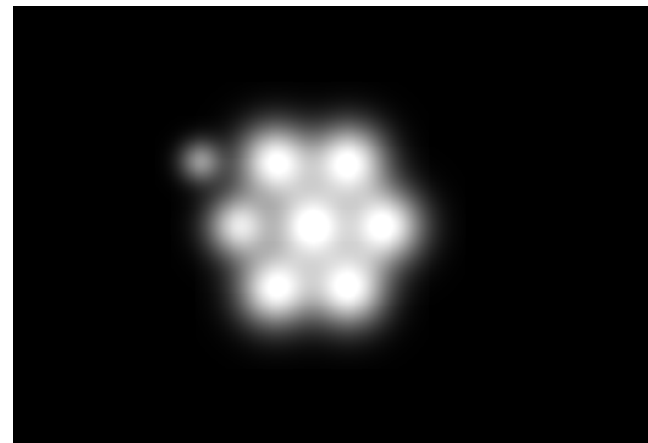
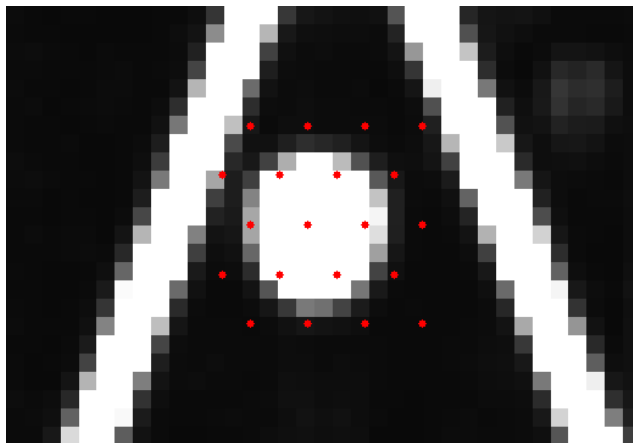
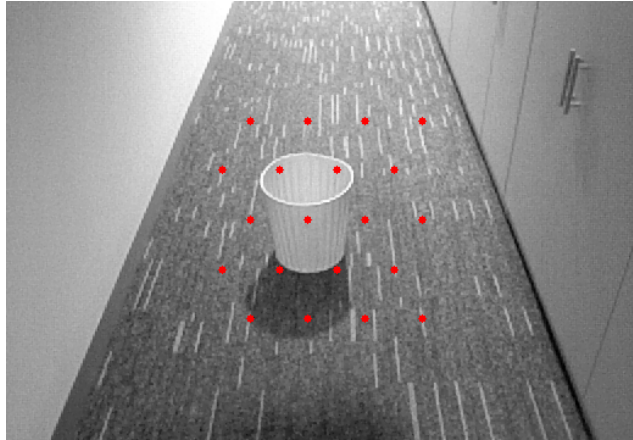
First prototype: Wide-view neurostimulator



BIONICVISION
AUSTRALIA



Application: Prosthetic Vision

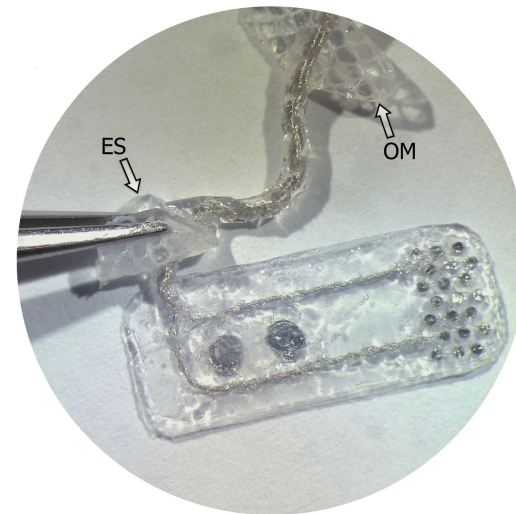
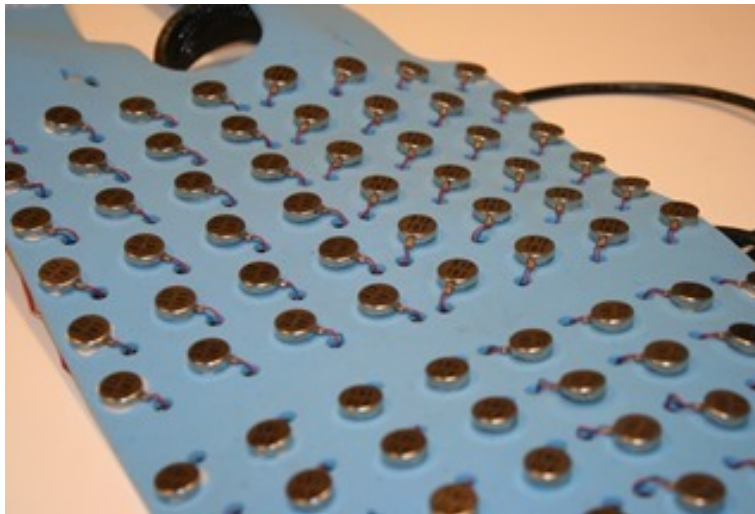


(a) Input Image

(b) Displayed Image

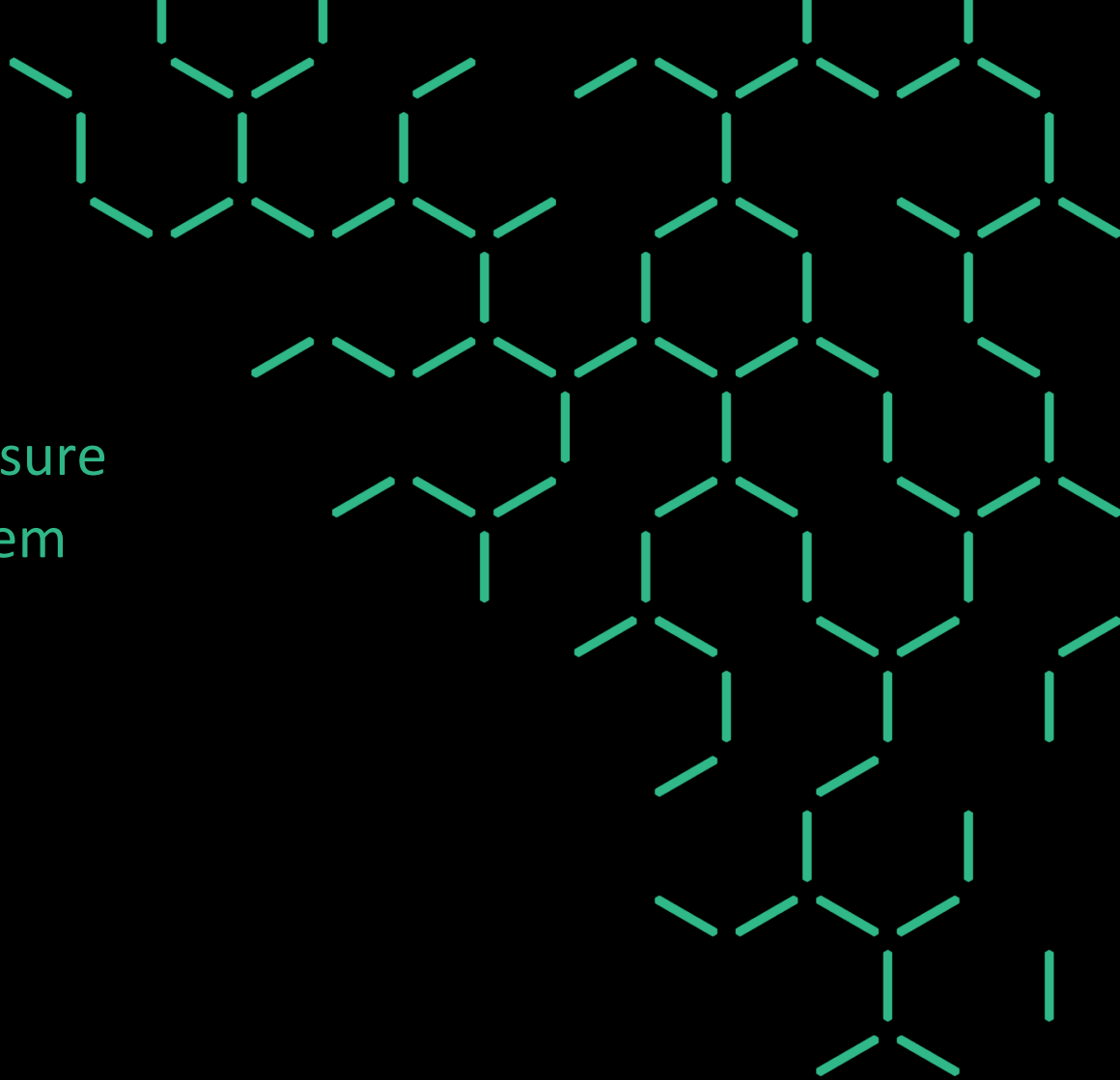
Application: Prosthetic Vision

- Vision Processing for Prosthetic Vision
 - Ensure key information is displayed on low-resolution visual aids.



Overview

- Overview of saliency
- **Previous work**
- Local Background Enclosure
- Saliency Detection System
- Results



RGB-D Salient Object Detection



(a) RGB

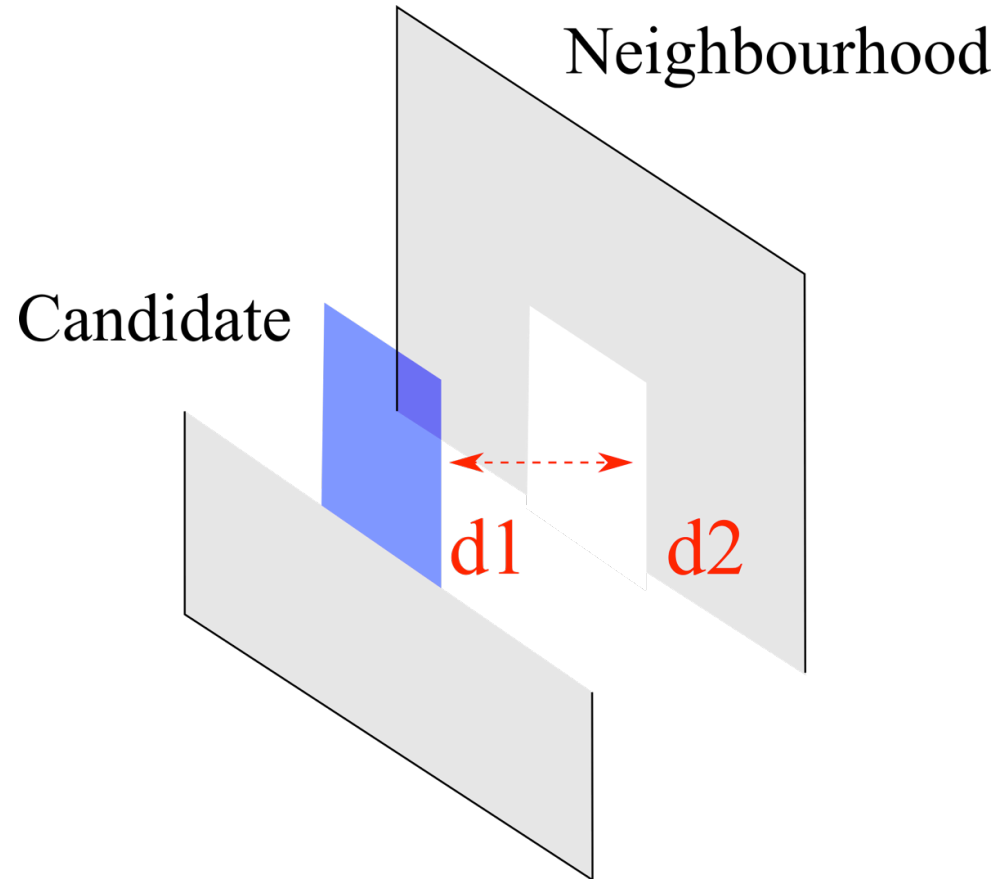


(b) Depth



(c) Ground Truth

Previous Work



Depth Contrast: $d1 - d2$

Previous Work



- **GP:** Global Priors (Ren *et al.*, 2015)
 - Global depth contrast
- **ACSD:** Anisotropic Center Surround Difference (Ju *et al.*, 2014)
 - Depth contrast along scanlines
- **LMH:** Low Medium High (Peng *et al.*, 2014)
 - Multi-context depth contrast

Previous Work



- Problem: background often contains areas of high depth contrast.

Previous Work

- Problem: background often contains areas of high depth contrast.



(a) Depth



(b) Ground Truth

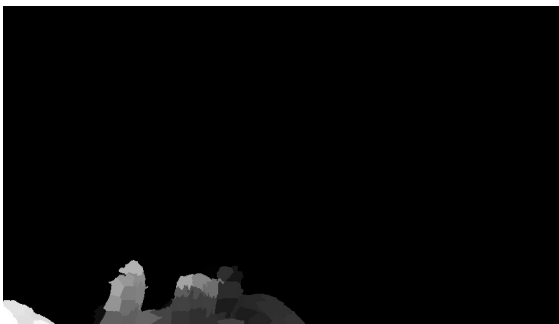
Previous Work



(a) Depth



(b) Ground Truth



(c) LMH



(d) GP



(e) ACSD

Previous Work



(a) Depth



(b) Ground Truth



(c) LMH-DC

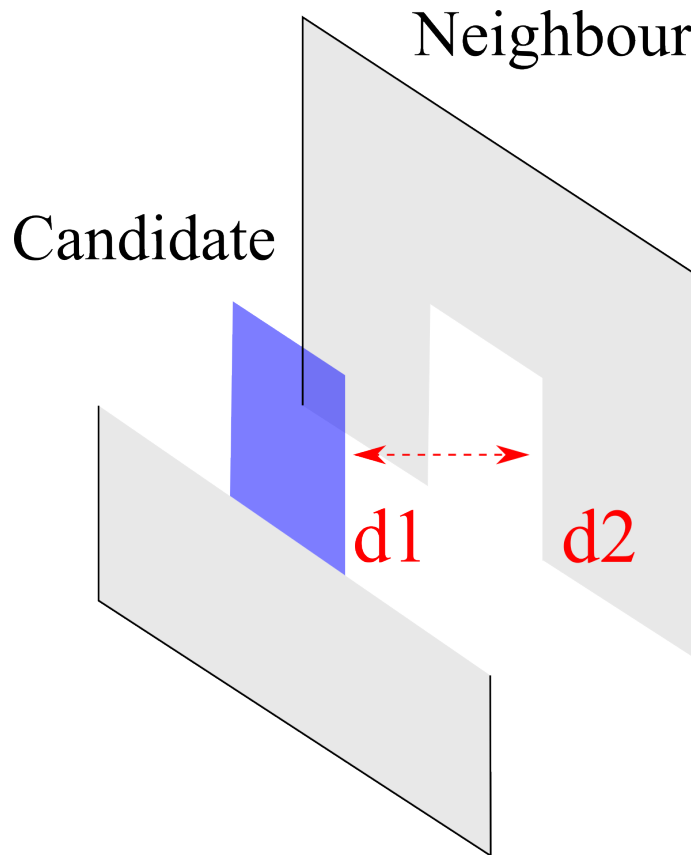


(d) GP-DC

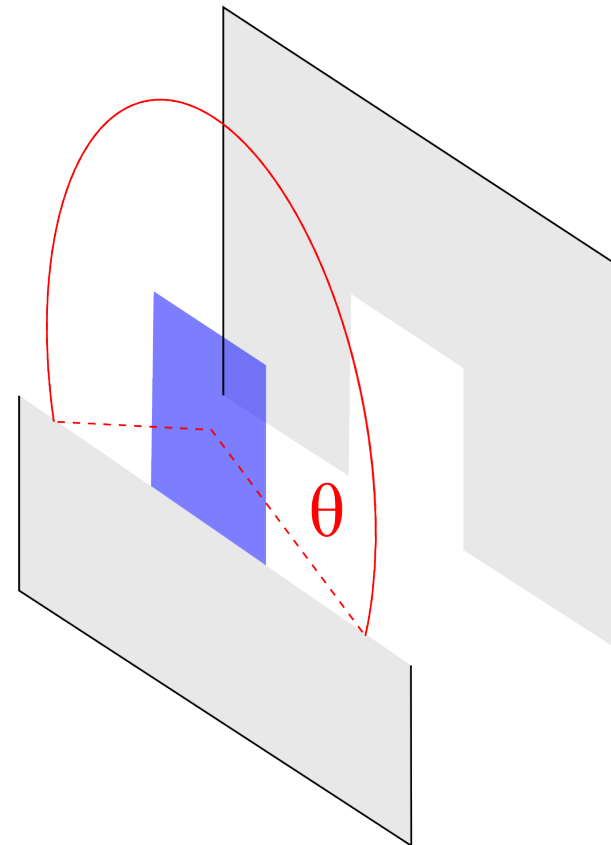


(e) ACSD-DC

Local Background Enclosure



Depth Contrast: $d1 - d2$



Background Enclosure: $\theta / 2\pi$

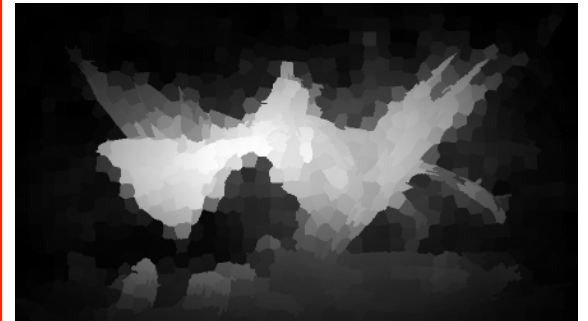
Local Background Enclosure (LBE)



(a) Depth



(b) Ground Truth



(c) LBE



(c) LMH-DC



(d) GP-DC



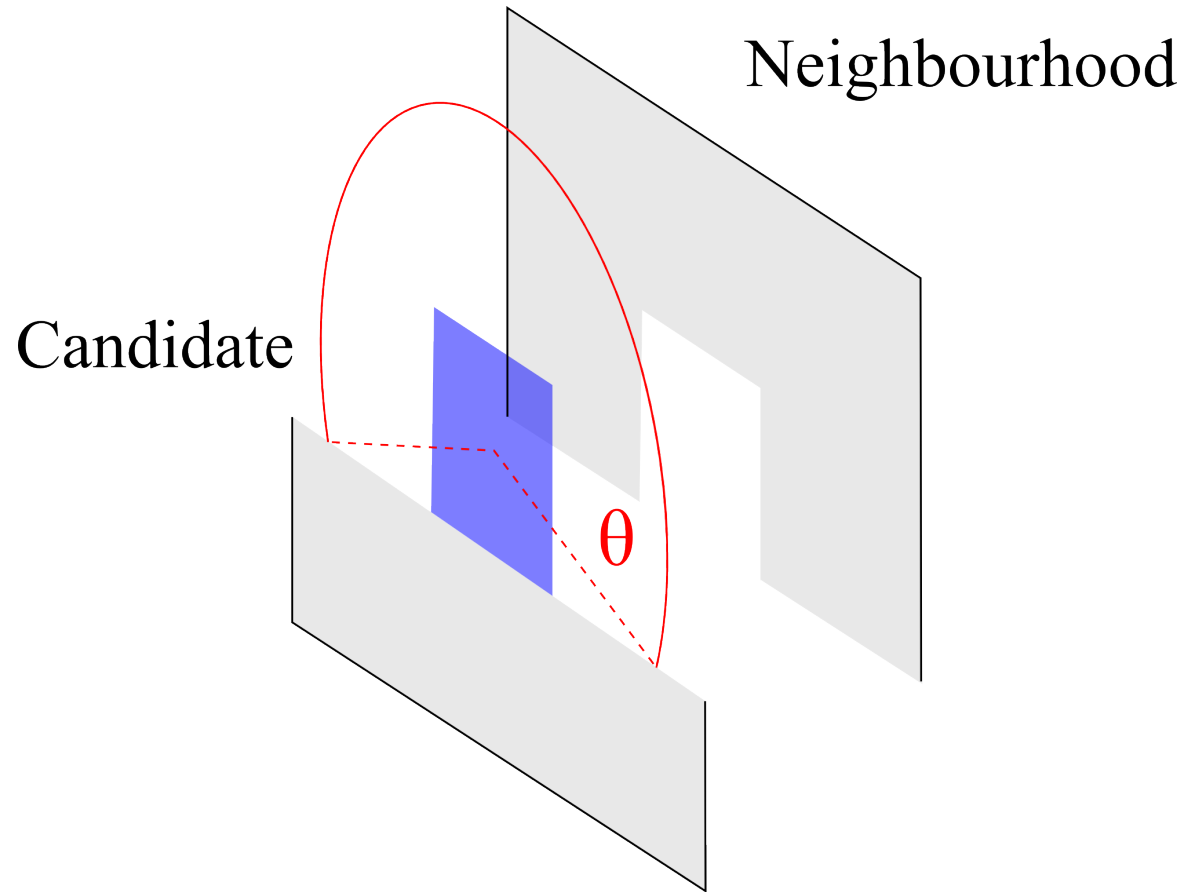
(e) ACSD-DC

Overview

- Overview of saliency
- Previous work
- **Local Background Enclosure**
- Saliency Detection System
- Results



Local Background Enclosure (LBE)



Local Background Set

- Given a patch P and local neighbourhood N_P , background set denoted by $B(P, t)$:

$$B(P, t) = \bigcup \{P' \in N_P | D(P') > D(P) + t\}, \quad (1)$$

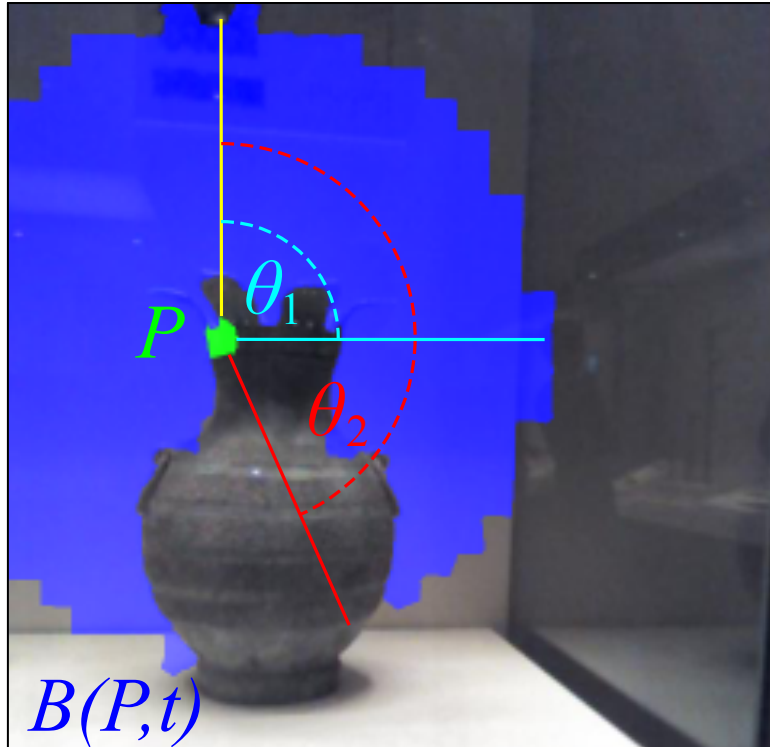


Local Background Enclosure (LBE)



- We employ two complementary measures of background enclosure:
 - Angular Fill
 - Angular Gap

Density Function for Angular Fill

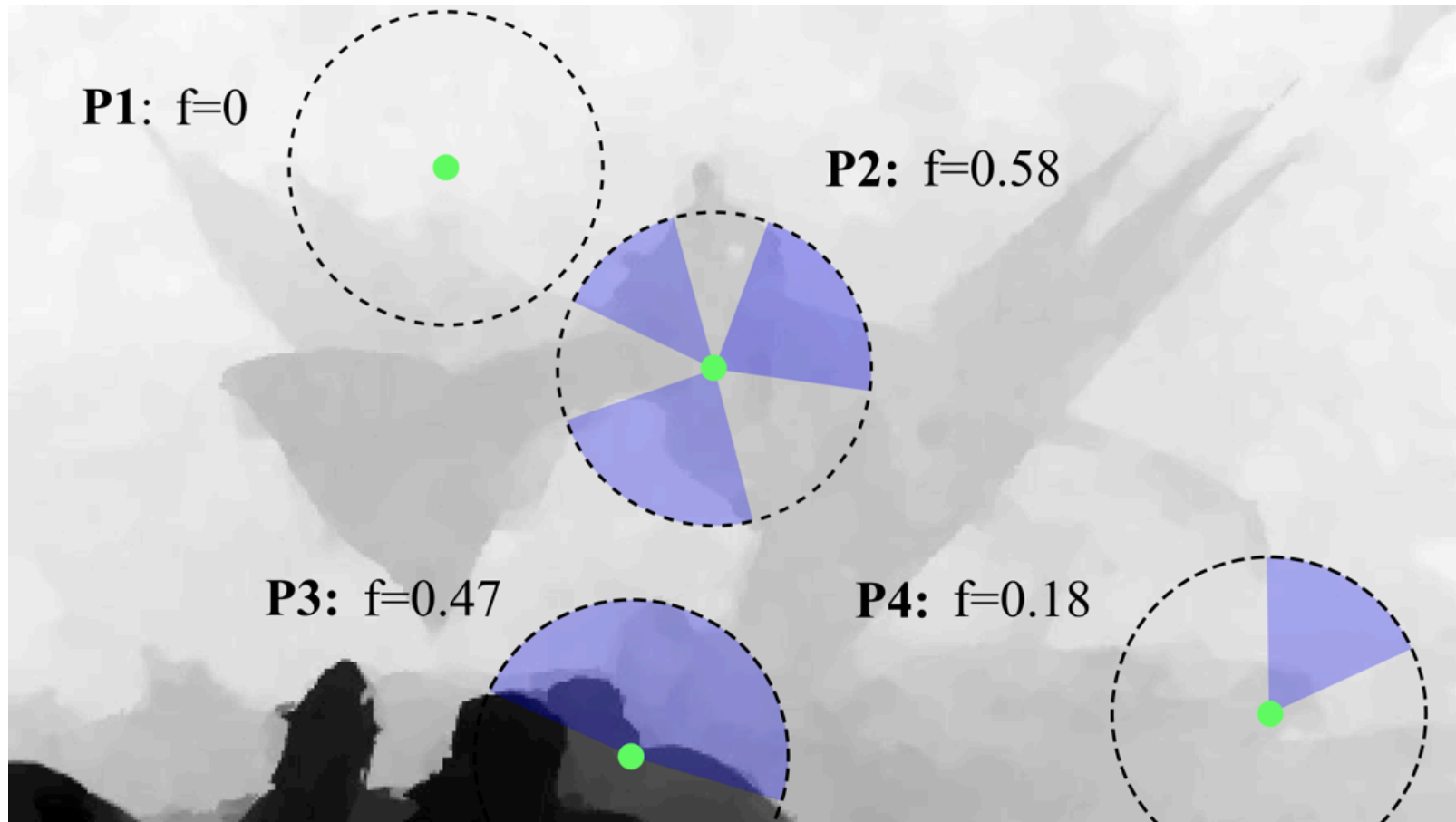


$$I(\theta_1, P, B(P, t)) = 1$$

$$I(\theta_2, P, B(P, t)) = 0$$

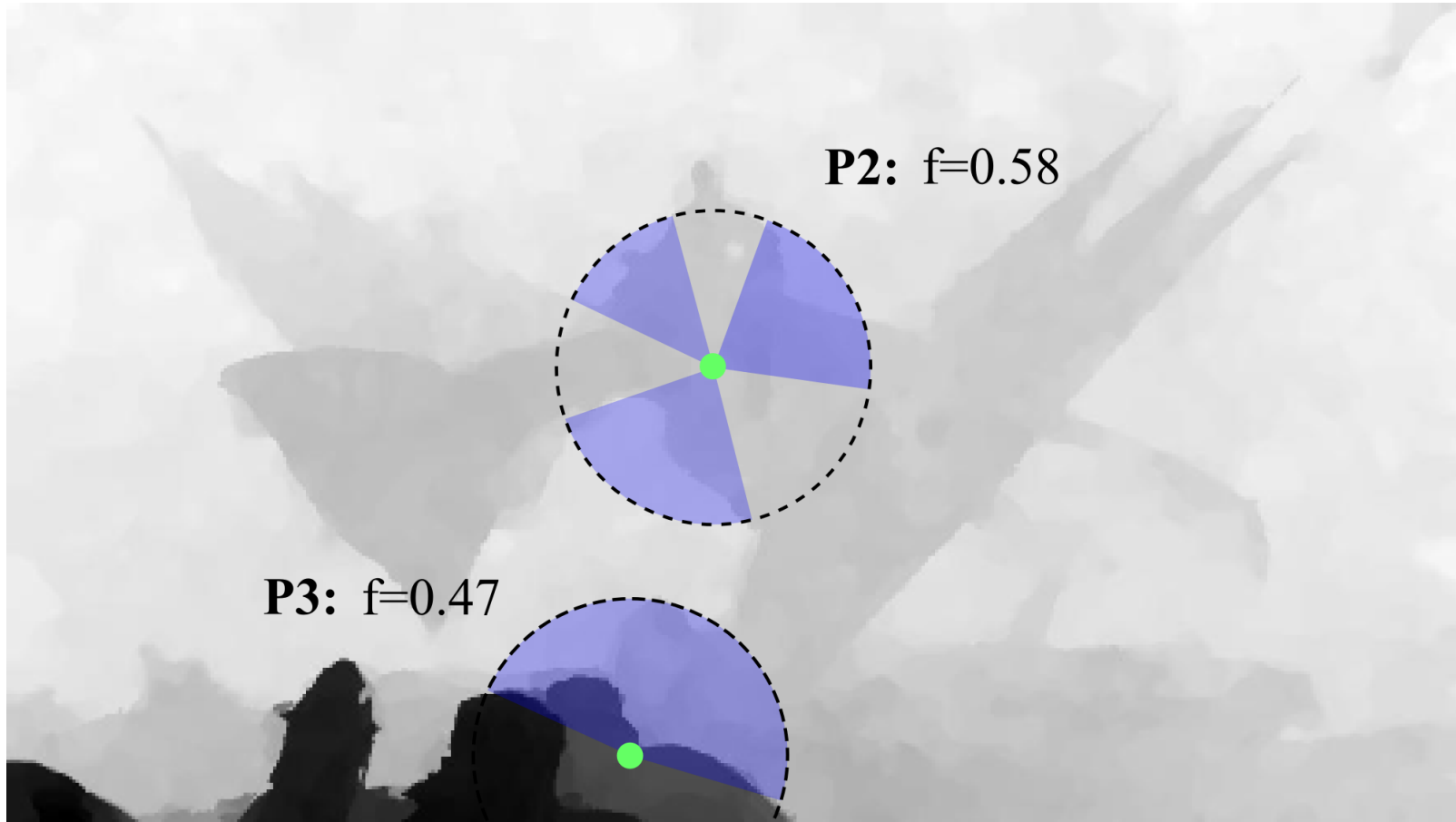
$$f(P, B(P, t)) = \frac{1}{2\pi} \int_0^{2\pi} I(\theta, P, B(P, t)) d\theta, \quad (2)$$

Density Function for Angular Fill



Angular Fill Density Function, $t = \sigma/2$

Density Function for Angular Fill



Density Function for Angular Gap



Density Function for Angular Gap



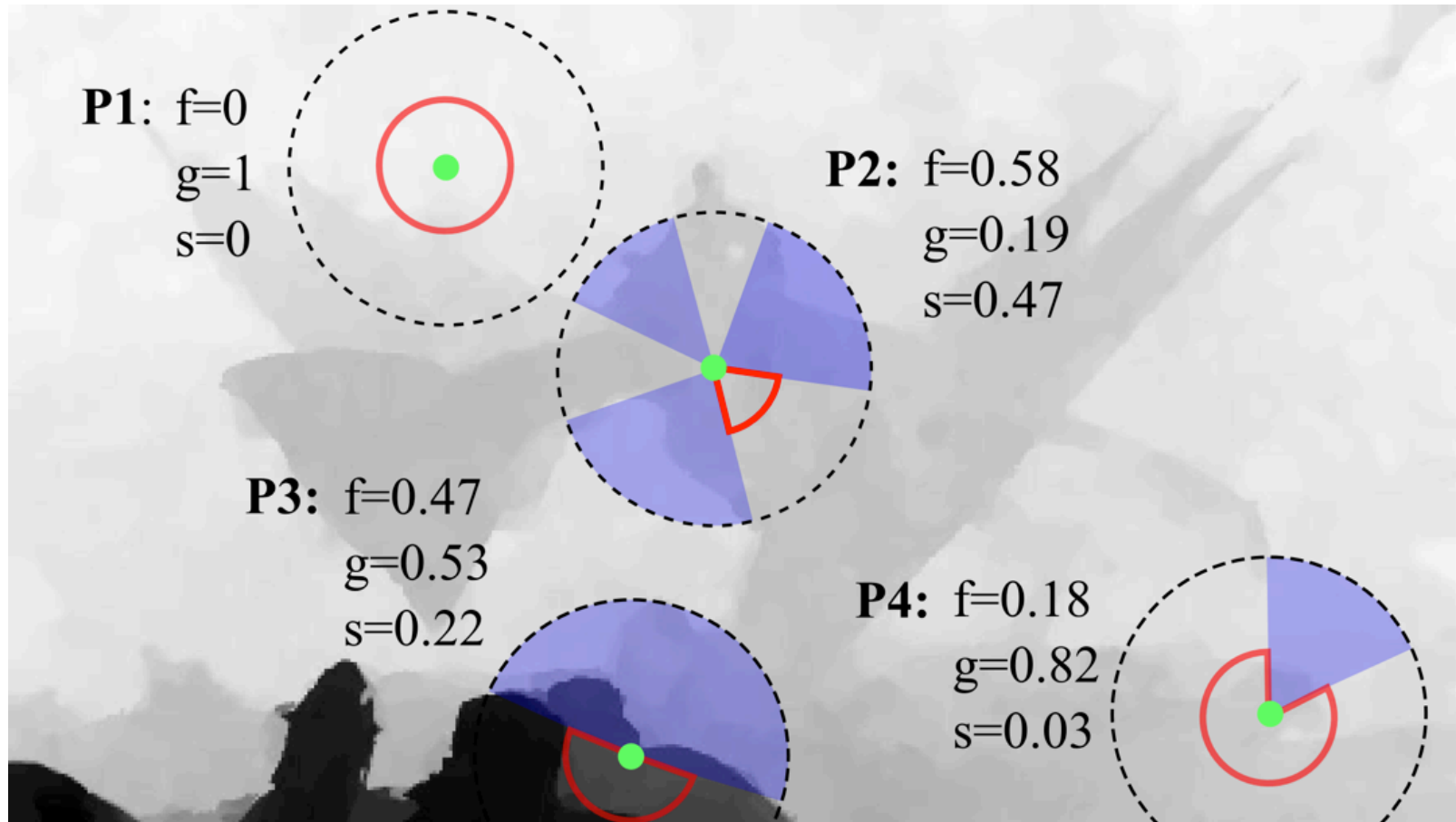
- Angular gap function defined by:

$$g(P, Q) = \frac{1}{2\pi} \cdot \max_{(\theta_1, \theta_2) \in \Theta} \{|\theta_1 - \theta_2|\}, \quad (4)$$

- Θ denotes set of angular boundaries not containing background:

$$\Theta = \{(\theta_1, \theta_2) \mid I(\theta, P, Q) = 0 \quad \forall \theta \in [\theta_1, \theta_2]\}. \quad (5)$$

Density Function for Angular Gap



Distribution Functions



- Distribution function of f :

$$F(P) = \int_0^{\sigma} f(P, B(P, t)) dt, \quad (3)$$

Distribution Functions

- Distribution function of f :

$$F(P) = \int_0^{\sigma} f(P, B(P, t)) dt, \quad (3)$$

- Distribution function of g :

$$G(P) = \int_0^{\sigma} 1 - g(P, B(P, t)) dt. \quad (6)$$

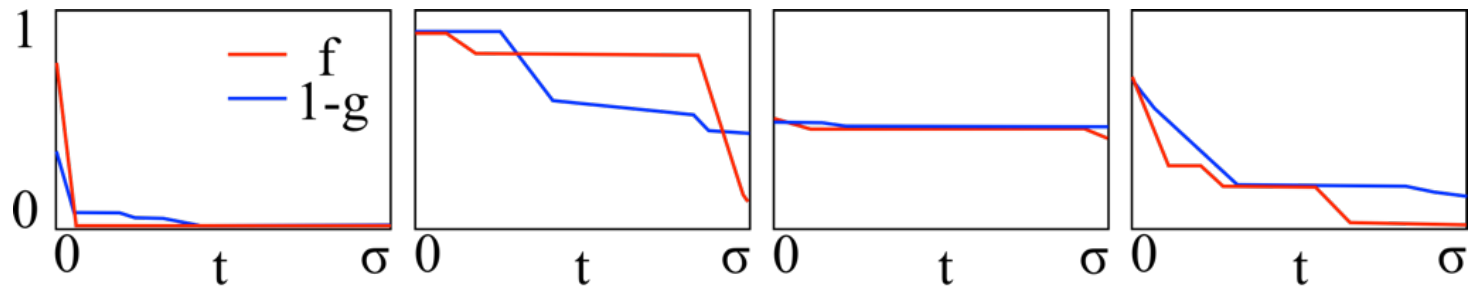
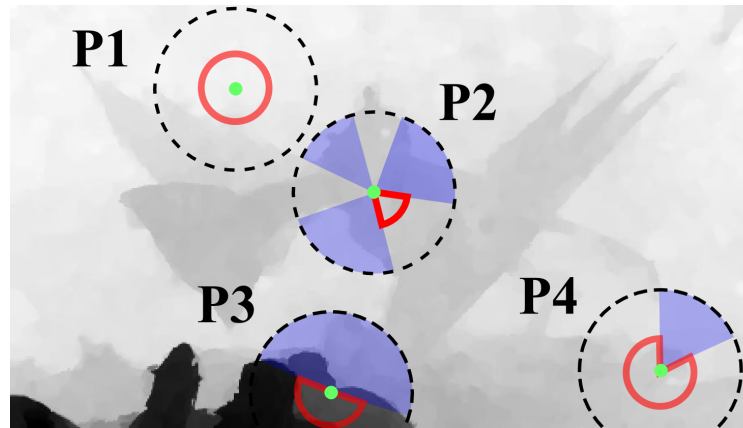
Local Background Enclosure



- LBE map given by

$$S(P) = F(P) \cdot G(P). \quad (7)$$

Local Background Enclosure



P1: $F = 0.05$	P2: $F = 0.66$	P3: $F = 0.48$	P4: $F = 0.22$
$G = 0.06$	$G = 0.78$	$G = 0.47$	$G = 0.16$
$S = 0.03$	$S = 0.51$	$S = 0.23$	$S = 0.04$

Angular Fill and Gap Distribution Functions, $t \in [0, \sigma]$

Datasets



- **RGBD1000**

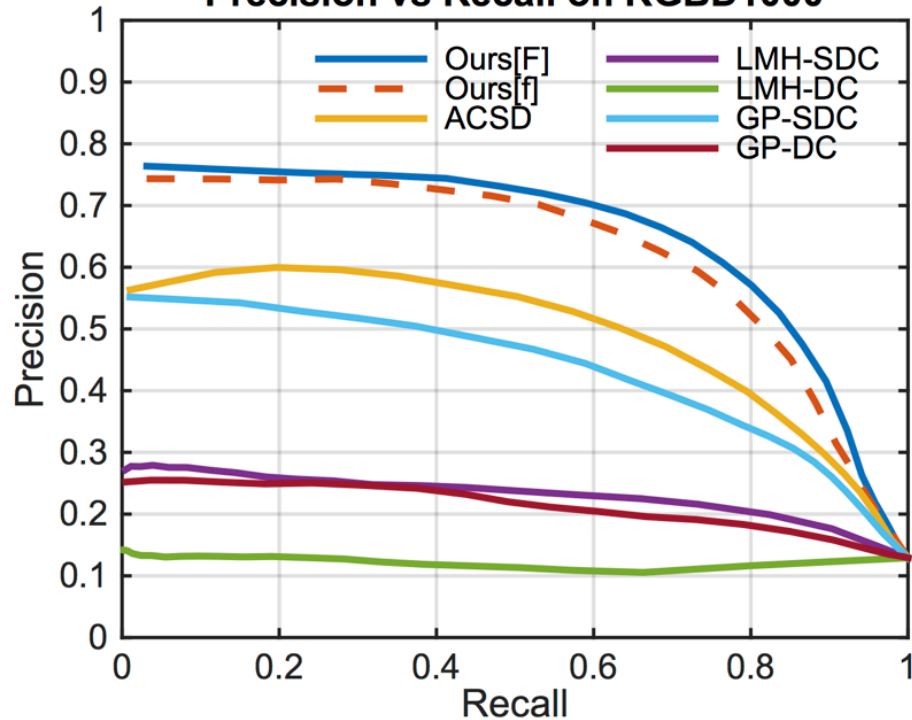
- 1000 RGB-D images
- Depth from structured light sensor
- Proposed by authors of LMH

- **NJUDS2000**

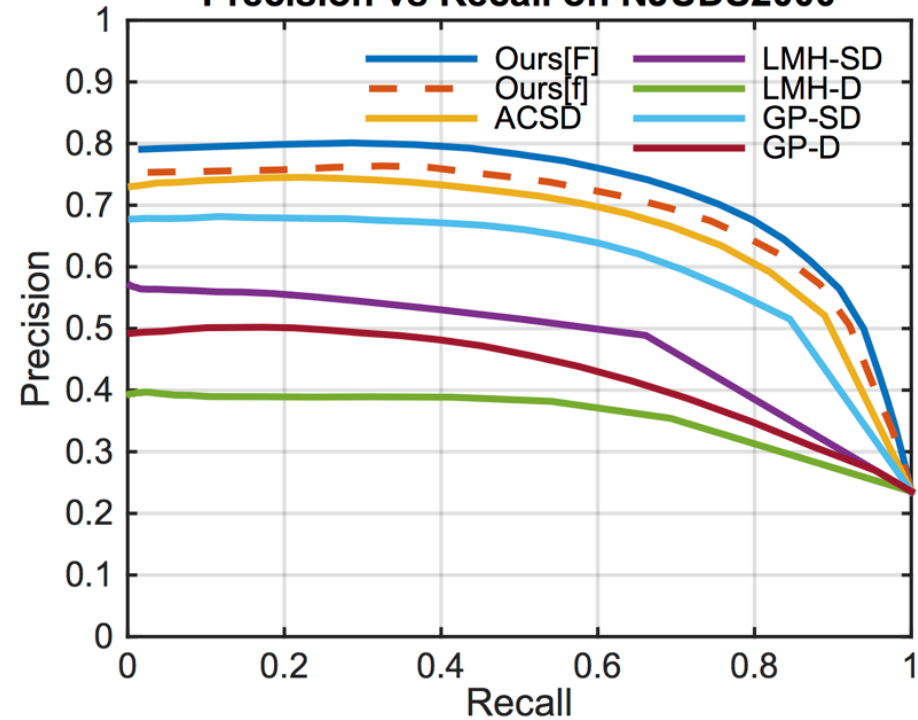
- 2000 RGB-D images
- Depth estimated from stereo
- Proposed by authors of ACSD

Comparison with Contrast Features

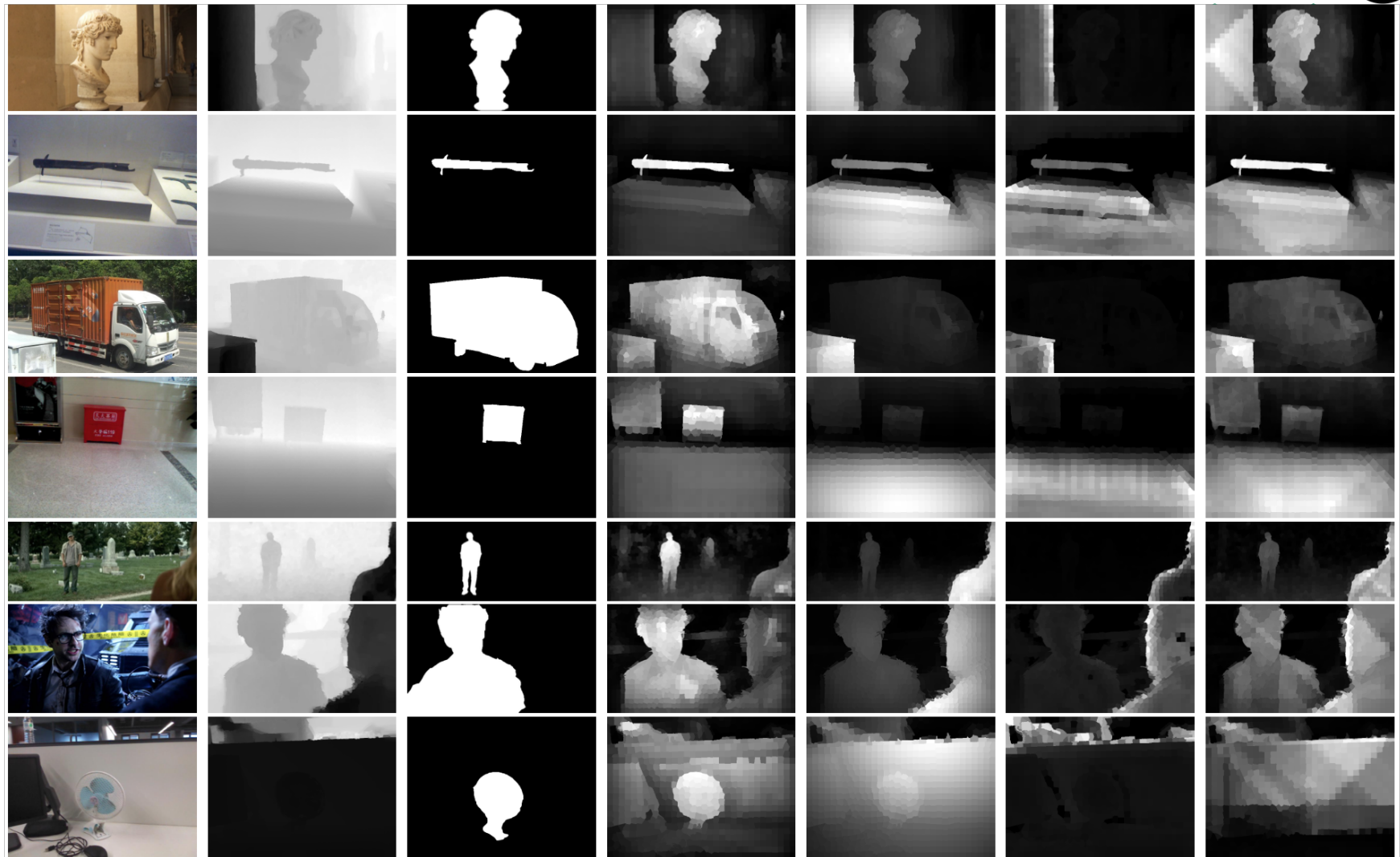
Precision vs Recall on RGBD1000



Precision vs Recall on NJUDS2000



Reducing FP: Low Contrast Foreground



(a) Colour

(b) Depth

(c) Groundtruth

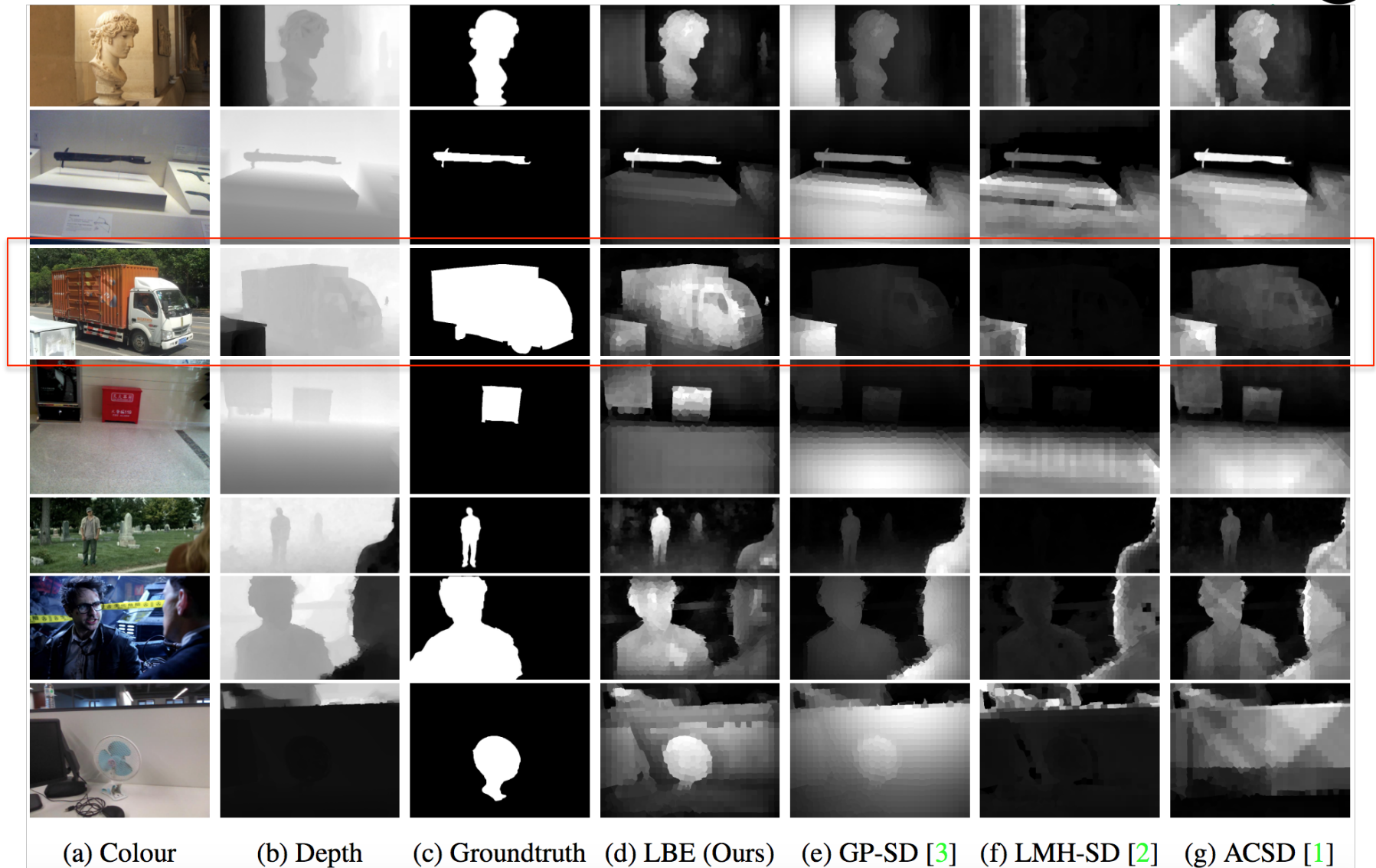
(d) LBE (Ours)

(e) GP-SD [3]

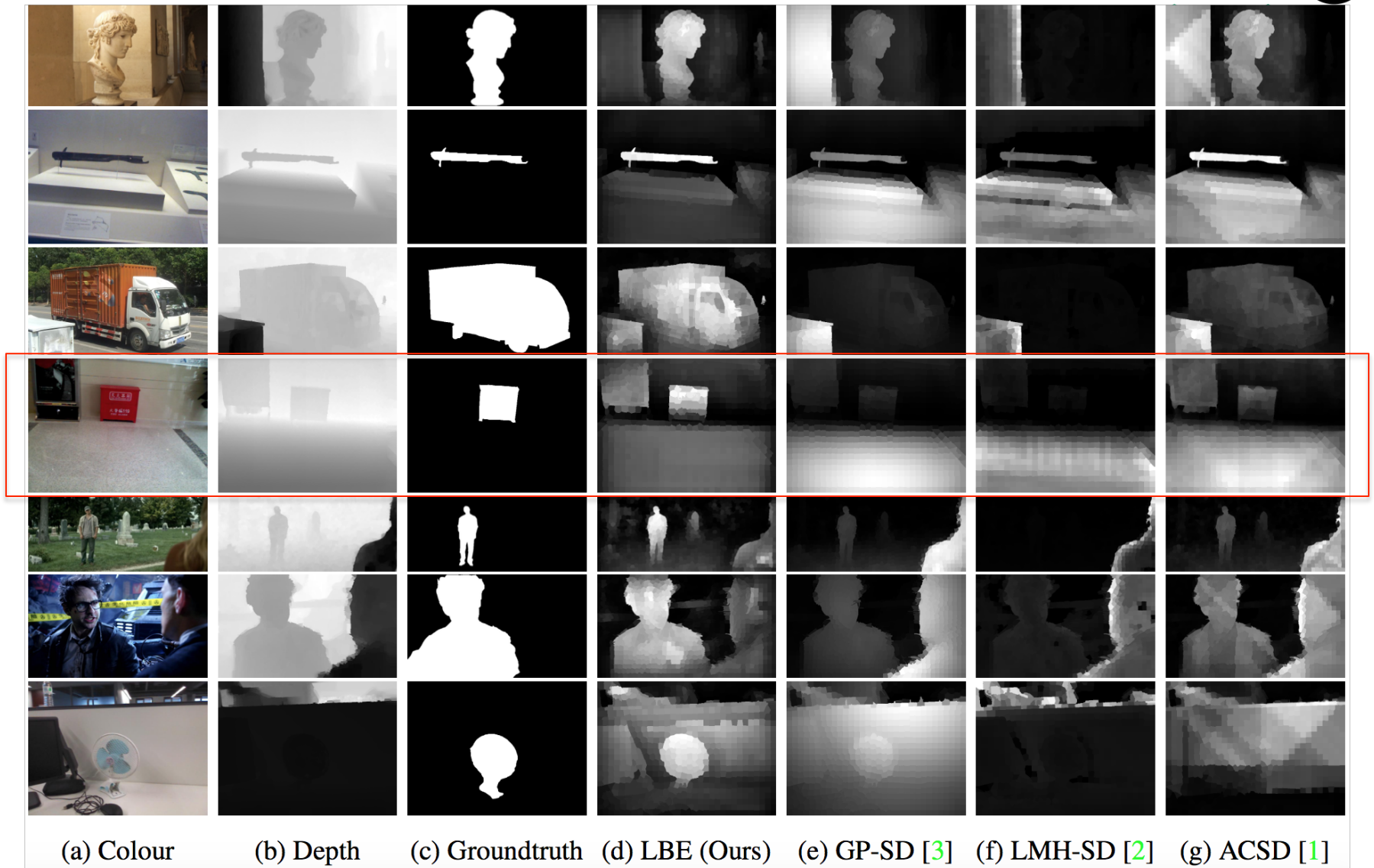
(f) LMH-SD [2]

(g) ACSD [1]

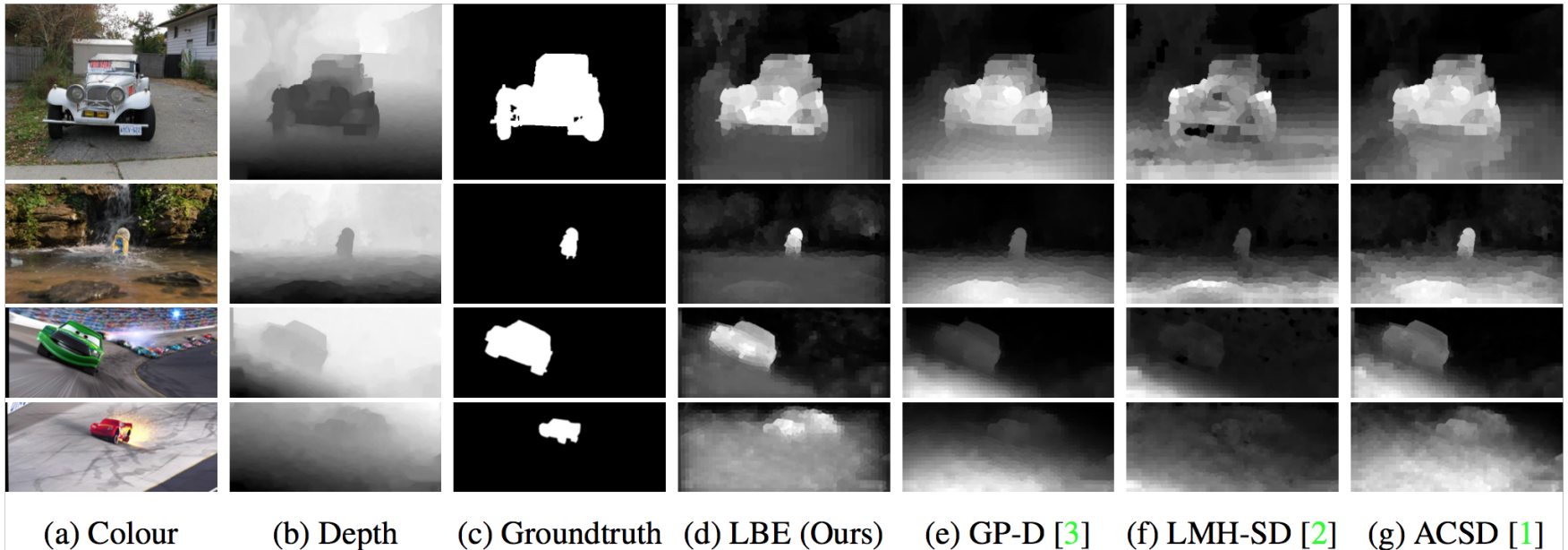
Reducing FP: Low Contrast Foreground



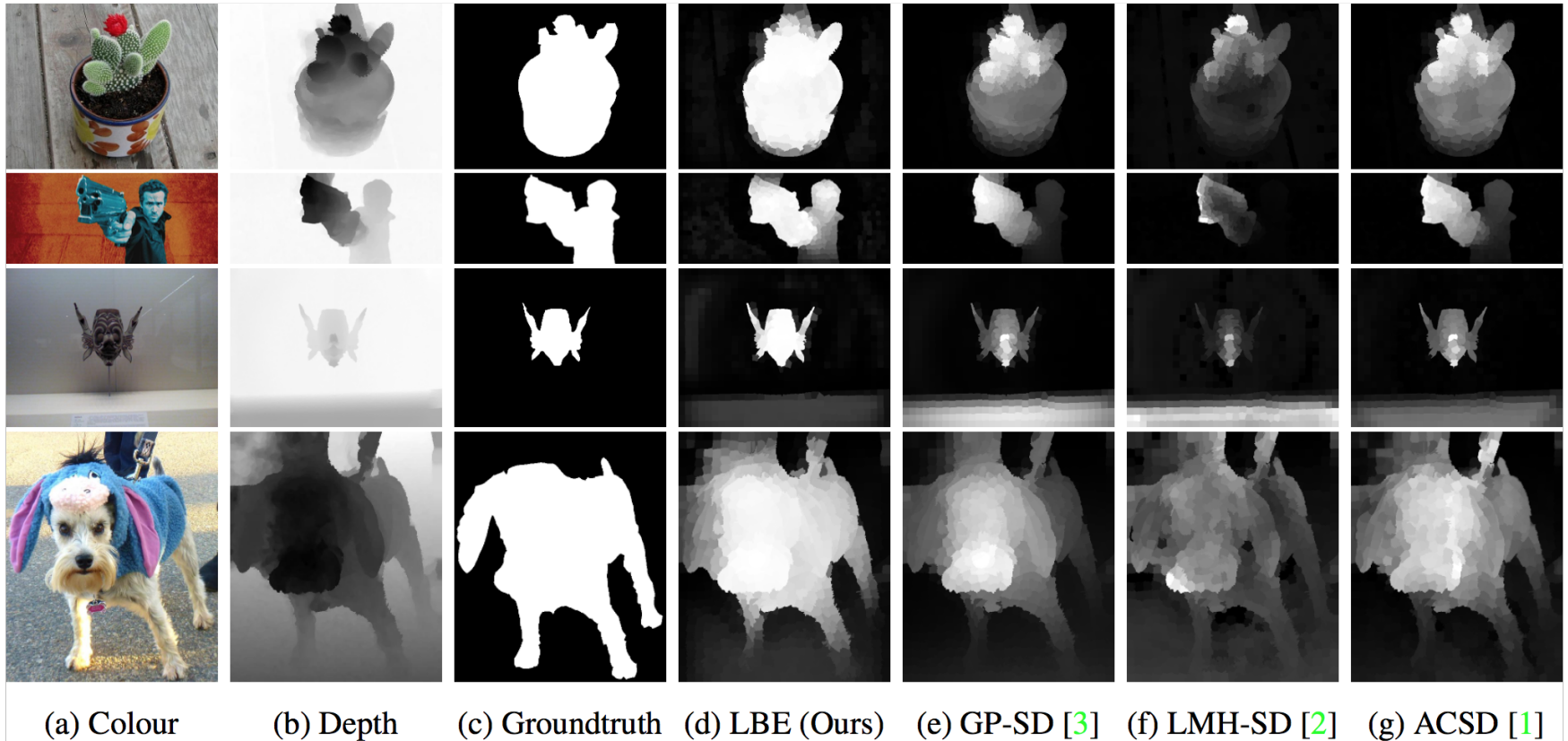
Reducing FP: Low Contrast Foreground



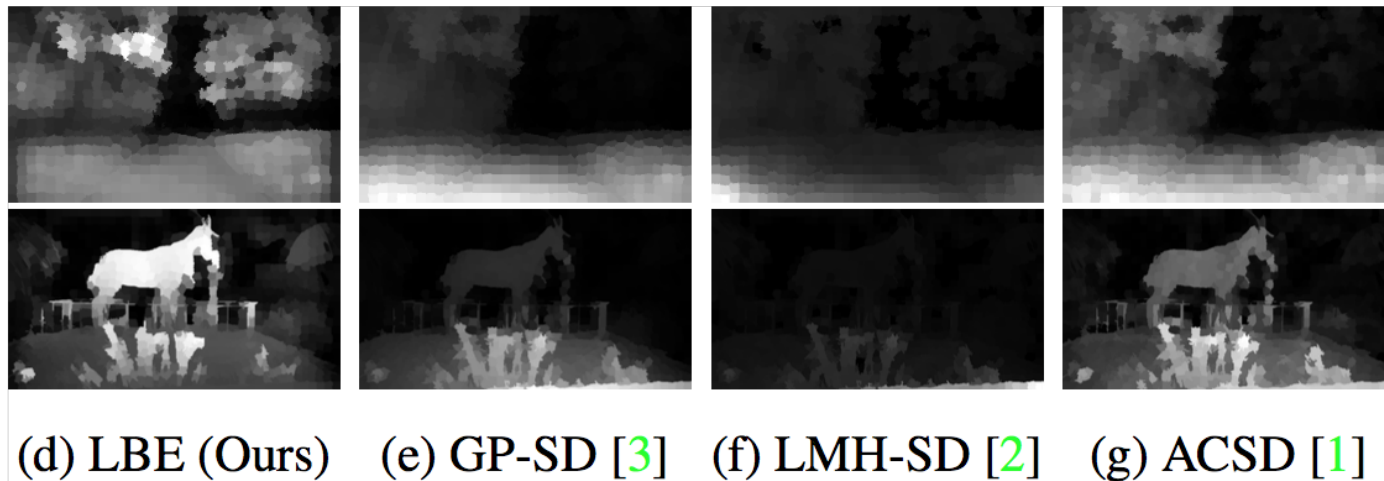
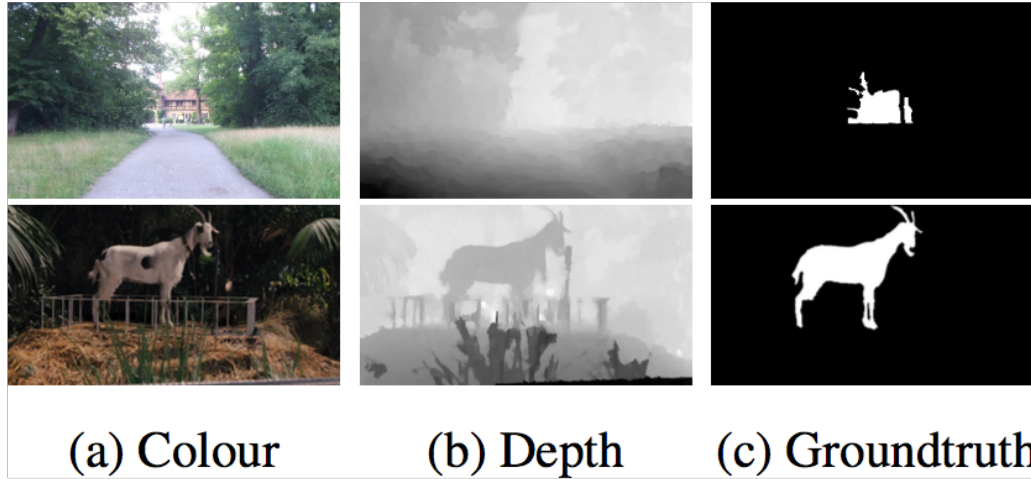
Reducing FP: Angled Planar Surfaces



Reducing FN

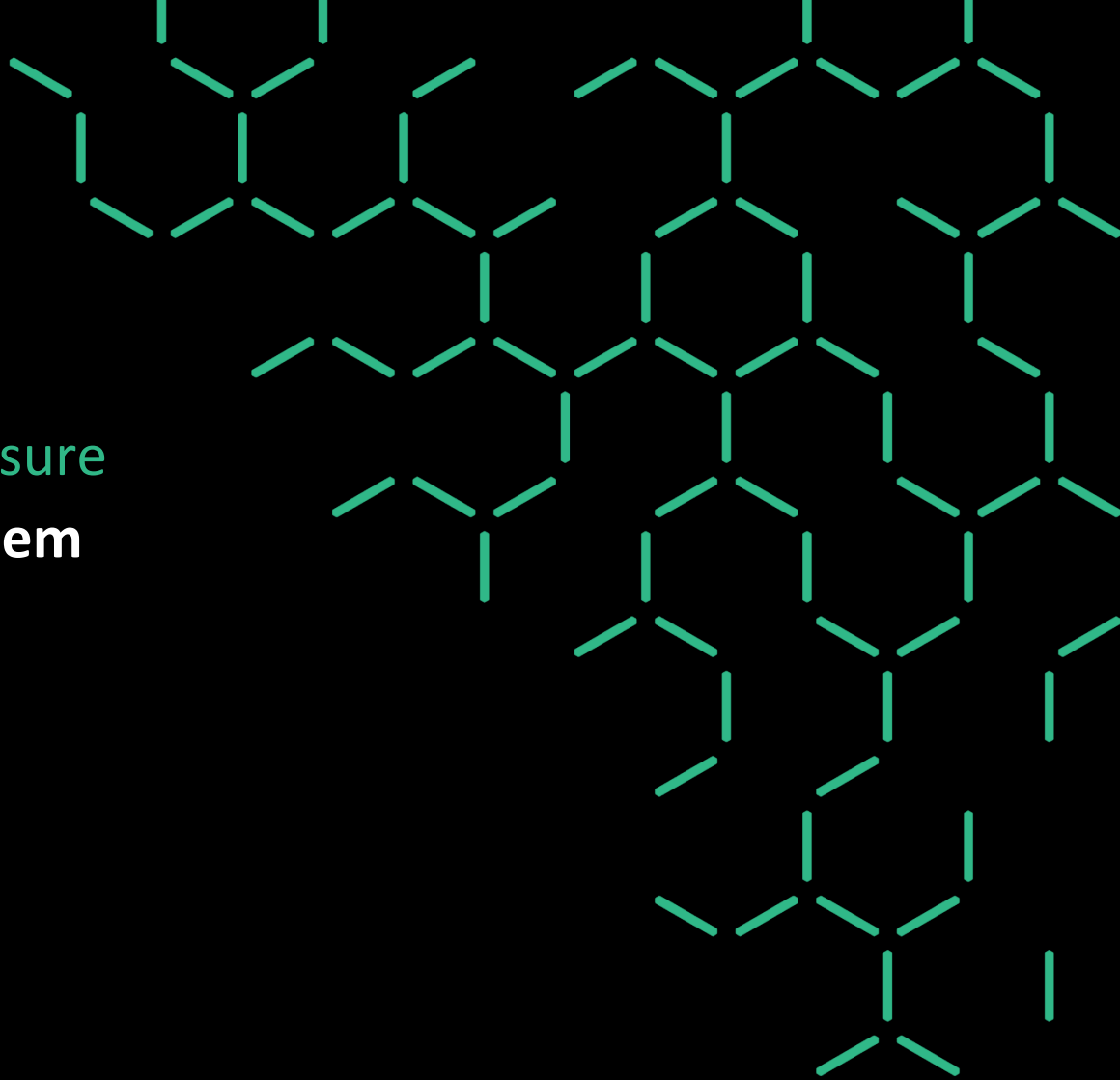


Fail Cases

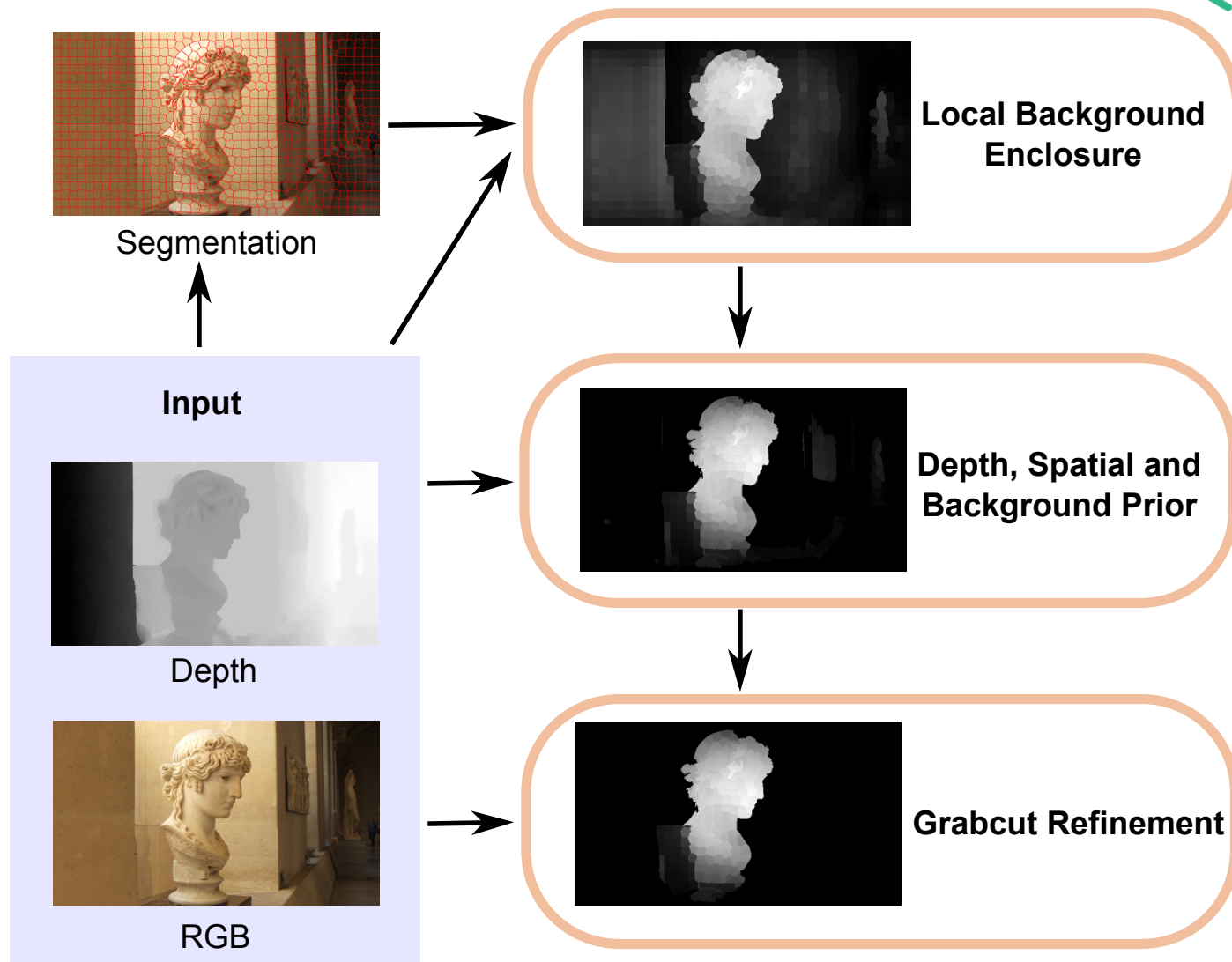


Overview

- Overview of saliency
- Previous work
- Local Background Enclosure
- **Saliency Detection System**
- Results



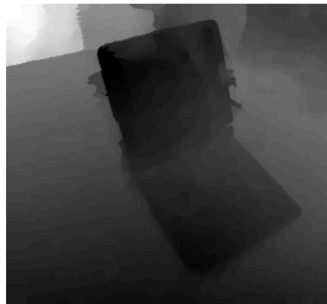
Saliency Detection System: 3 Steps



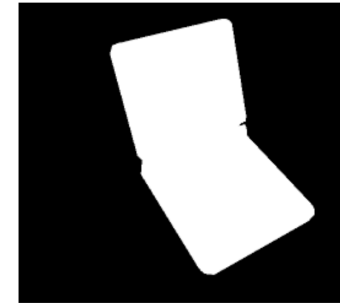
Step 1: LBE Computation



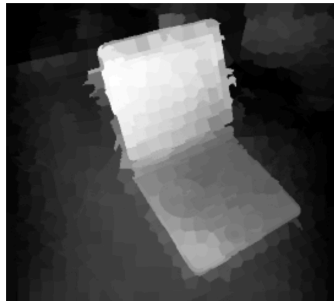
(i) RGB



(ii) Depth



(iii) Ground
Truth

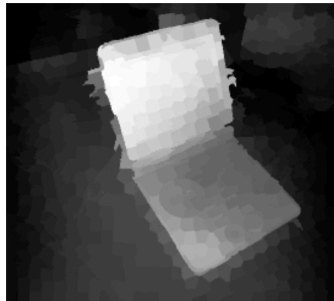


(a) LBE

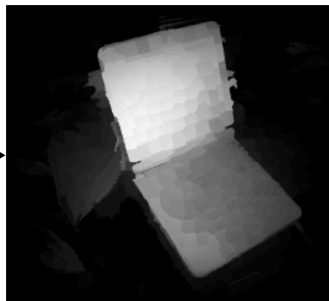
Step 2: Prior Refinement

- Center bias $\mathcal{G}$
- Depth prior $\mathcal{D}$
- Background prior $\mathcal{B}$

$$S_b = S \cdot \mathcal{D} \cdot \mathcal{G} \cdot \mathcal{B} \quad (8)$$



(a) LBE

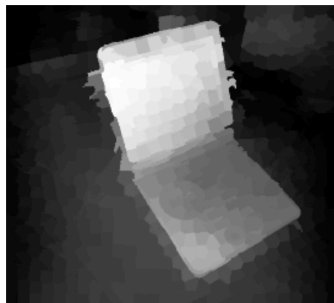


(b) LBE+P

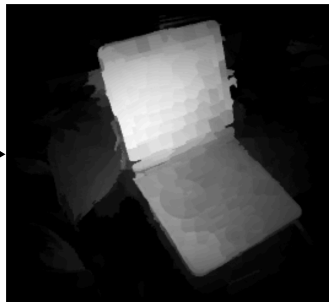
Step 3: Grabcut Refinement

- Obtain refinement mask A by running Grabcut with initialisation from thresholded S_b .

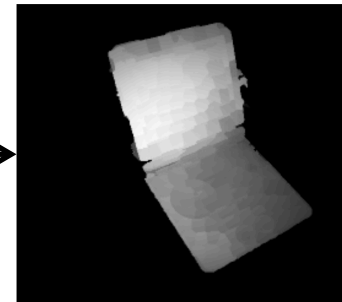
$$S_g = A \cdot S_b. \quad (9)$$



(a) LBE



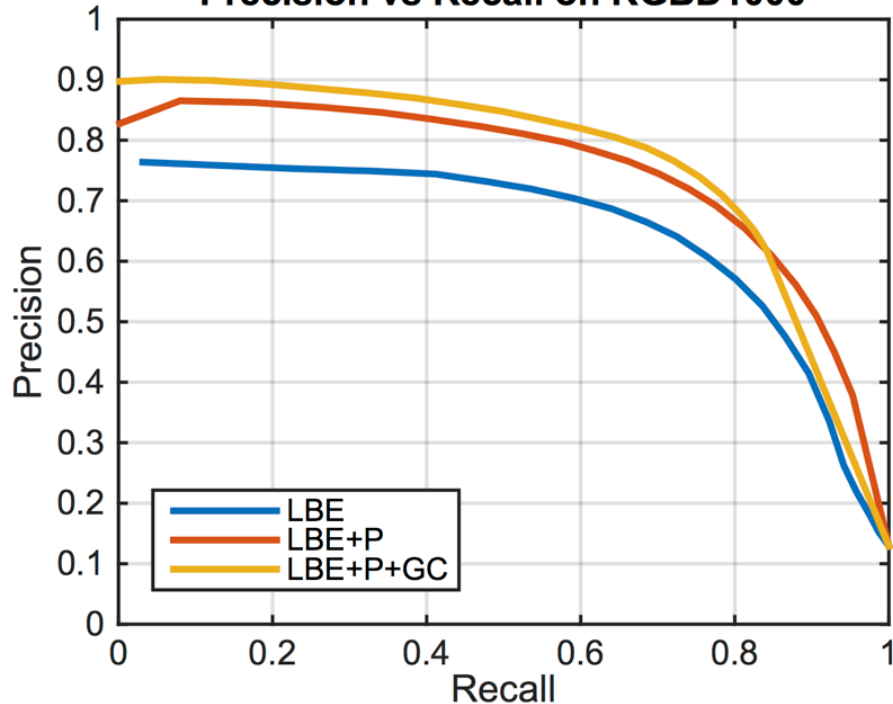
(b) LBE+P



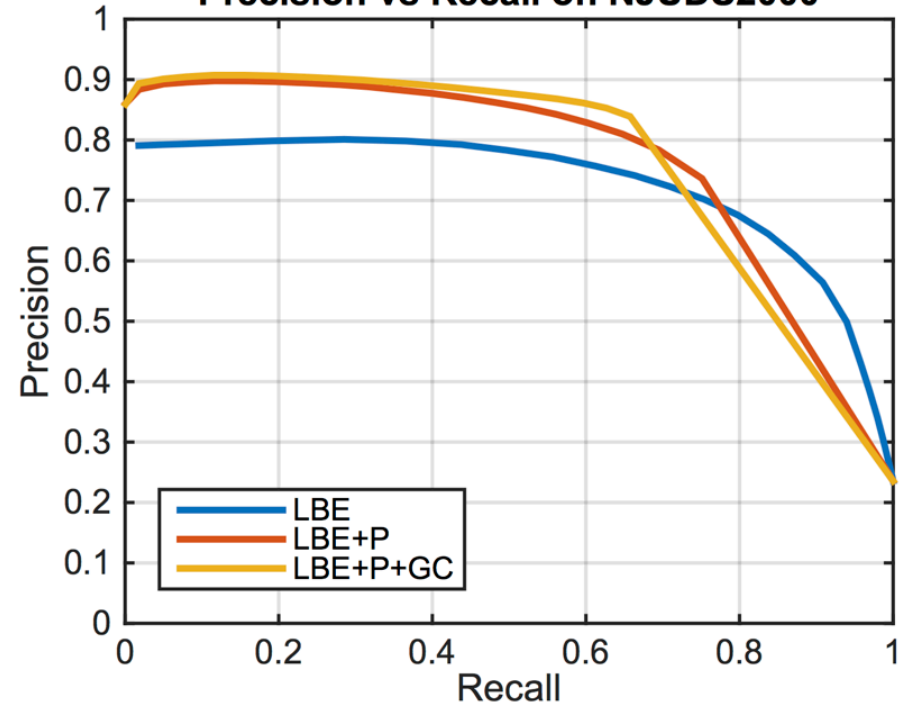
(c) LBE+P+GC

Saliency System Components

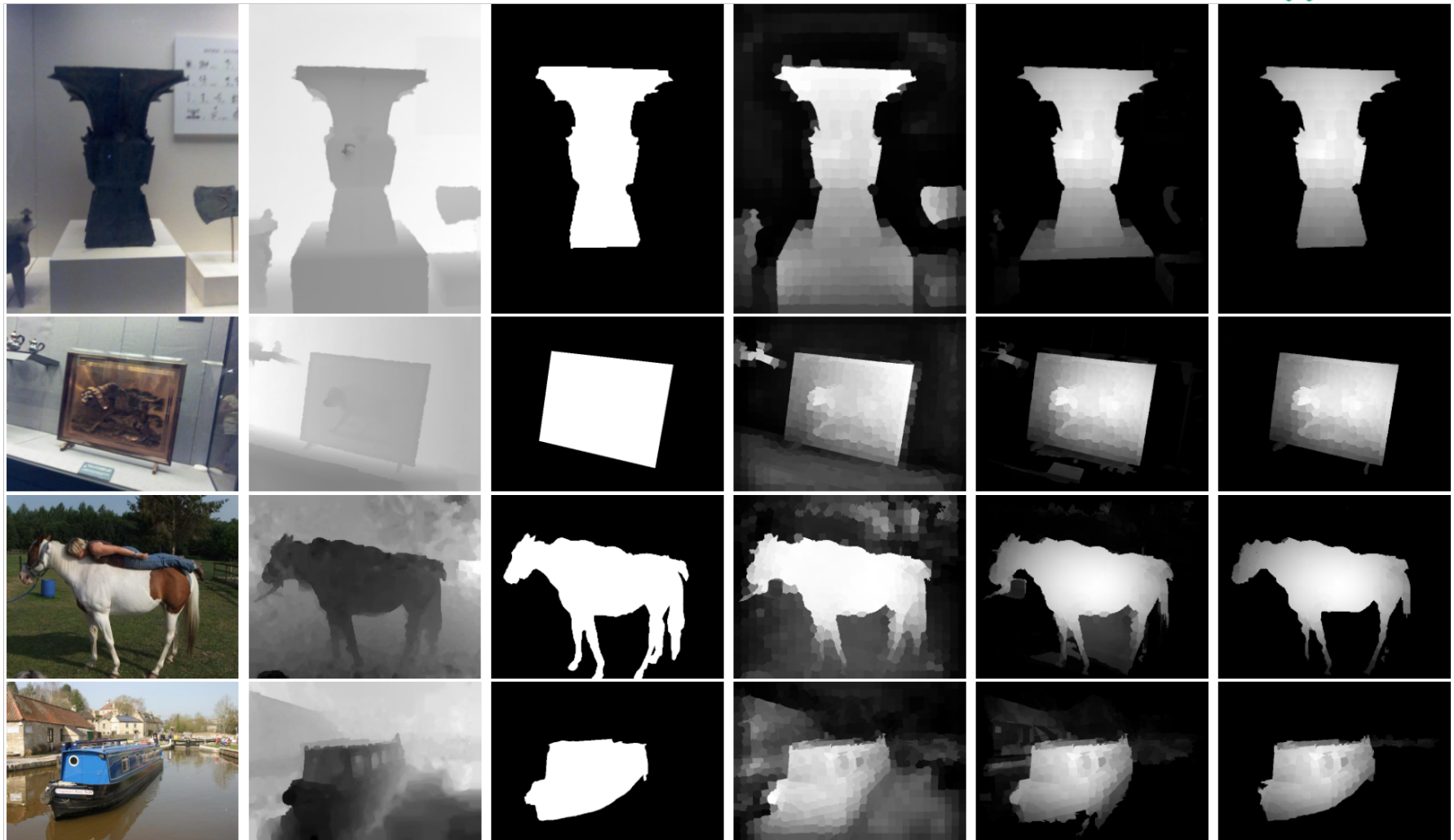
Precision vs Recall on RGBD1000



Precision vs Recall on NJUDS2000



Saliency System Components



(a) RGB

(b) Depth

(c) Ground Truth

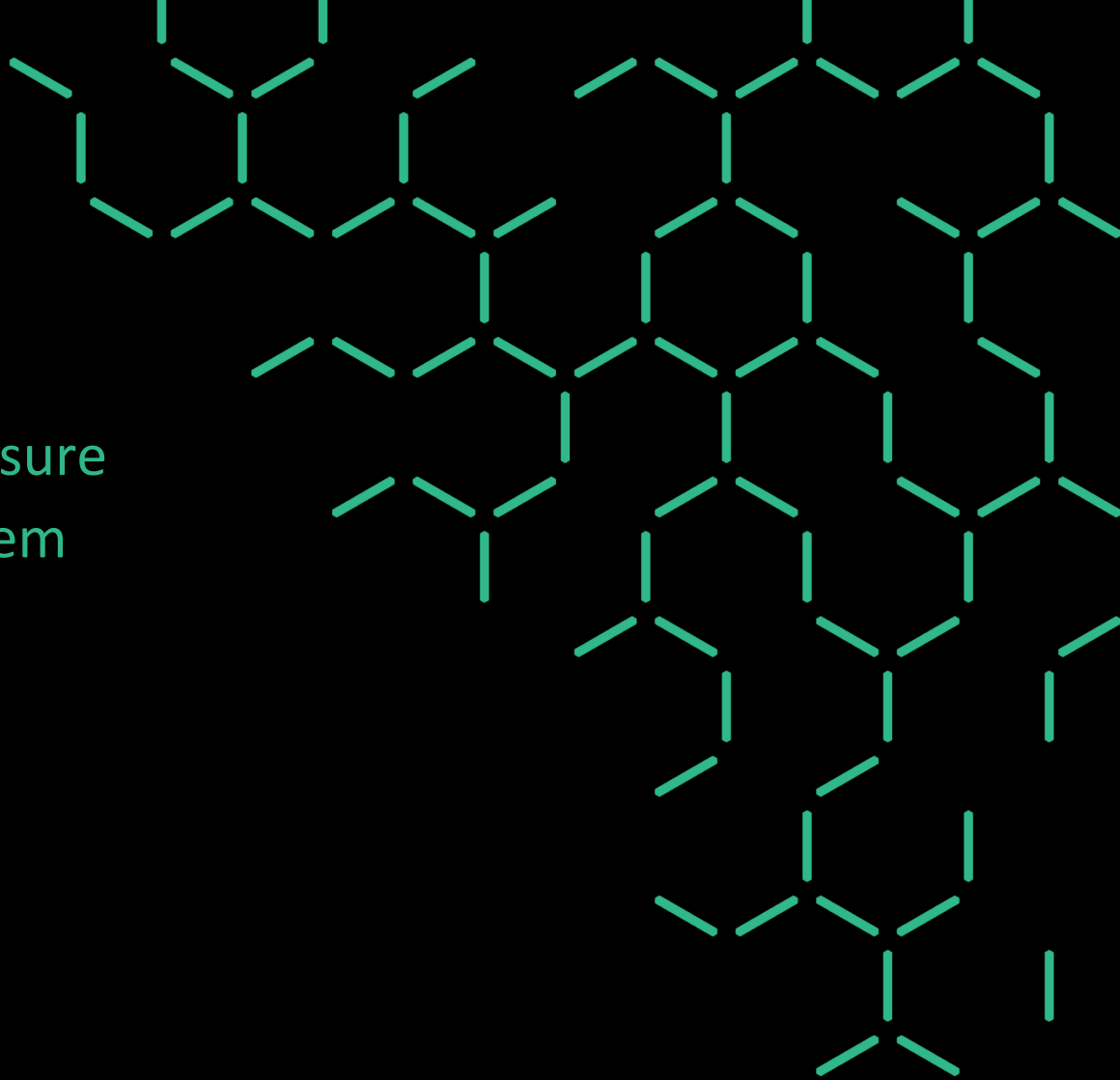
(d) LBE

(e) LBE+P

(f) LBE+P+G

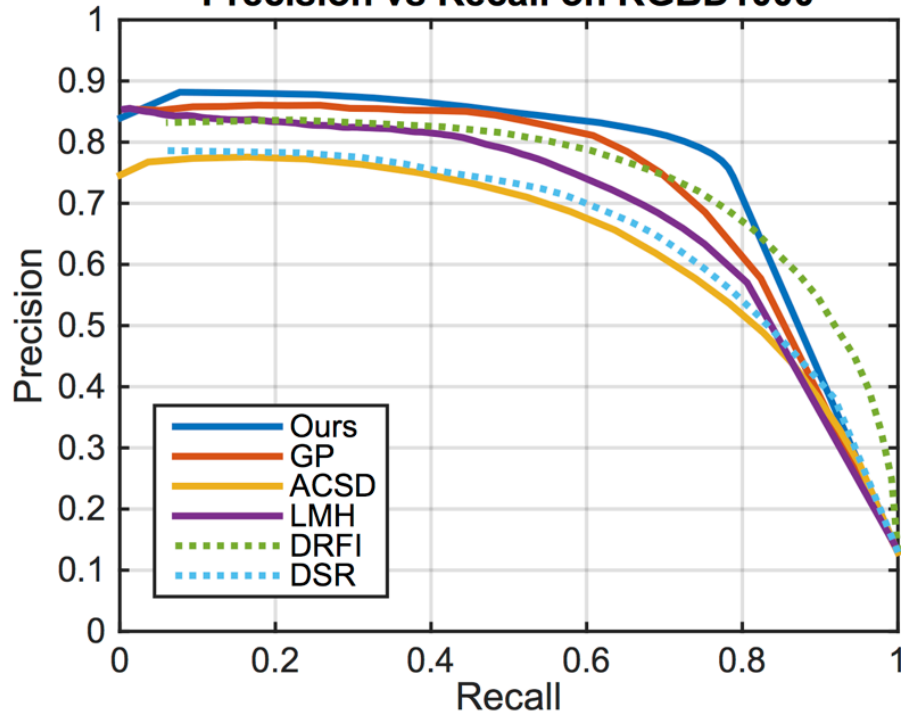
Overview

- Overview of saliency
- Previous work
- Local Background Enclosure
- Saliency Detection System
- **Results**

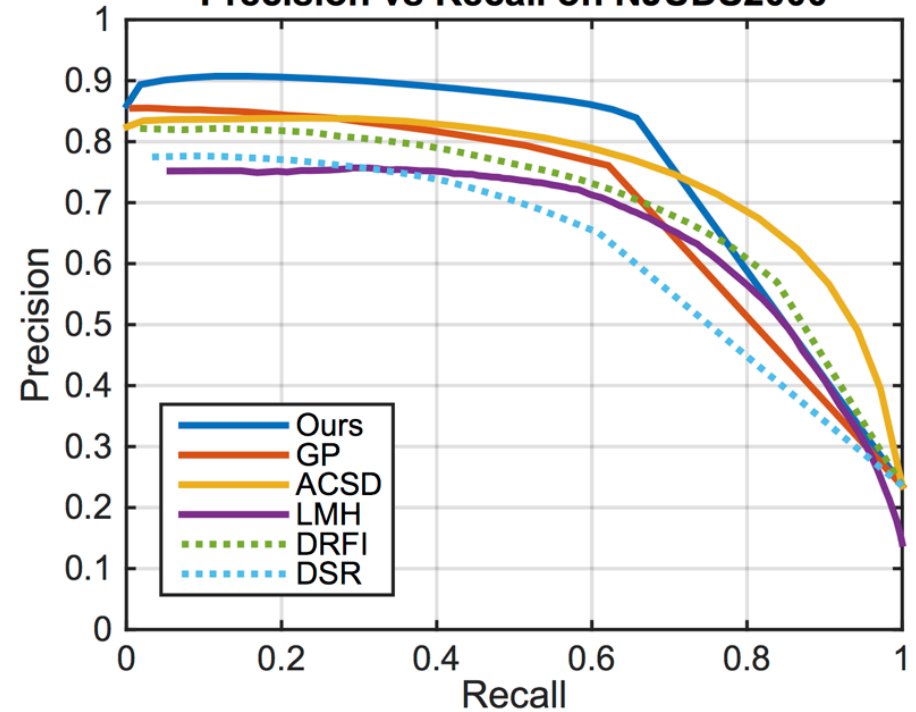


Results

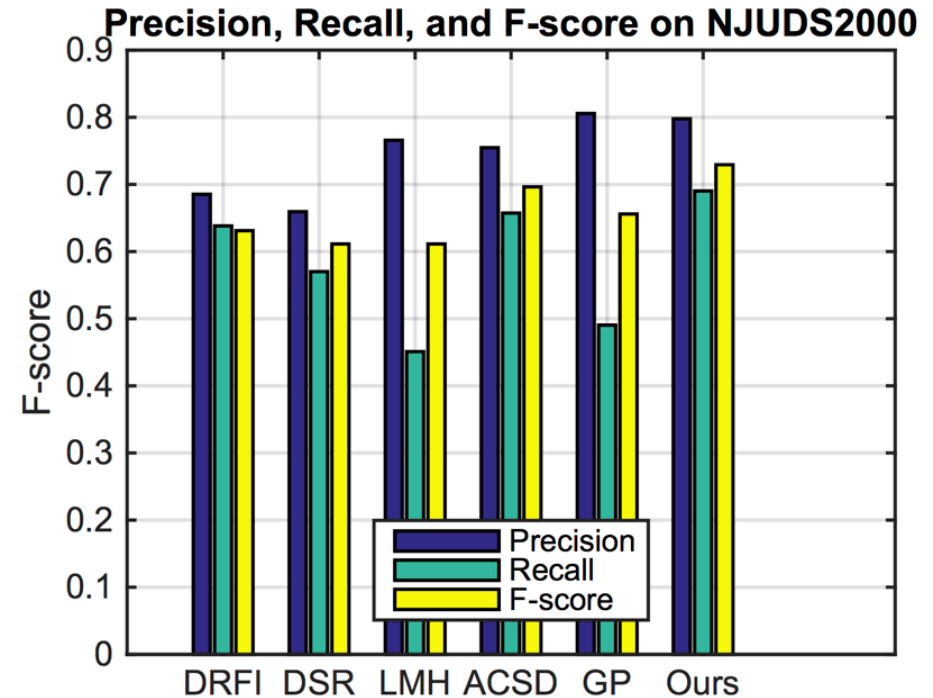
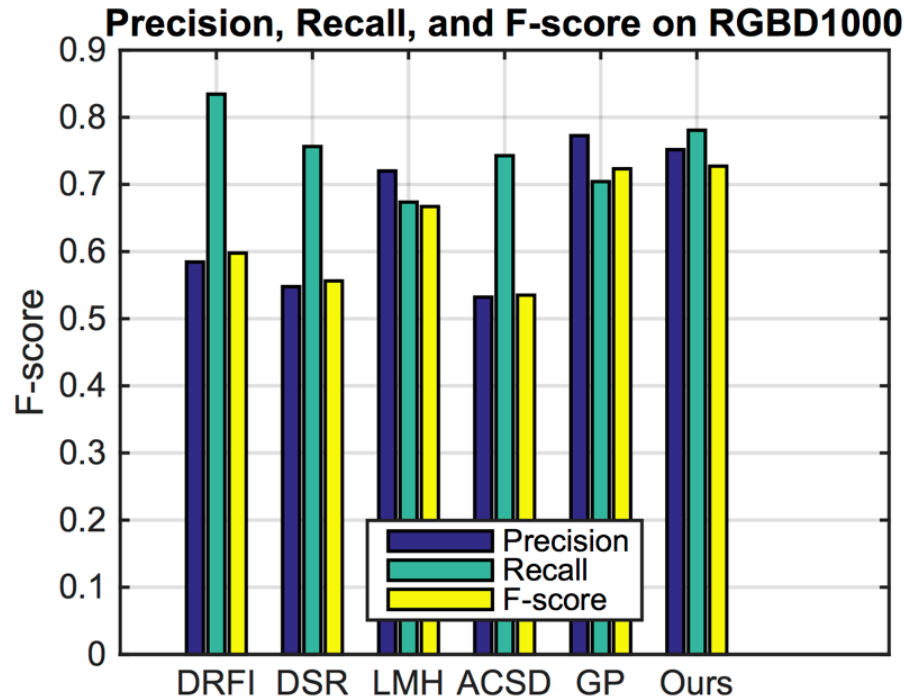
Precision vs Recall on RGBD1000



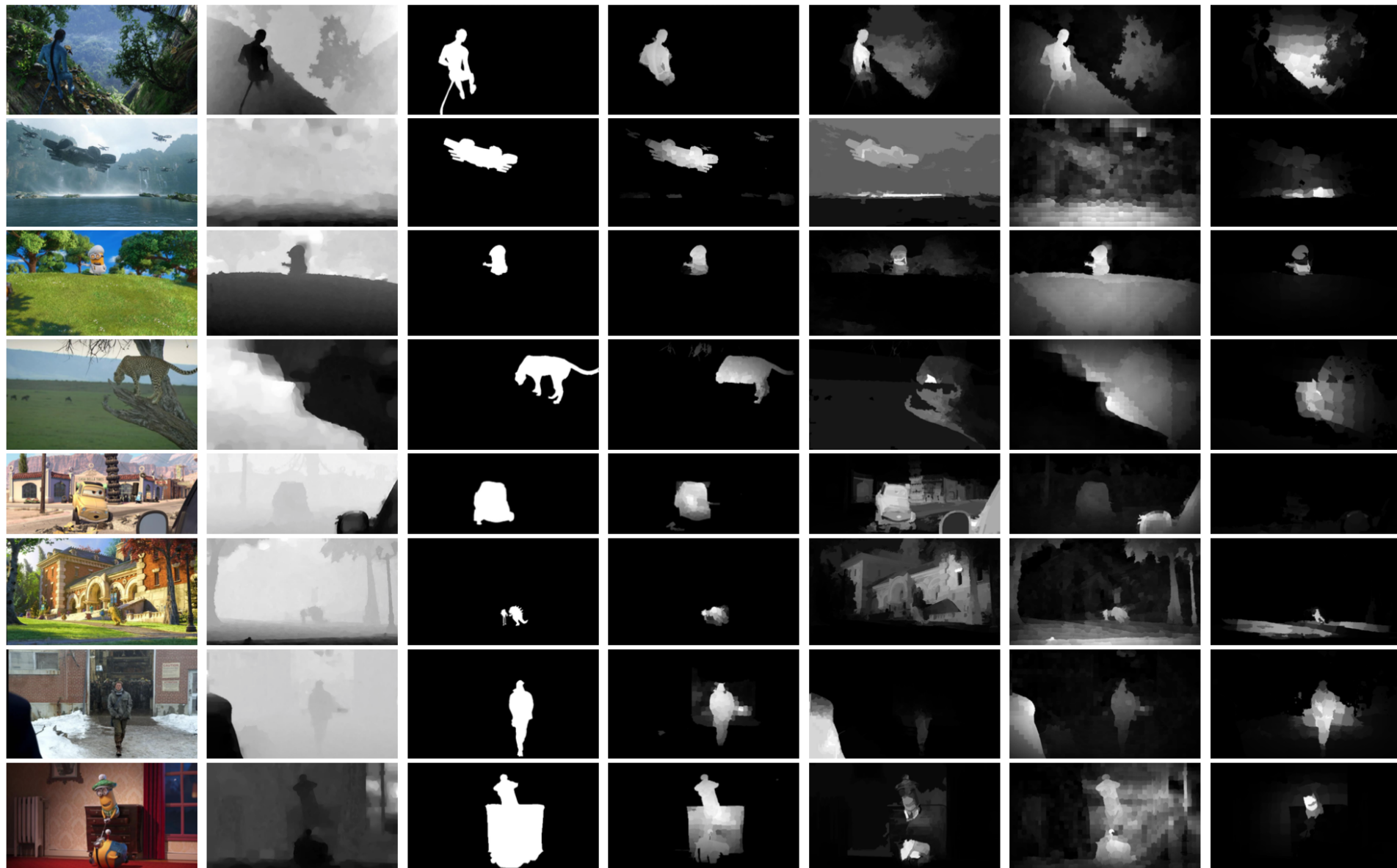
Precision vs Recall on NJUDS2000



Results



Results



(a) RGB

(b) Depth

(c) G. T.

(d) Ours

(e) GP [23]

(f) ACSD [15]

(g) LMH [22]

Summary



- Proposed the depth LBE feature, which measures structural enclosure rather than depth contrast.
- Demonstrated improved performance over depth contrast features.
- Integrated LBE feature into saliency system to achieve superior results compared to state-of-the-art methods on two datasets.