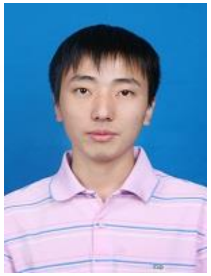
PCM -Nanjing 2013

Stereo GrabCut: Object Extraction for Stereo Images

Ran Ju



Xiangyang Xu

Yang Yang

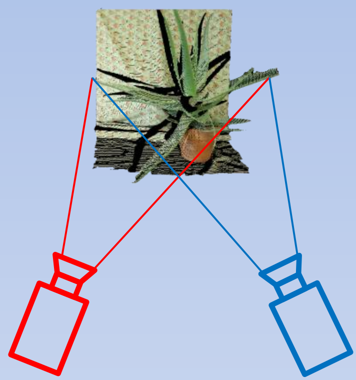
Gangshan Wu

State Key Laboratory for Novel Software Technology, Nanjing University

juran@smail.nju.edu.com

Stereo image: an old and young media

- A stereo image is made up of two images taken from two slightly different views to simulate the human stereo vision.

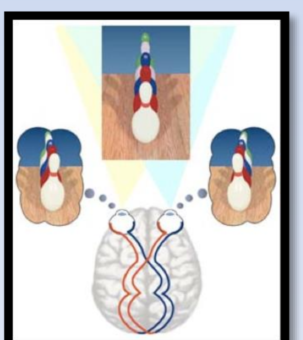


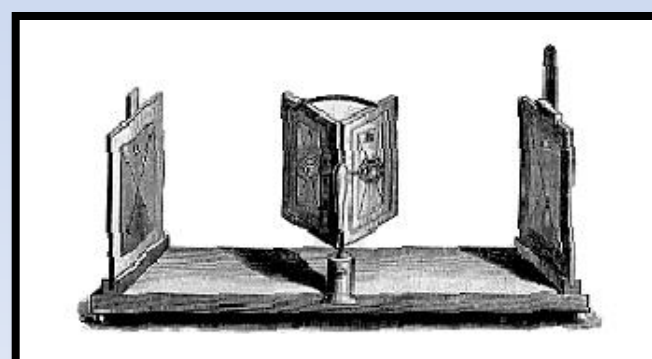
Left camera Right camera

Left image

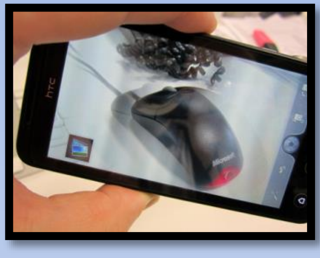
Right image

- Based on binocular disparity, which is first described by Sir Charles Wheaston in 1838.





Easier to acquire, display and transfer



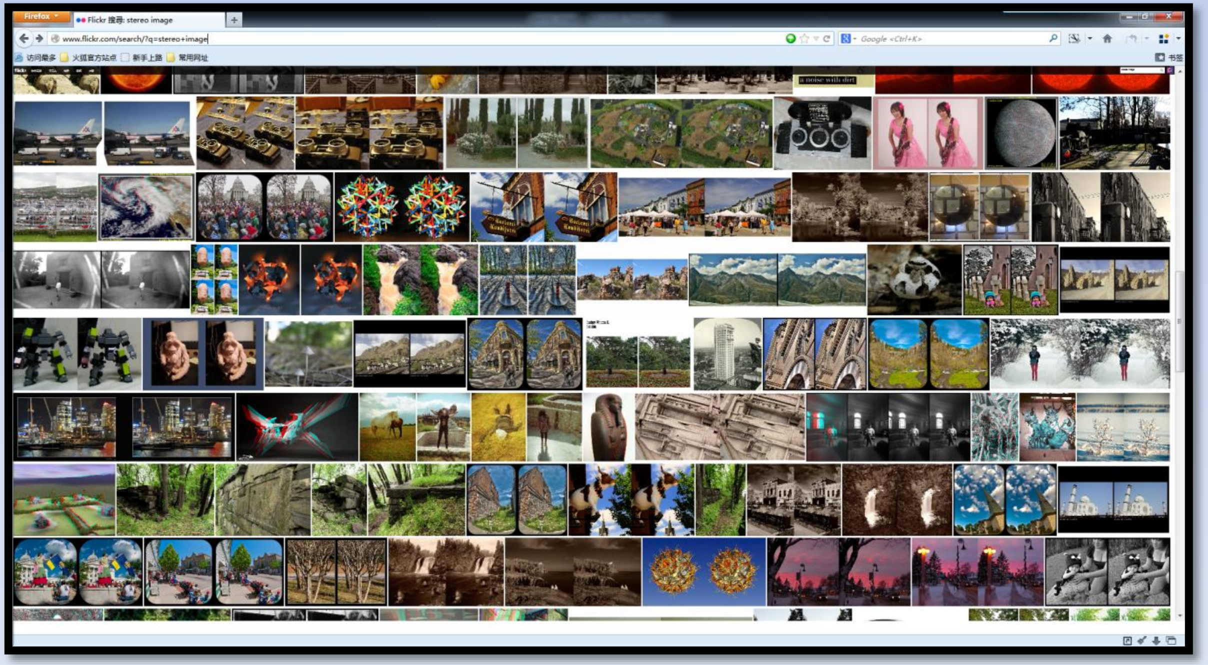
HTC Evo 3D——the first 3D smartphone

Stereo camera

Stereo digital video camera

3D projector

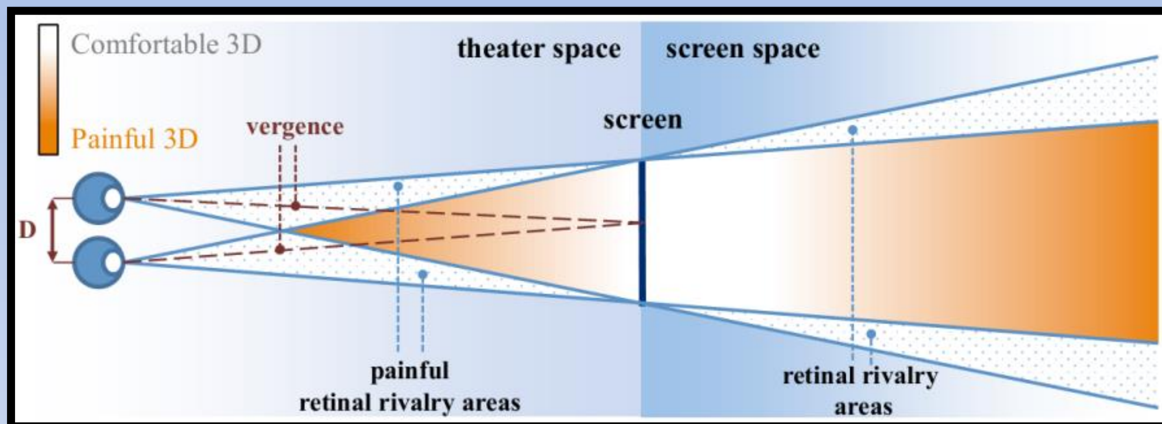
3D TV

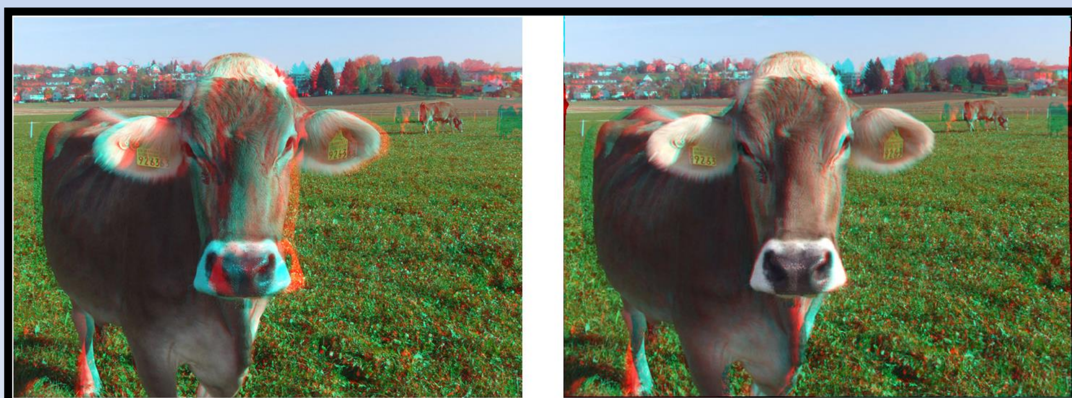


Search results on Flickr

Applications

- 3D movie production

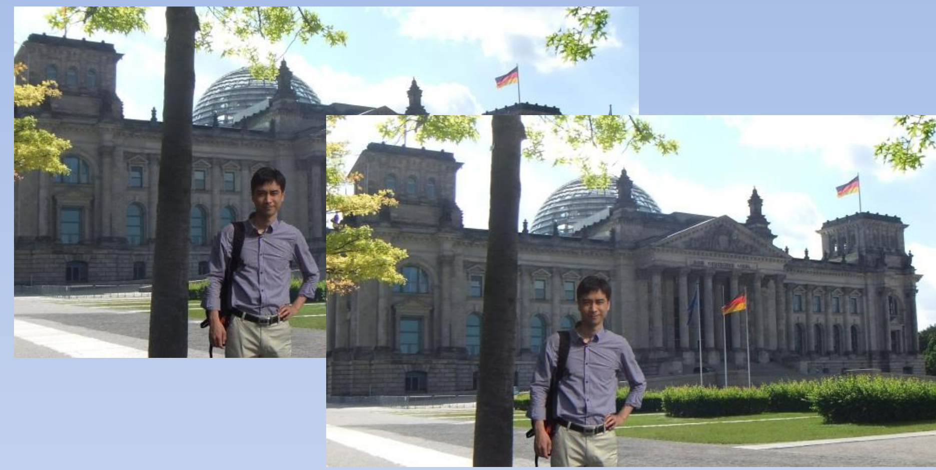




Remap the disparity range for comfortable 3D viewing experience.

Applications

- Image editing



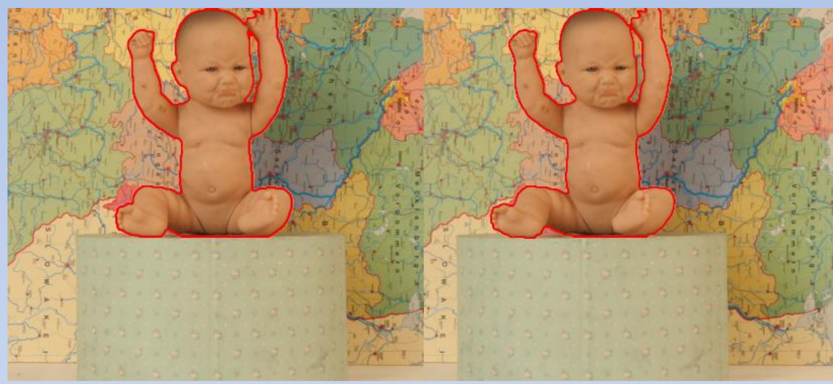
Background replacement

Stereo photo

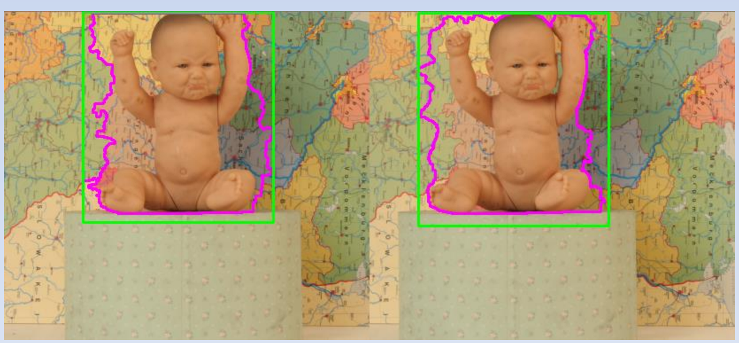
Color popping

Object extraction for stereo images

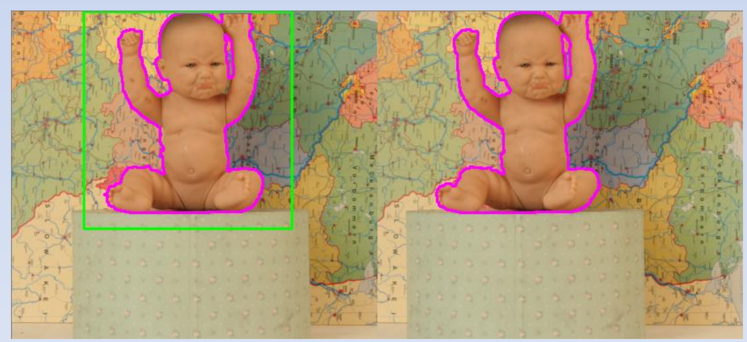
- Segmentation should be consistent for both views.
- Stereo images have implicit depth information.



- Comparison of GrabCut and Stereo GrabCut



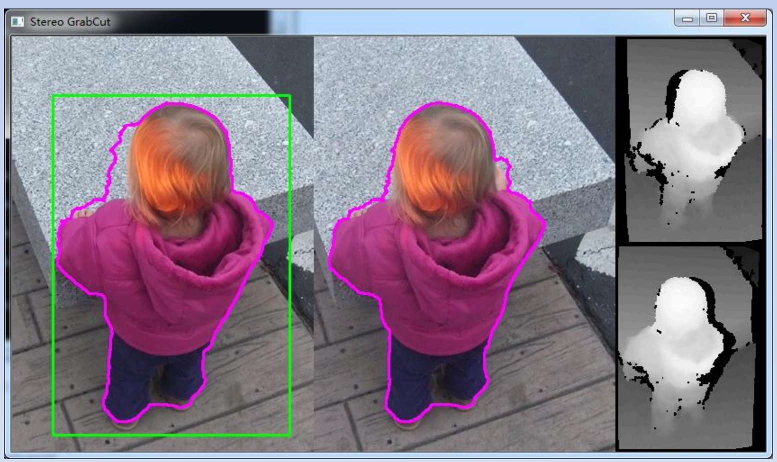
GrabCut[1]



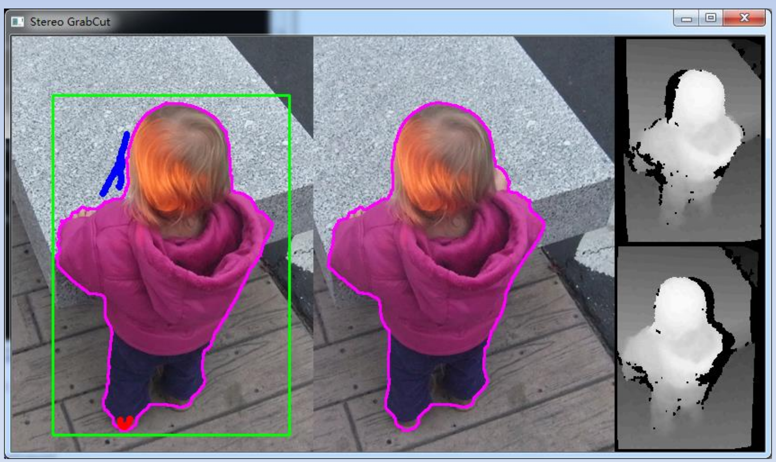
Stereo GrabCut

User interface

Step 1. The user drags a compact rectangle around an object to get an initial segmentation.
Step 2. The user scribble with a foreground and background brush to revise the initial result.



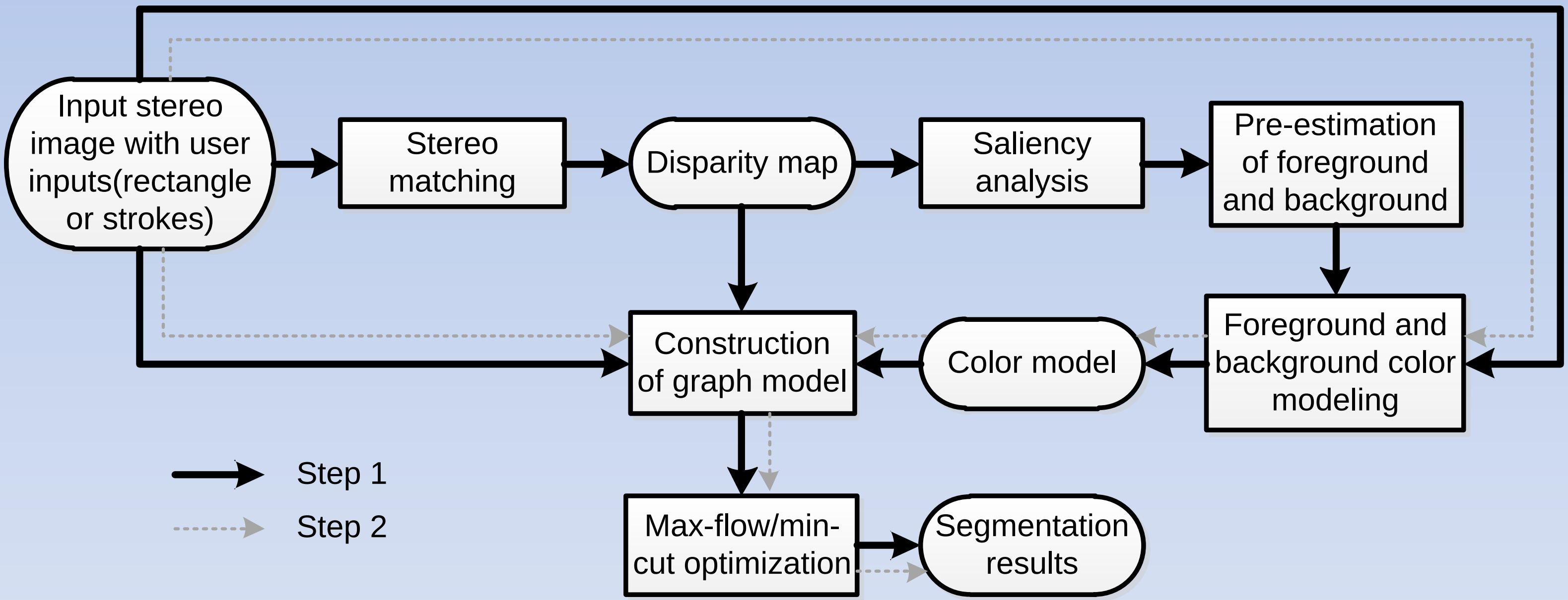
Step 1



Step 2

Approach

- Establish a correspondence term using stereo matching.
- Pre-estimation of foreground/background from depth map using saliency analysis.



Stereo matching

- Stereo matching
  - Accurate
  - Fast

Left image　　Right image　　Matching points　　Disparity map

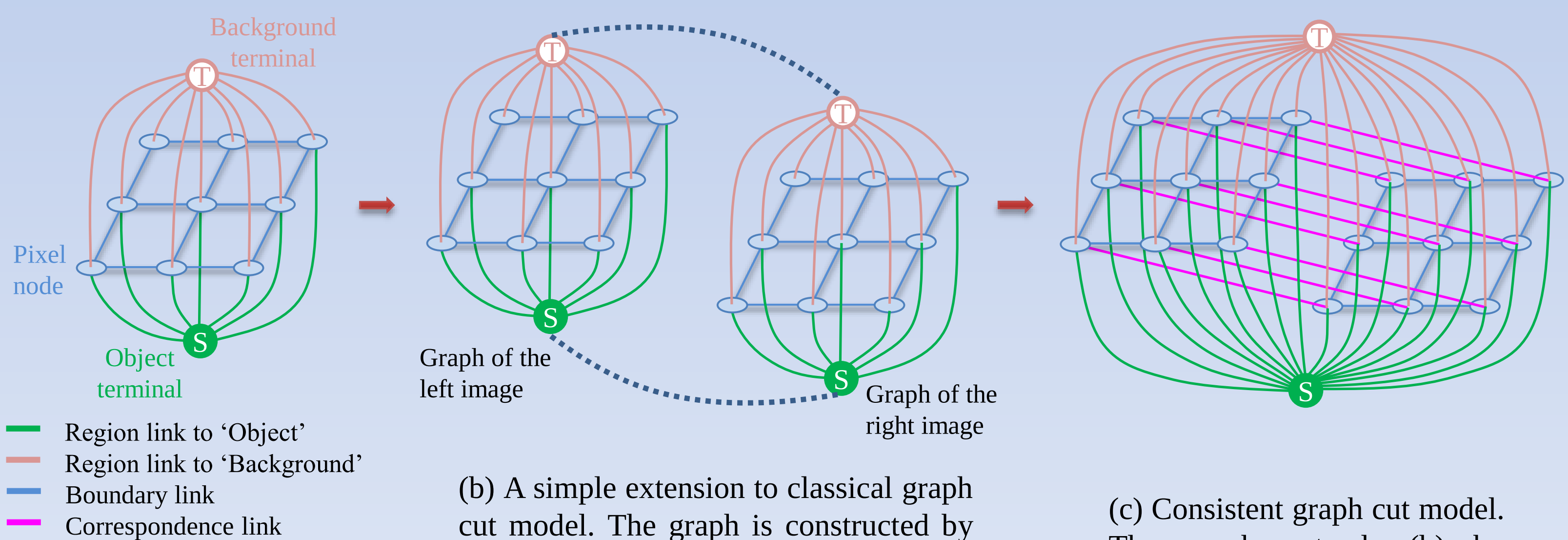
- ELAS[2] is a GCP(ground control points) based algorithm and works in nearly real time.

Consistent graph cut

- Global energy function

$$E(A) = \underbrace{\sum_{p\in P_l \cup P_r} R_p(A_p)}_{\text{Region term}} + \underbrace{\lambda_B \sum_{\{p,q\}\in N_B} B_{\{p,q\}}|A_p - A_q|}_{\text{Boundary term}} + \underbrace{\lambda_C \sum_{\{p_l,q_r\}\in N_C} C_{\{p_l,q_r\}}|A_{p_l} - A_{q_r}|}_{\text{Consistency term}}$$

- Graph cut model[3][4]



(a) Classical graph cut model

(b) A simple extension to classical graph cut model. The graph is constructed by simply linking the graphs of the left image and the right image at the terminal nodes.

(c) Consistent graph cut model. The graph extends (b) by adding correspondence edges.

Depth saliency and pre-estimation

- Basic assumption: salient regions are more likely to differ from background in depth.
- Saliency value definition:

$$S(p_i) = \sum_{p_k \in S_B} |d(p_i) - d(p_k)|, \qquad p_i \in S_O$$

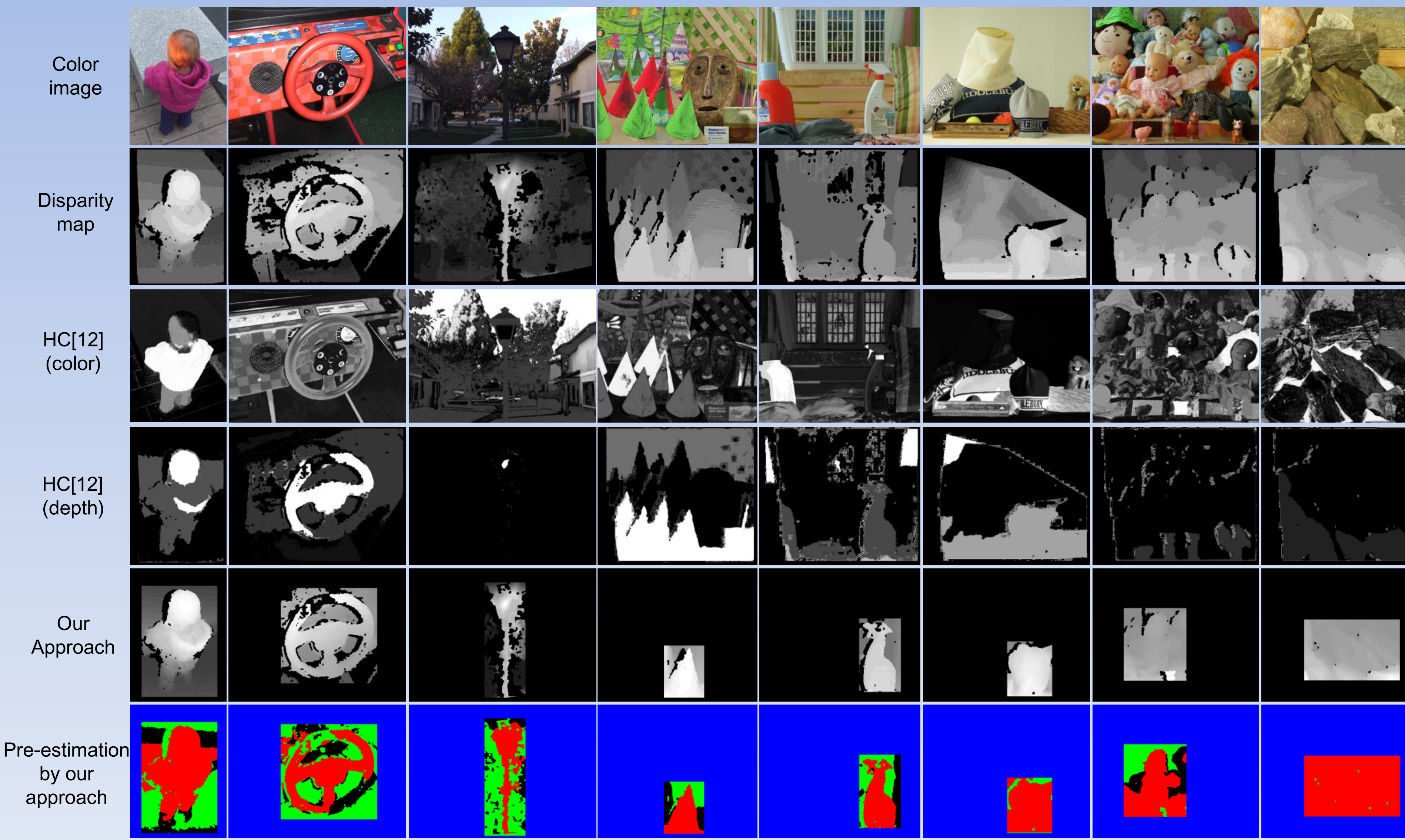
- Histogram based speed up:

$$S_D(d(p_i)) = \sum_{0 \le p_k \le D_{max}} f_{d_k} |d_k - d(p_i)|, \qquad p_i \in S_O$$

- Pre-estimation of foreground/background:

$$L_p = \begin{cases} \text{"Object"} & if\ S_D(p_i) \ge S_f \\ \text{"Background"} & if\ S_D(p_i) \le S_b \\ \text{"Unsure"} & otherwise \end{cases}$$

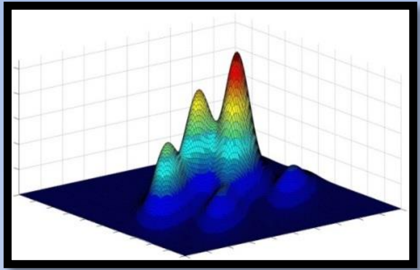
Depth saliency and pre-estimation



FG/BG color modeling

- Gaussian mixture model:

$$P(c_i|\mu,\Sigma) = \sum_{k=1}^{K} \pi_k N(c_i|\mu_k,\Sigma_k)$$



- Use the pre-estimation of FG/BG as samples and initialize the color model using K-means

- Global energy function

$$E(A) = \sum_{p\in P_l \cup P_r} R_p(A_p) + \lambda_B \sum_{\{p,q\}\in N_B} B_{\{p,q\}}|A_p - A_q| + \lambda_C \sum_{\{p_l,q_r\}\in N_C} C_{\{p_l,q_r\}}|A_{p_l} - A_{q_r}|$$

- Region term

$$R_p(A_p) = \begin{cases} -\log P(c_p|\mu_F,\Sigma_F), & if\ A_p \in Foreground \\ -\log P(c_p|\mu_B,\Sigma_B), & if\ A_p \in Background \end{cases}$$

Boundary and correspondence term

- Global energy function

$$E(A) = \sum_{p \in P_l \cup P_r} R_p(A_p) + \lambda_B \sum_{\{p,q\} \in N_B} B_{\{p,q\}} |A_p - A_q| + \lambda_C \sum_{\{p_l,q_r\} \in N_C} C_{\{p_l,q_r\}} |A_{p_l} - A_{q_r}|$$

- Boundary term

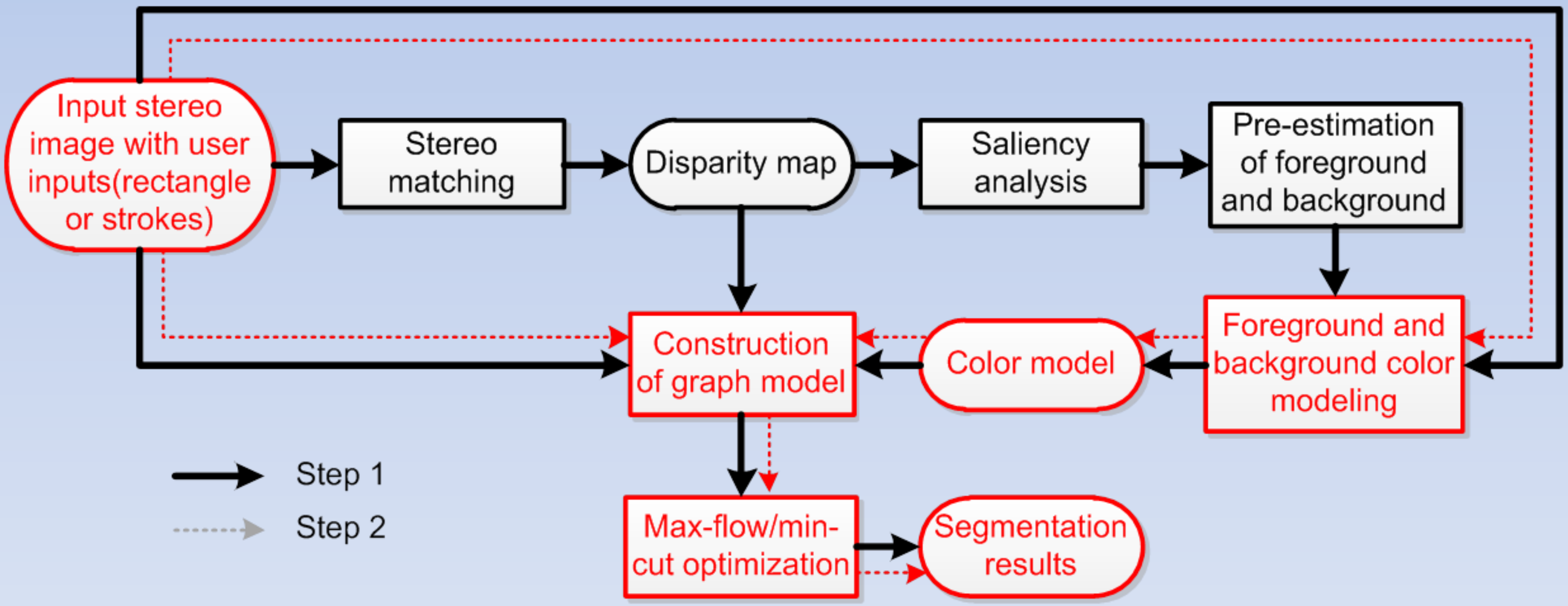
$$B_{\{p,q\}} = \exp\left(-\left\| c_p - c_q \right\|_2\right)$$

- Correspondence term

$$C_{\{p_l,q_r\}} = \exp\left(-\left\| c_{p_l} - c_{q_r} \right\|_2\right)$$

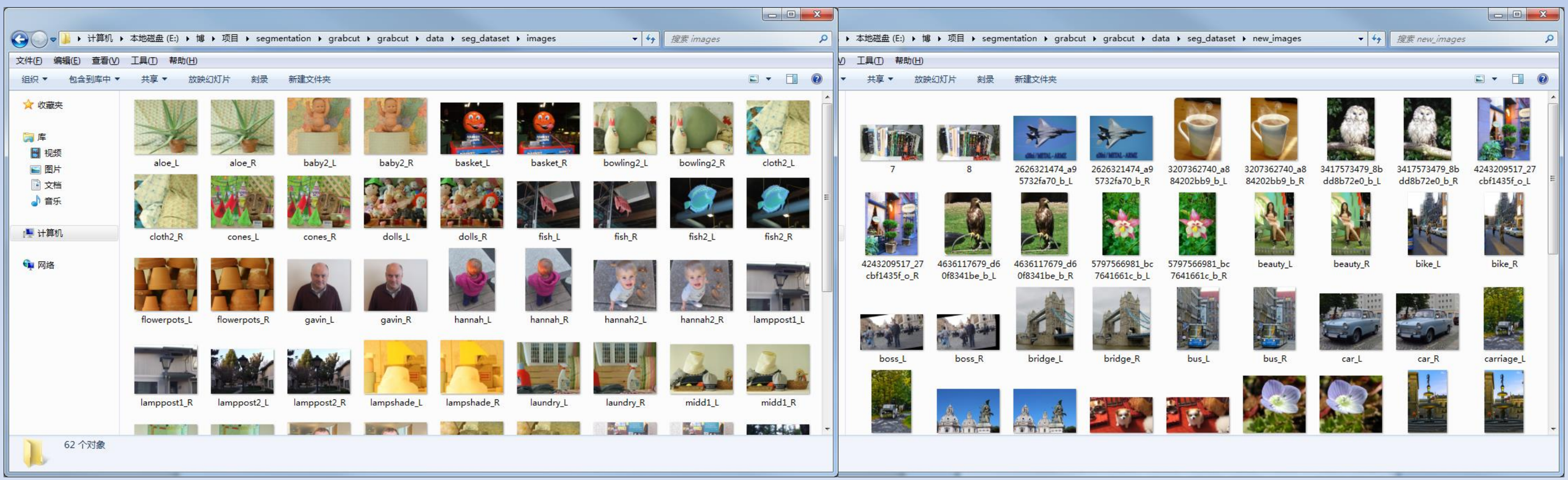
Further editing

- Users scribble on the initial result with a foreground and background brush.
- The color models, consistent graph model and optimal flow are re-computed



Evaluation

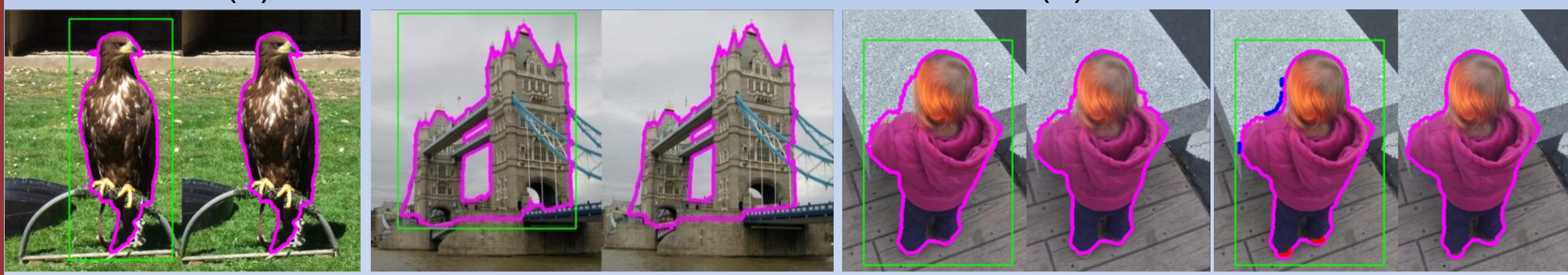
- Dataset
  - ➢ www.adobe.com/go/datasets[5]
  - ➢ www.flickr.com
  - ➢ Stereo photos taken in real world



Evaluation

Stereo GrabCut

(a) (b)



(c) (d) (e)

(f) (g)

Consistency and accuracy

- Consistency evaluation $\frac{|C_l| + |C_r|}{|N_l| + |N_r|}$

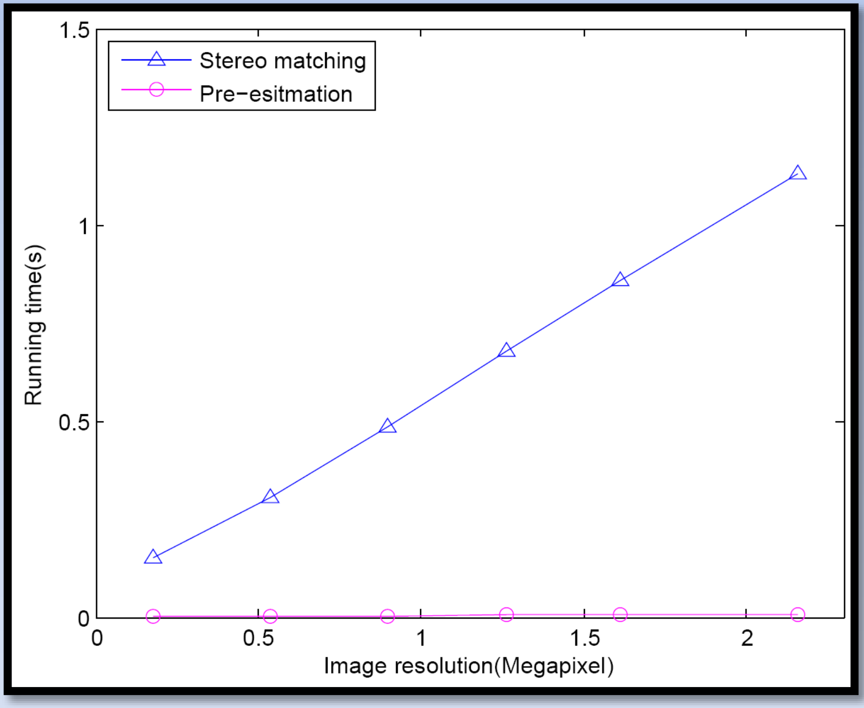
Approach	Input by machine			Input by user		
	Rect	Stroke	Rect+Stroke	Rect	Stroke	Rect+Stroke
GrabCut[5]	95.93	-	98.33	85.72	-	97.20
StereoCut[4]	-	99.13	-	-	99.12	-
Stereo GrabCut	99.44	-	99.38	99.25	-	99.27

- Accuracy evaluation $\frac{N_{gr} \cap N_{rs}}{N_{gr} \cup N_{rs}}$

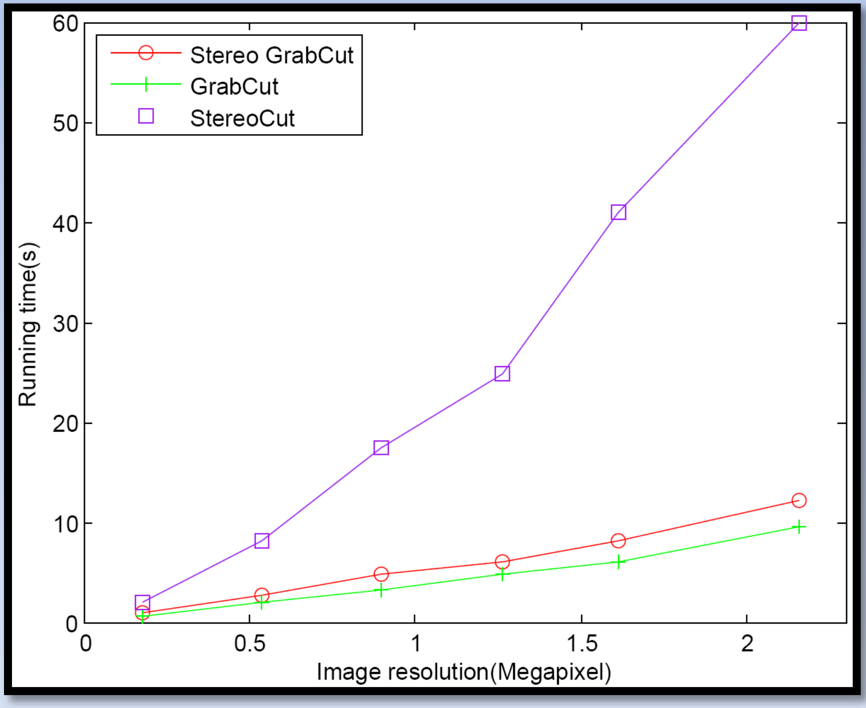
Approach	Input by machine			Input by user		
	Rect	Stroke	Rect+Stroke	Rect	Stroke	Rect+Stroke
GrabCut[5]	81.52	-	94.85	74.01	-	95.02
StereoCut[4]	-	91.49	-	-	94.44	-
Stereo GrabCut	87.36	-	96.62	85.89	-	98.16

Running time

- Test on a 2.4GHz Intel T8300 CPU with 2GB RAM
- Acceptable for user interaction



Running time of pre-processing modules

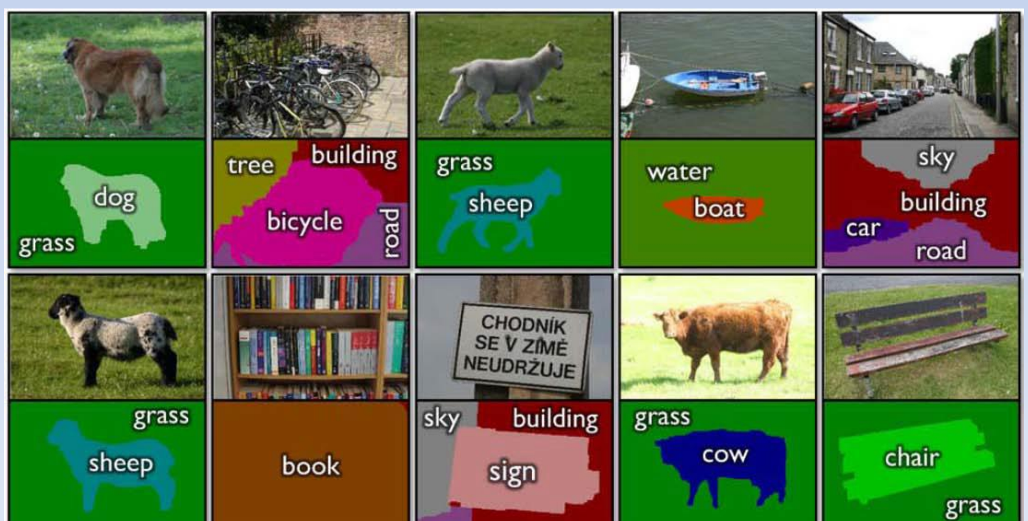


Total time comparison

Future work

- Apply Stereo GrabCut to further applications
  - Stereo image editing

  - Supervised segmentation and labeling



References

[1] Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graph cuts. In: ACM Transactions on Graphics (TOG). Volume 23., ACM (2004) 309–314

[2] Geiger, A., Roser, M., Urtasun, R.: Efficient large-scale stereo matching. In: Computer Vision–ACCV 2010. Springer (2011) 25–38

[3] Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. Volume 1., IEEE (2001) 105–112

[4] Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 23(11) (2001) 1222–1239

[5] Price, B.L., Cohen, S.: Stereocut: Consistent interactive object selection in stereo image pairs. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE (2011) 1148–1155

Thank you!

Q&A