Combining Contrast Saliency and Region Discontinuity for Precise Hand Segmentation in Projector-Camera System

Jingwen Dai *Ronald Chung*

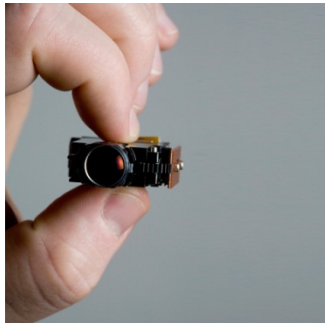
Computer Vision Laboratory

Dept. of Mech. and Automation Engineering

The Chinese University of Hong Kong

ICPR2012, Tsukuba Science City, Japan, 14 Nov 2012

Motivation



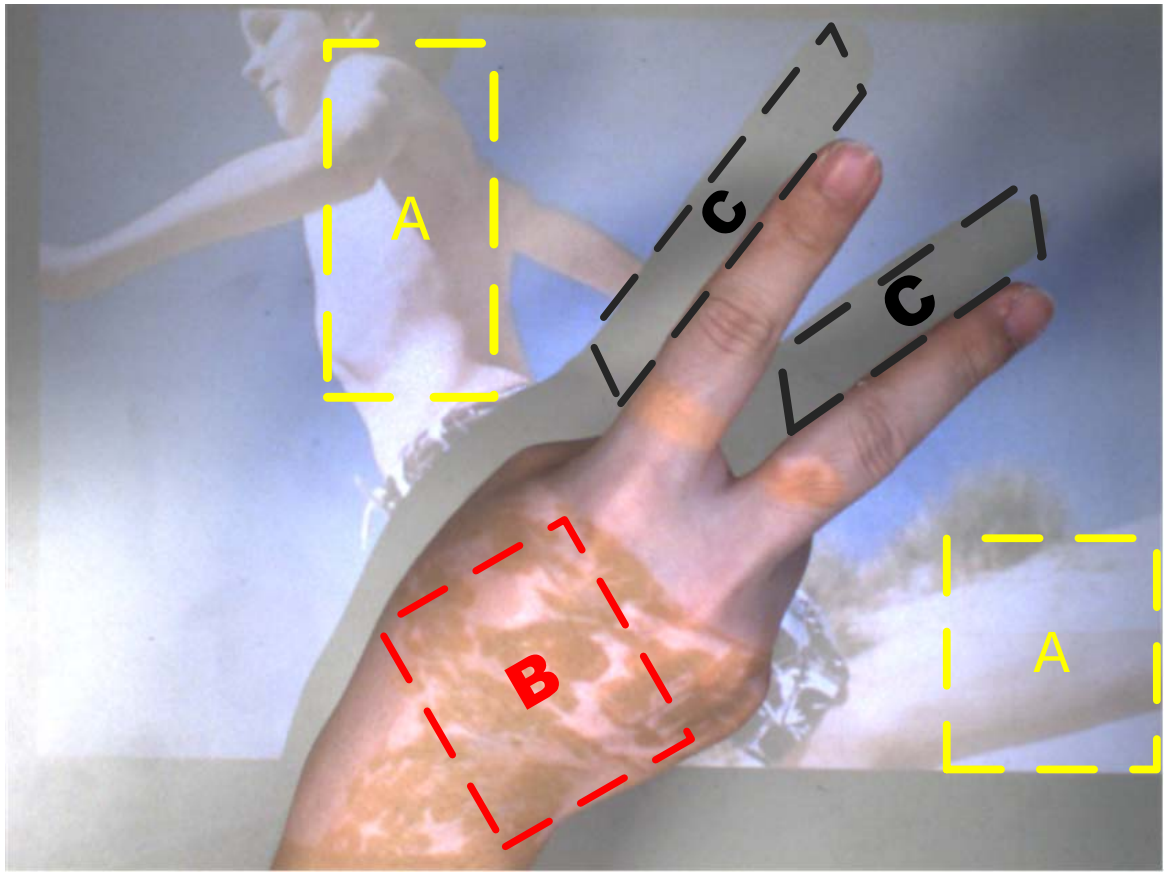
DLP Pico Projector

Mobile Phone

DC

DV

Previews Works



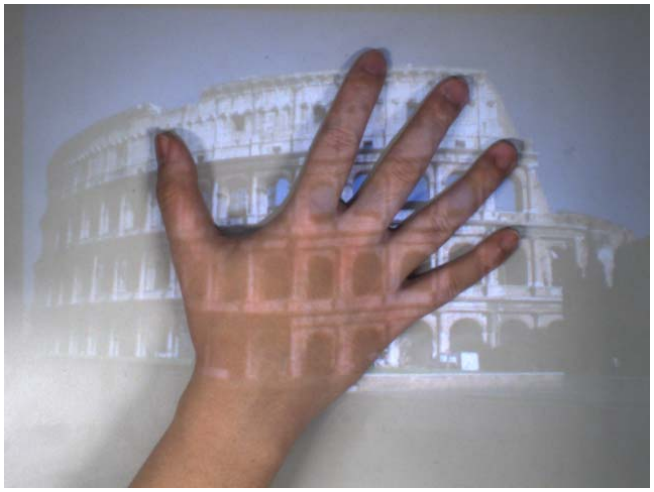
- **Skin-color Detection**
- **Background Subtraction**

Previews Works

- **Graph-based Approaches**
  - *Time Consuming*
  - *User's Interaction*
- **Additional Instruments**
  - *Infrared Camera*
  - *Stereo Camera*
  - *Depth Sensor (Time-of-Fight Camera, Kinect etc.)*
- **Our Method**

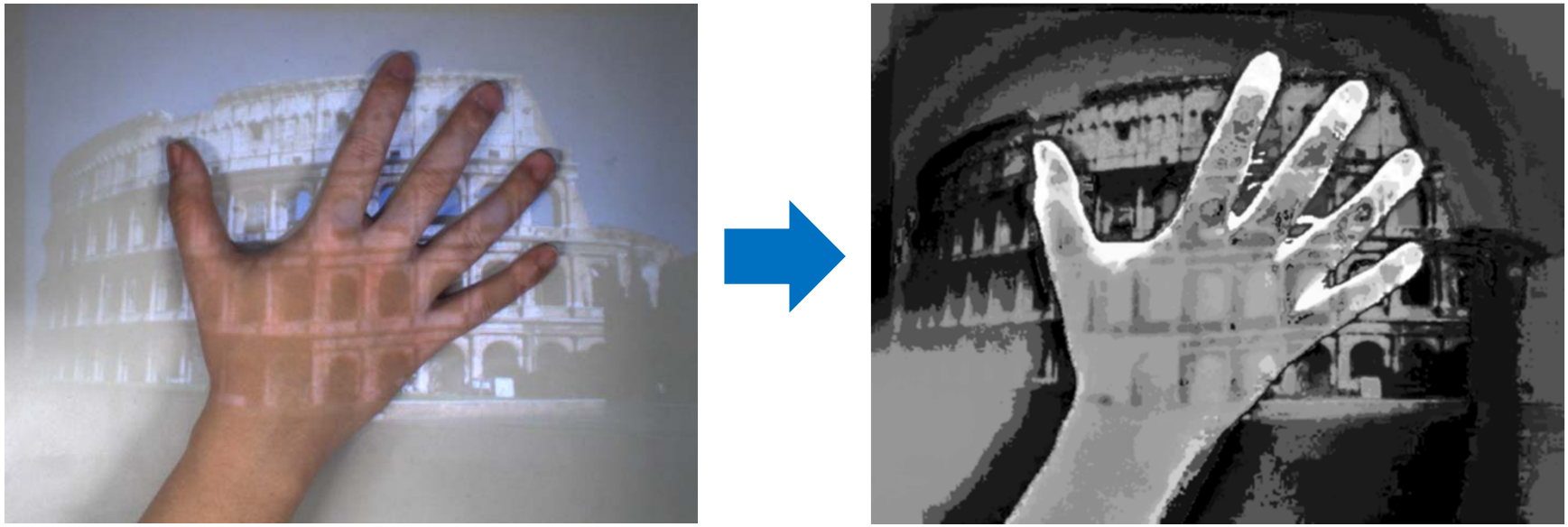
  - ***w/o pre-calibration & pre-training***

Rough Segmentation by Contrast Saliency



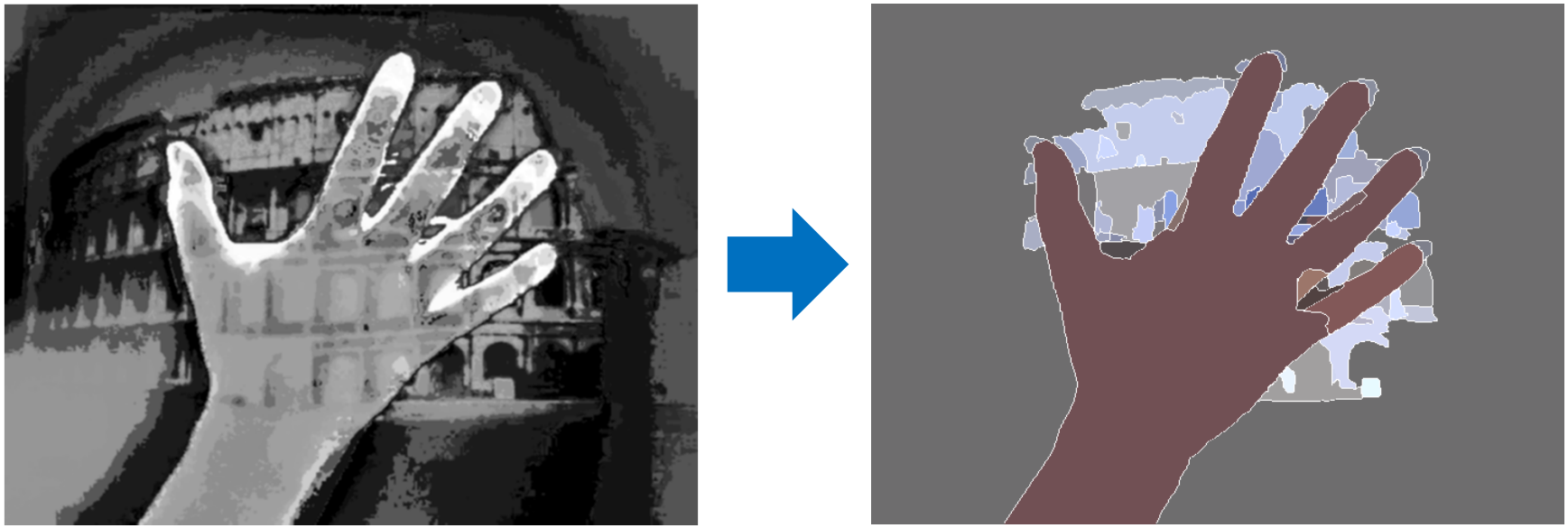
- **Hand - Most *noticeable* object**
- ***Salient* Region Detection**
- **Saliency Detector Requirements:**
  - *Emphasizing the **largest** salient objects*
  - ***Uniformly** highlighting whole salient regions*
  - ***Disregarding artifacts** arising from projection content and ambient illumination*
  - *Accomplishing detection less than **15ms***

Histogram-based Contrast Saliency



- *M. Cheng et al. Global contrast based salient region detection, CVPR2011.*

Mean-Shift Region Smoothing



Precise Segmentation by Fusing

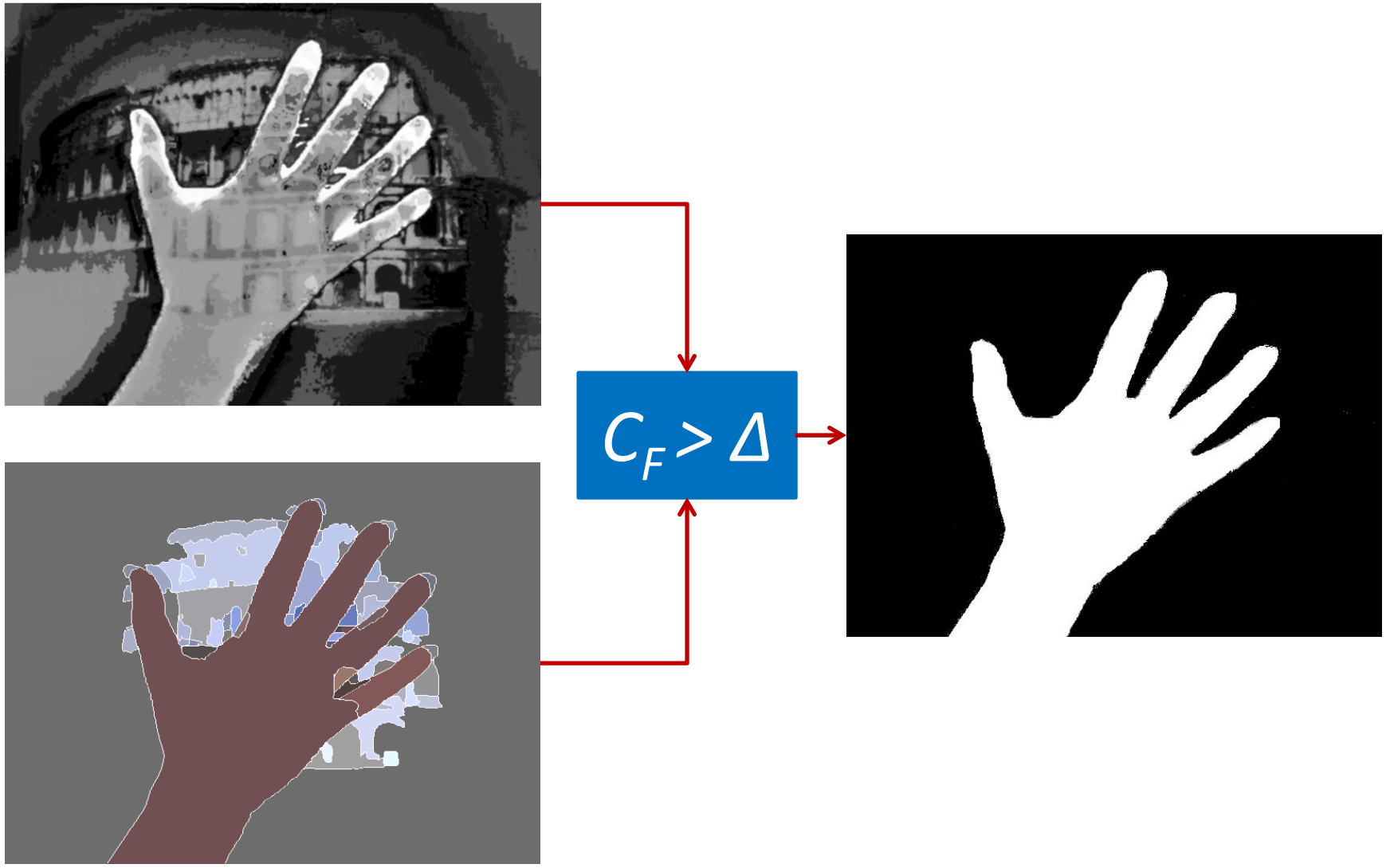
- **Confidence Function**

$$C_F(k) = \frac{1}{e^{(L-1)}}\left[\alpha \bar{S}(k) + \beta \bar{S}_N(k) + \gamma A(k)\right]$$

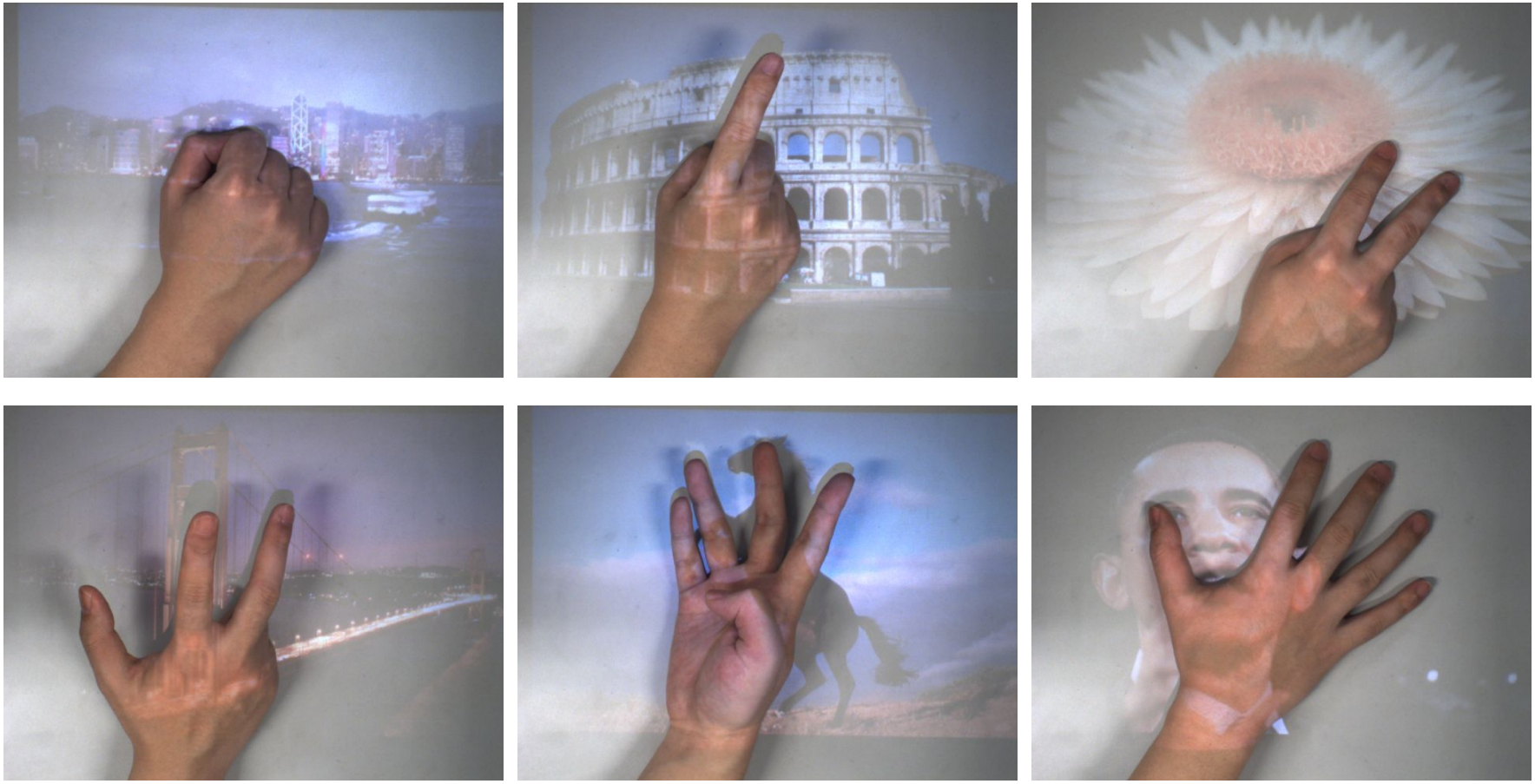
  - *The **average salient value** of the pixels in the partition*
  - *The **number** of the neighbor partitions and the **average salient value** of neighbor partitions*
  - *The **area** of the partition*
  - *Whether the partition is on the **image boundaries***

$$\begin{cases} \beta = \frac{1}{2}, \alpha = \gamma = \frac{1}{4}, & \text{when N} = 1 \\ \alpha = \frac{1}{2}, \beta = \gamma = \frac{1}{4}, & otherwise \end{cases}$$

Precise Segmentation by Fusing



Experiments – *Data Collection*

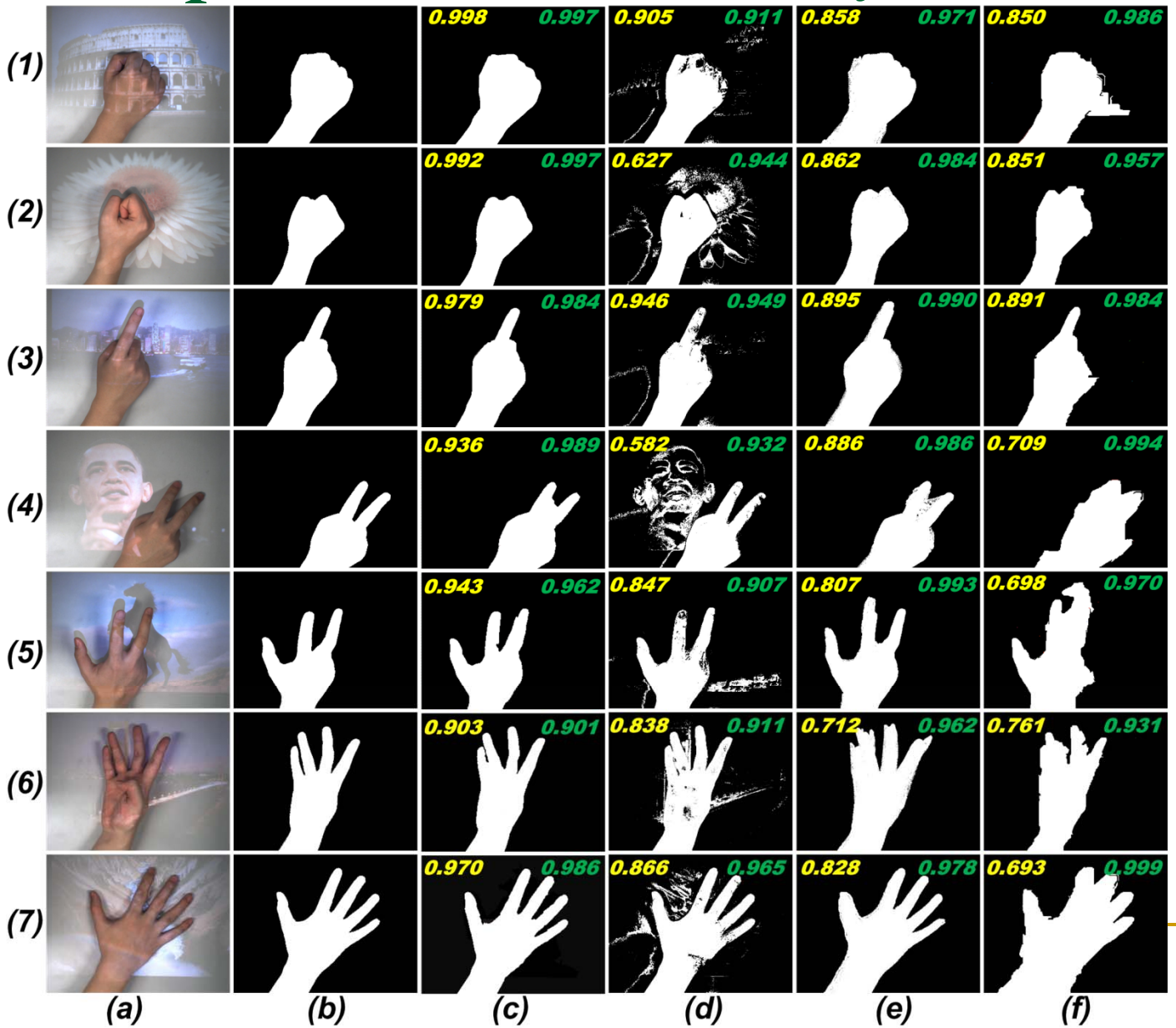


Experiments – *Accuracy Evaluation*

- **Comparison with related methods**
  - *The classic approach of **statistical color model-based** method (most cited in literature)*
  - *The **background subtraction** method (most precise)*
  - *The **graph based** method (most recent)*
- **Evaluation Criterion**
  - *Precision:* $p = \frac{N_C}{N_R}$
  - *Recall:* $r = \frac{N_C}{N_G}$
  - *F-beta score:* $F_\beta = \frac{(1+\beta^2)pr}{\beta^2 p + r}$

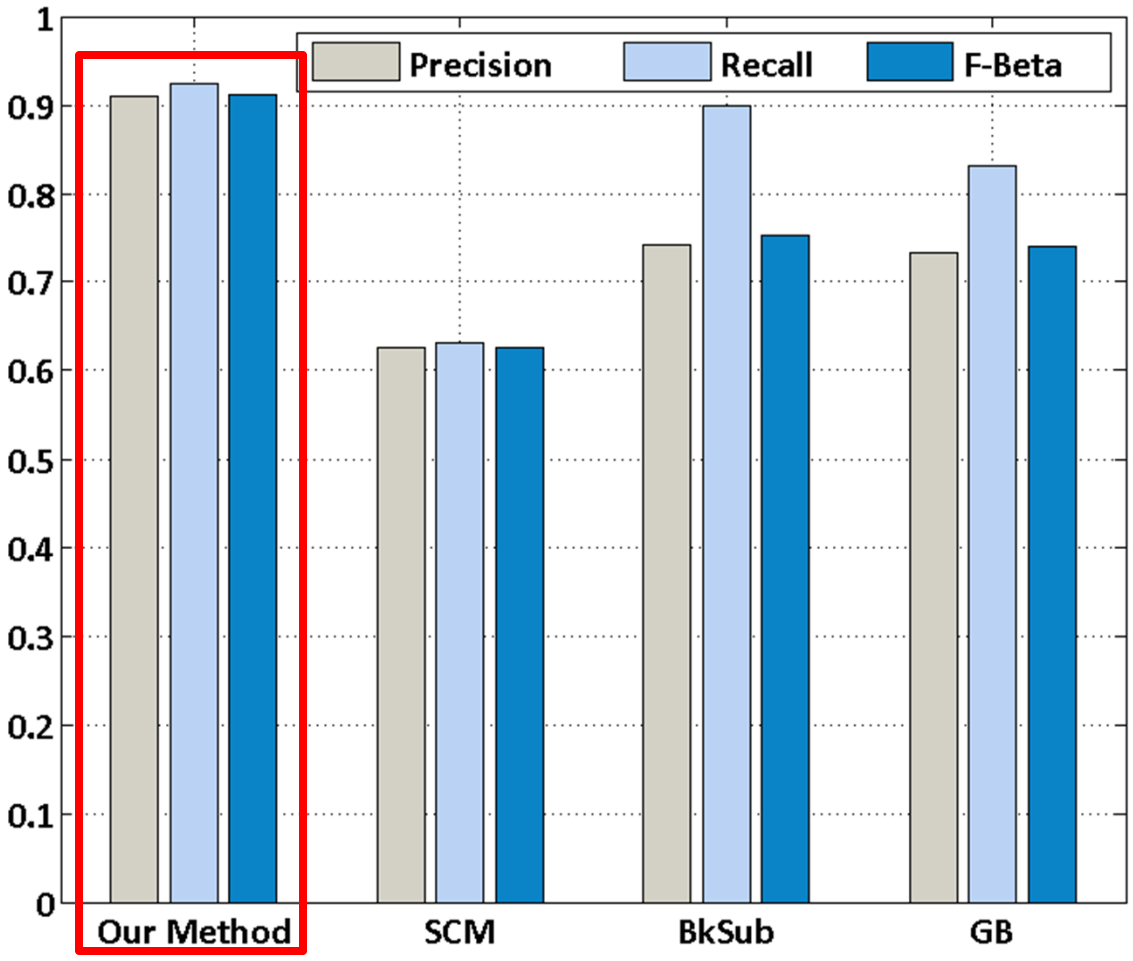
N_C, N_R, N_G *are the number of **correct segmented** pixels, **all segmented** pixels and **ground-truth** pixels respectively.*

Experiments – *Accuracy Evaluation*



(a) Original Image
(b) Ground-truth
(c) Our Method
(d) SCM
(e) BkSub
(f) GB

Experiments – *Accuracy Evaluation*



Experiments – *Efficiency Evaluation*

Method	Ours	SCM	BkSub	GB
Time (ms/frame)	*29.6*	*10.9*	*2.3*	*115.2*

Conclusion and Future Works

- **A novel *coarse-to-fine* approach for hand segmentation in projector-camera system, which puts together *contrast saliency* and *region discontinuity* information through a *confidence function*.**

- **Future Works**
  - *Improve the formulation of confidence function*
  - *Reduce algorithm complexity*

THANK YOU!!

If you have any questions, please contact

Dr. Jingwen Dai

dai@cs.unc.edu