Active 3D Scene Segmentation Using Kinect

Carlos Hernández Matas

Royal Institute of Technology

carloshm@kth.se

July 5, 2011

Outline

- Introduction
- Background
- Novelties
- Experiments
- Discussion

Introduction

What is image segmentation?

Active 3D Scene Segmentation

Björkman and Kragic: *Active 3D scene segmentation and detection of unknown objects*. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2010.

- Markov Random Fields.
- 3-label model: Foreground, Surface and Background.
- Surface hypothesis.
- Iterative segmentation.
- 2 sets of stereo cameras $\rightarrow$ disparity.

Binary Labeling

$$E(\bar{x}) = \sum_{i=1,\ldots,N} \left(E(x_i) + \sum_{x_j \in N(x_i)} (x_i, x_j)\right)$$

Markov Random Fields

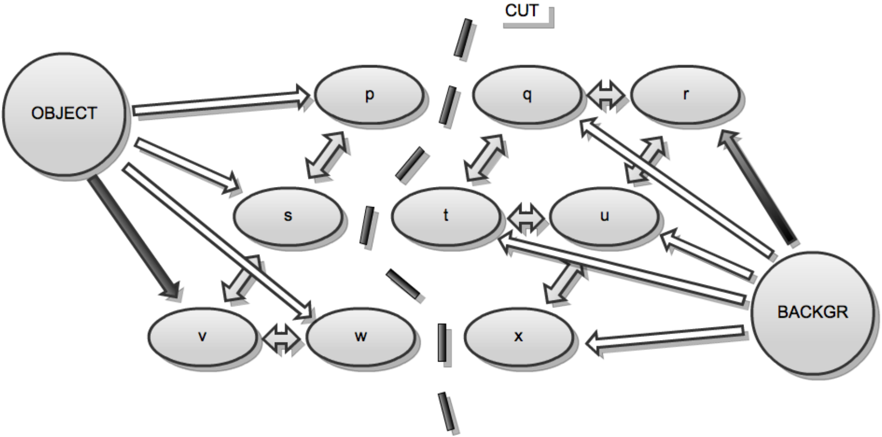


Figure: Markov Random Field

Kinect

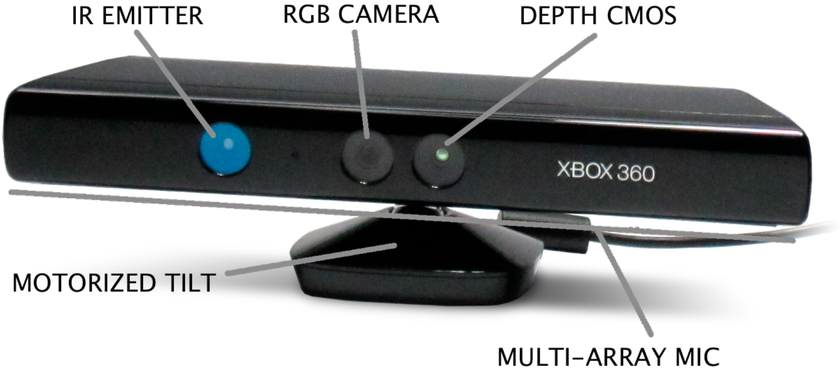
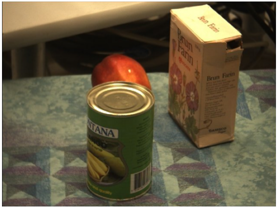
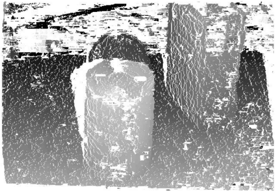


Figure: Kinect Sensor

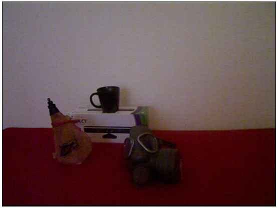
Hardware differences



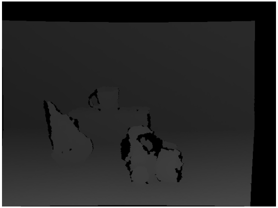
(A)



(B)



(C)



(D)

Figure: Disparities. **(A)**: Color image. **(B)**: Stable matching disparity. **(C)**: Color image. **(D)**: Kinect disparity.

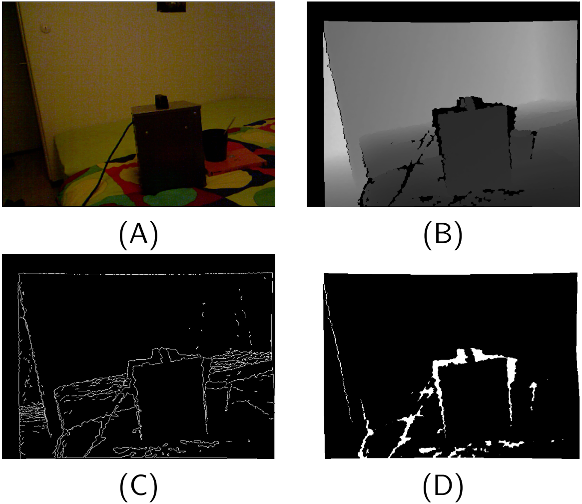
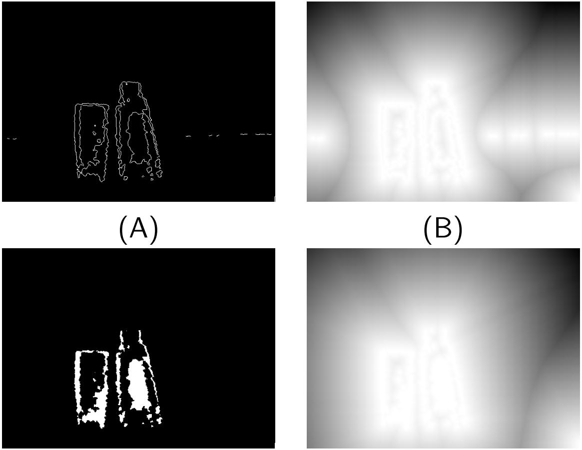


Figure: Cues. **(A)**: Color image. **(B)**: Depth image. **(C)**: Canny cue. **(D)**: Halo cue.

Cue transformation



(A) (B)

(C) (D)

Figure: D.T. **(A)**: Canny. **(B)**: Canny D.T. **(C)**: Halo. **(D)**: Halo D.T.

Energy functions

Potts model for penalty (energy / cost function)

$$p(l_i, l_j)_{l_i \neq l_j} = \exp^{-V_{i,j}} \tag{1}$$

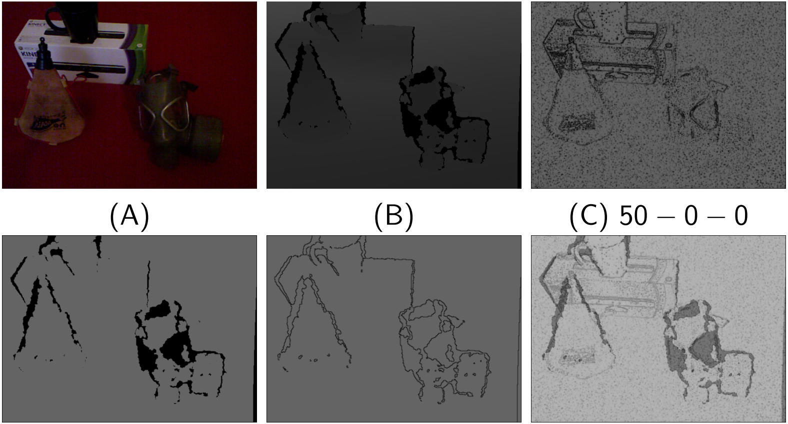
Direct application and Gauss

$$\begin{aligned} V_{i,j} &= V_{i,j}^L + V_{i,j}^C + V_{i,j}^L = \\ &= \lambda^L \exp^{-\beta^L (v_i^L - v_j^L)^2} + \lambda^C \exp^{-(v_i^C v_j^C)} + \lambda^H \exp^{-(v_i^H v_j^H)} \\ \beta^L &= (2\langle (v_i^L - v_j^L)^2 \rangle)^{-1} \end{aligned} \tag{2}$$

Distance transform

$$V_{i,j} = \lambda^L \exp^{-\beta^L (v_i^L - v_j^L)^2} + \lambda^C \exp^{-\frac{v_i^C v_j^C}{255^2}} + \lambda^H \exp^{-\frac{v_i^H v_j^H}{255^2}} \tag{3}$$

Energy map



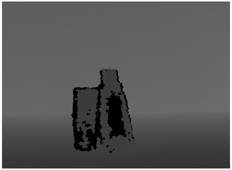
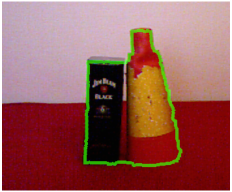
(A) (B) (C) $50 - 0 - 0$

(D) $0 - 50 - 0$ (E) $0 - 0 - 50$ (F) $30 - 30 - 30$

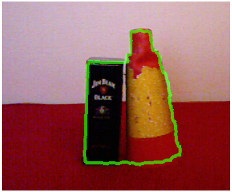
Figure: Clutterednb energy map. **(A)**: Color image. **(B)**: Depth map. **(C)** – **(F)**: Energy maps for different values of $\lambda^L - \lambda^H - \lambda^C$.

Iterative segmentation

http://www.youtube.com/watch?v=NrdCriH6xIA

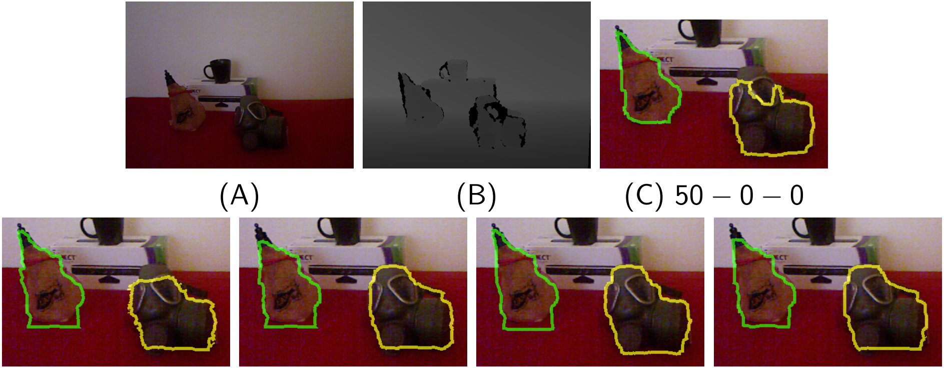
(A) (B) (C) $50 - 0 - 0$

(D) $0 - 30 - 0$ (E) $0 - 0 - 30$ (F) $0 - 30 - 30$ (G) $30 - 30 - 30$

Figure: JimConeClose segmentations. **(A)**: Color image. **(B)**: Depth image. **(C)** $-$ **(G)**: Segmentation results for different values of $\lambda^L - \lambda^H - \lambda^C$.

Cues comparison Cluttered



(A) (B) (C) 50 – 0 – 0

(D) 0 – 50 – 0 (E) 0 – 0 – 50 (F) 0 – 10 – 10 (G) 10 – 10 – 30

Figure: Cluttered segmentations. **(A)**: Color image. **(B)**: Depth image. **(C)** – **(G)**: Segmentation results for different values of $\lambda^L - \lambda^H - \lambda^C$.

Canny Gauss

(A) 0 – 0 – 30 Normal

(B) 0 – 0 – 30 Normal

(C) 0 – 0 – 50 Normal

(D) 0 – 0 – 30 Gauss

(E) 0 – 0 – 30 Gauss

(F) 0 – 0 – 50 Gauss

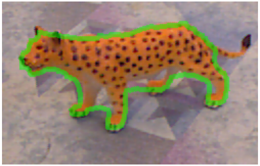
Figure: Canny Gauss. **(A)** – **(C)**: Segmentation results for different images using normal cues. **(D)** – **(F)**: Segmentation results for different images using Gaussian blur.

Canny Distance Transform

(A) $0 - 0 - 30$ Normal

(B) $0 - 0 - 30$ Normal

(C) $0 - 50 - 0$ Normal



(D) $0 - 0 - 30$ D.T.

(E) $0 - 0 - 30$ D.T.

(F) $0 - 50 - 0$ D.T:

Figure: Canny distance transform. **(A)** – **(C)**: Segmentation results for different images using normal cues. **(D)** – **(F)**: Segmentation results for different images using distance transform.

Halo Gauss

(A) 0 – 30 – 0 Normal

(B) 0 – 30 – 0 Normal

(C) 0 – 50 – 0 Normal

(D) 0 – 30 – 0 Gauss

(E) 0 – 30 – 0 Gauss

(F) 0 – 50 – 0 Gauss

Figure: Halo Gauss. **(A)** – **(C)**: Segmentation results for different images using normal cues. **(D)** – **(F)**: Segmentation results for different images using Gaussian blur.

Halo Distance Transform

(A) 0 – 30 – 0 Normal

(B) 0 – 30 – 0 Normal

(C) 0 – 50 – 0 Normal

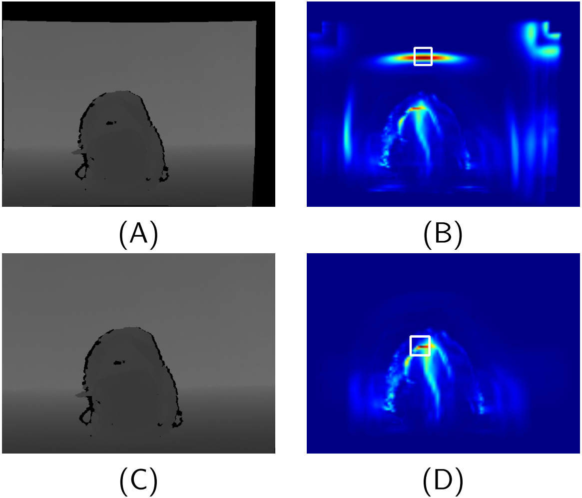
(D) 0 – 30 – 0 D.T.

(E) 0 – 30 – 0 D.T.

(F) 0 – 50 – 0 D.T.

Figure: Halo distance transform. **(A)** – **(C)**: Segmentation results for different images using normal cues. **(D)** – **(F)**: Segmentation results for different images using distance transform.

Saliency cropping



(A) (B)

(C) (D)

Figure: Backpack symmetry saliency. **(A)** and **(B)**: Original depth map and its symmetry saliency. **(C)** and **(D)**: Cropped depth map and its symmetry saliency.

Saliency

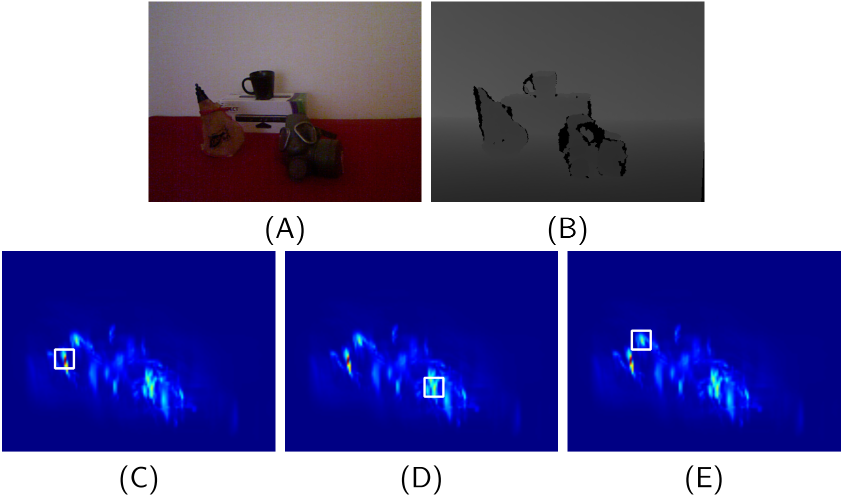


Figure: Cluttered symmetry saliency. **(A)**: Color image. **(B)**: Depth map. **(C)** – **(E)**: Symmetry saliency for the first three peaks.

Discussion

- Symmetry saliency.
- Canny.
- Halo.
- Cue combination.

Questions

Thank you for listening.

Any questions?