Use of Projector-Camera System for Human-Computer Interaction

Ph.D. Oral Defense

DAI, Jingwen

Thursday, Sept. 6th 2012

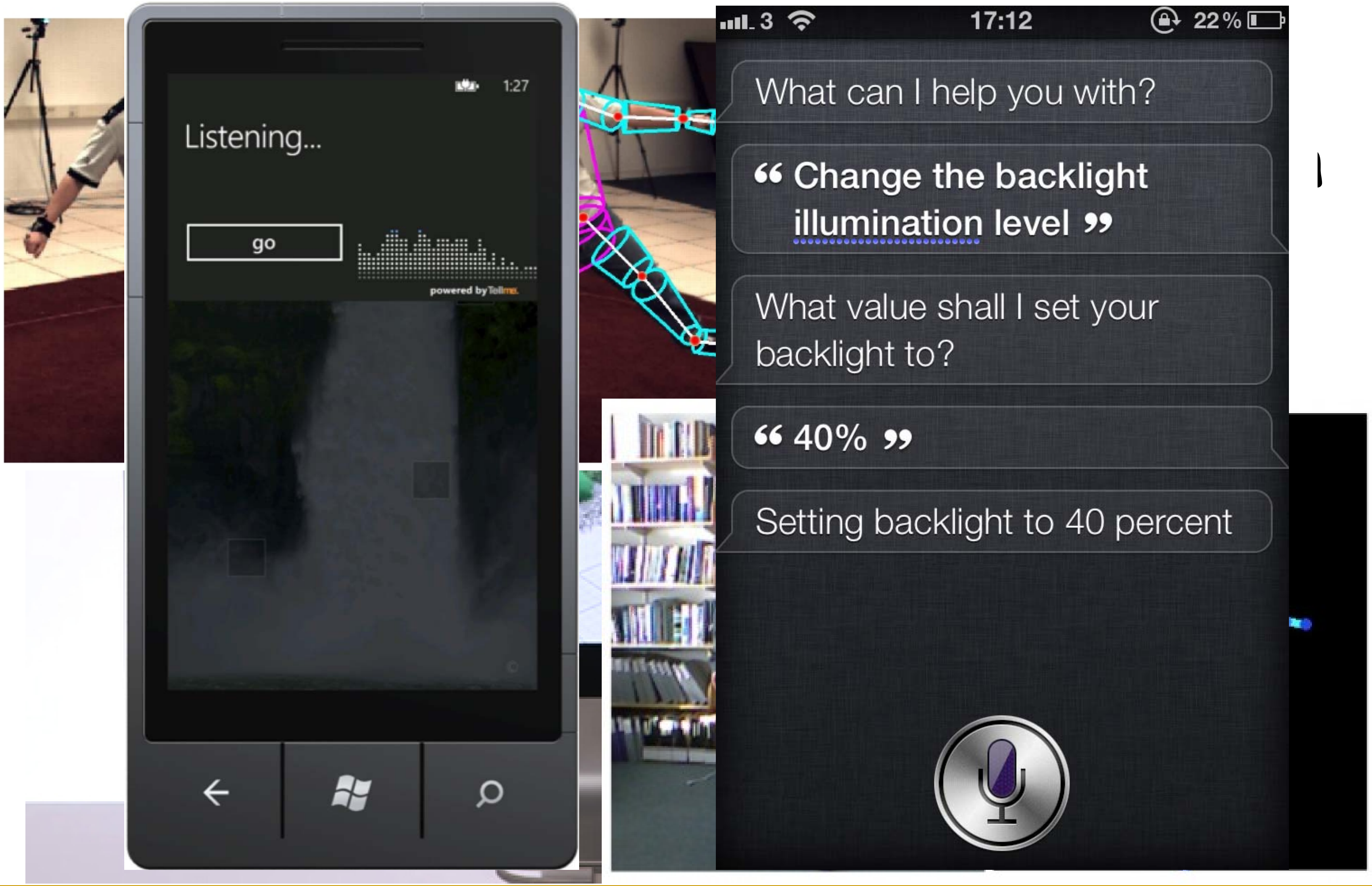
Computer Vision Lab, Mechanical and Automation Engineering,
The Chinese University of Hong Kong

Thesis Supervisor: Prof. Ronald Chung
Committee Members: Prof. Yun-hui Liu
Prof. Charlie C.L. Wang
Prof. Edmund Lam (University of Hong Kong)

Contents

- Motivation & Challenges
- Head Pose Estimation by ISL
- Embedding Invisible Codes into Regular Video Projection
- Hand Segmentation in ProCams
- Touch-Sensitive Display in Arbitrary Planar Surface
- Conclusion and Future Work

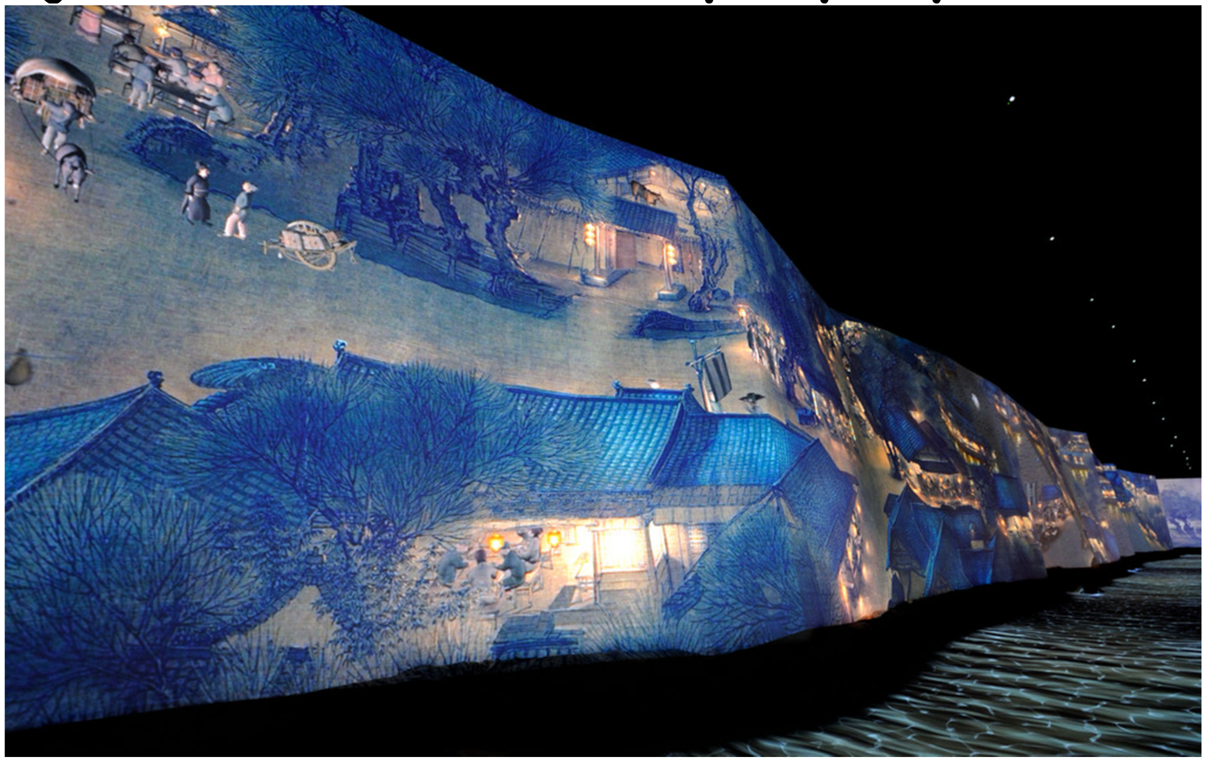
Motivation



1:27
Listening...
go
17:12
22 %
What can I help you with?
“ Change the backlight illumination level ”
What value shall I set your backlight to?
“ 40% ”
Setting backlight to 40 percent

Projector-Camera System

- **Display**
  - Projector-based display systems



[Shanghai Expo 2010]

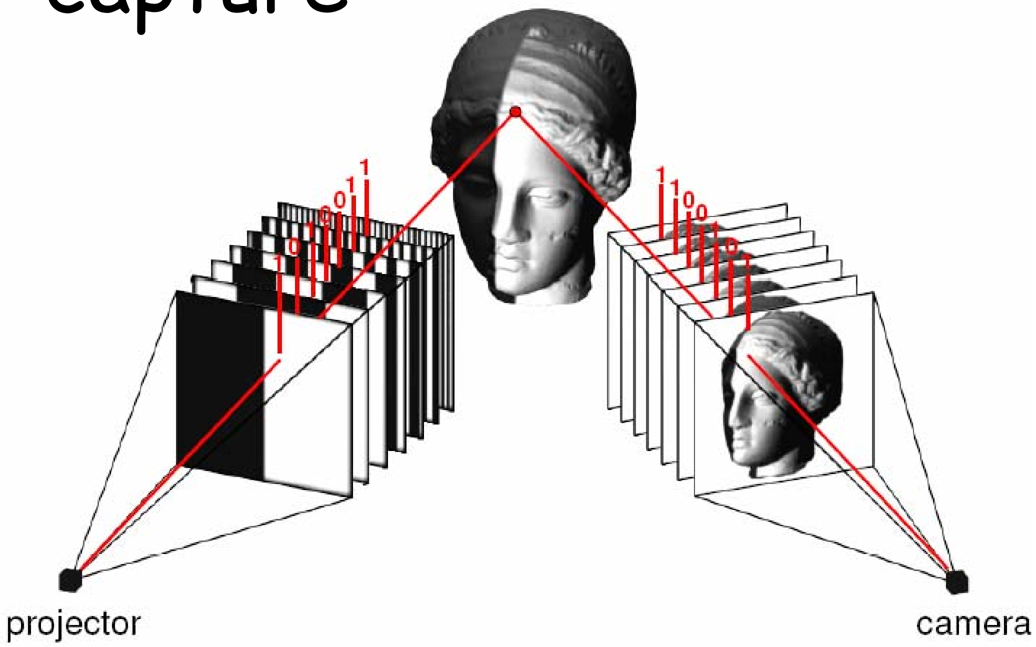
Projector-Camera System

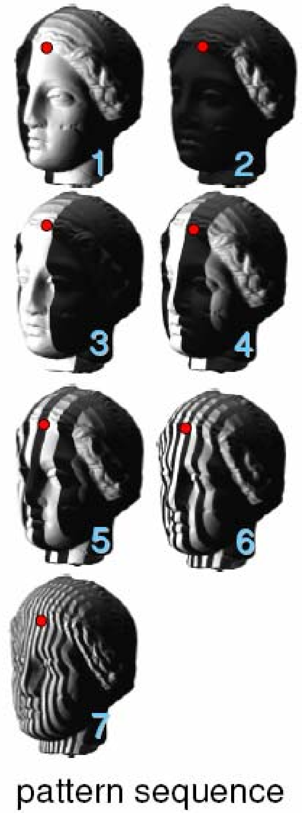
- **Display**
  - Geometric and radiometric calibration

[Fujii2005]

Projector-Camera System

- **Scene Capture**
  - Structured-light systems for scene capture

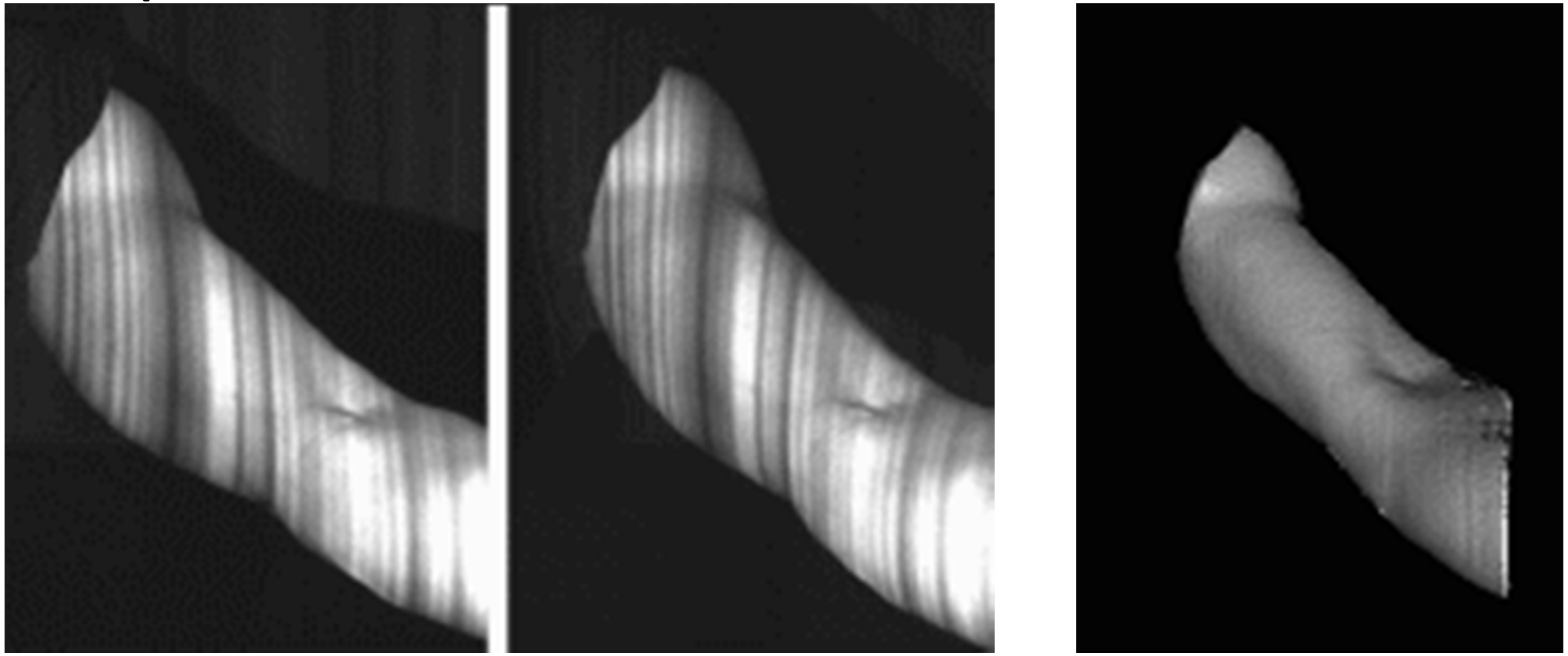




[Salvi2010]

Projector-Camera System

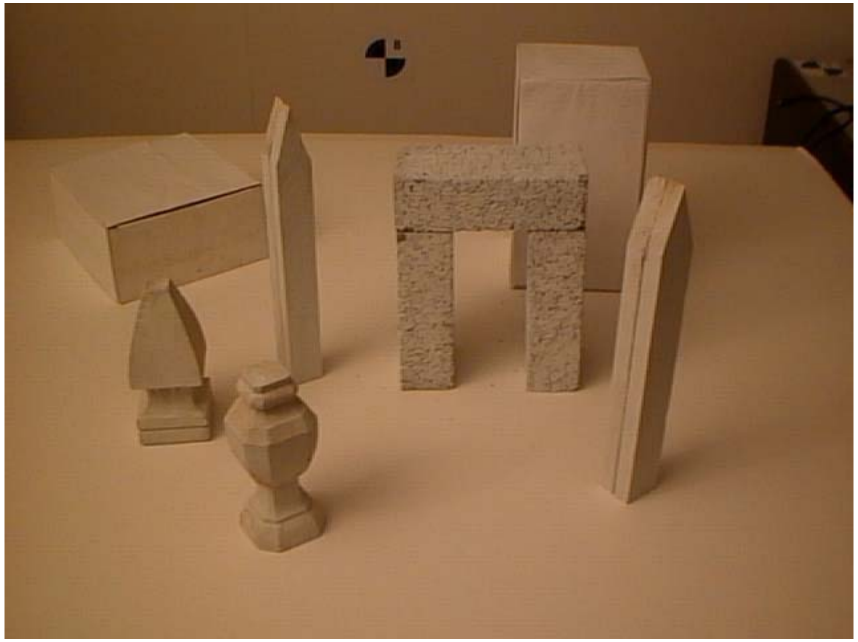
- **Scene Capture**
  - 3D geometry, reflectance, and motion capture

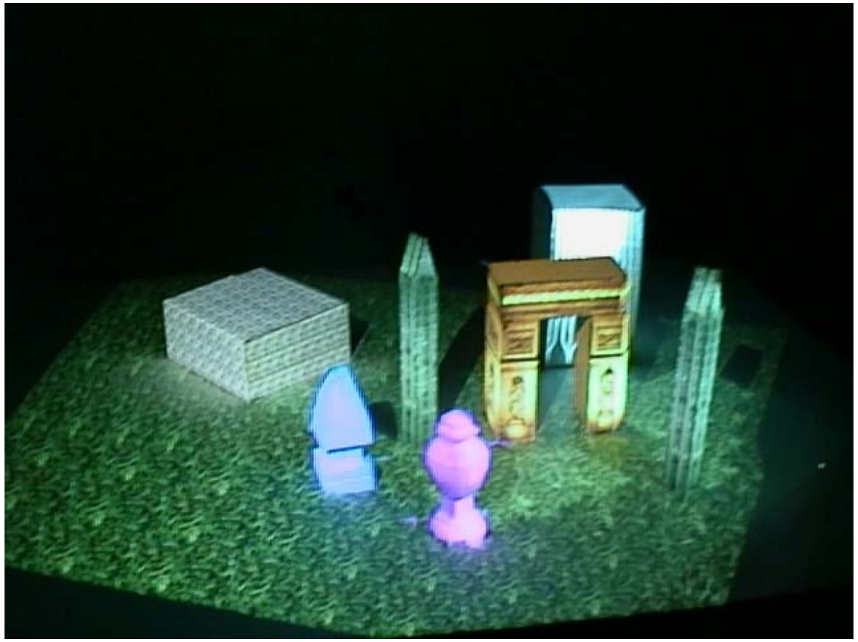


[Zhang2003]

Projector-Camera System

- **Augmentation**
  - Spatially augmented reality

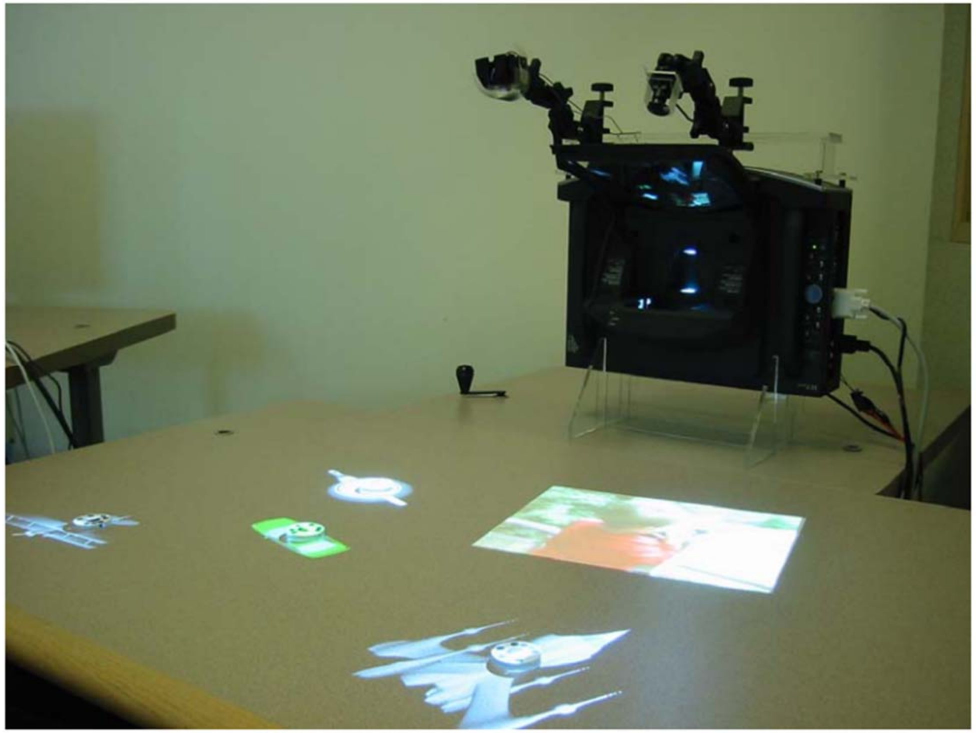




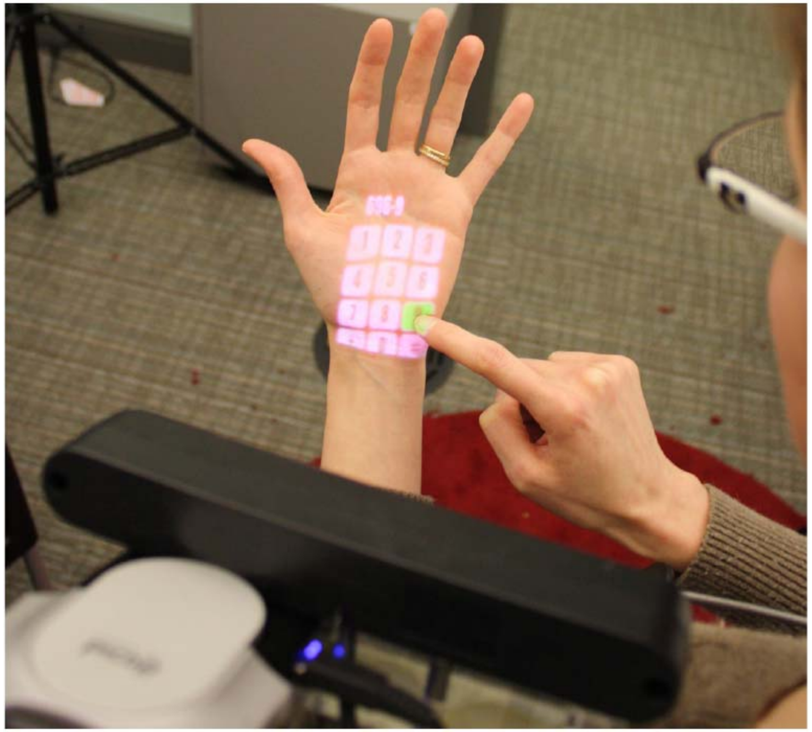
[Raskar1999]

Projector-Camera System

- **Interaction**
  - Novel interfaces



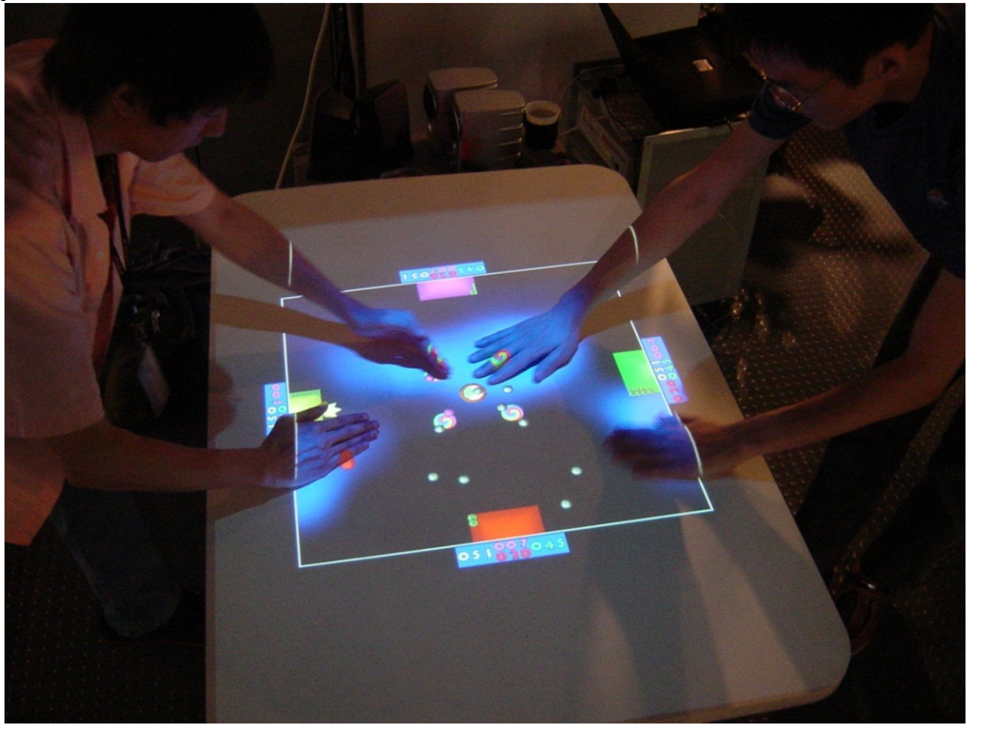
[PlayAnywhere2005]



[OmniTouch2011]

Projector-Camera System

- **Interaction**
  - Multiple user interaction



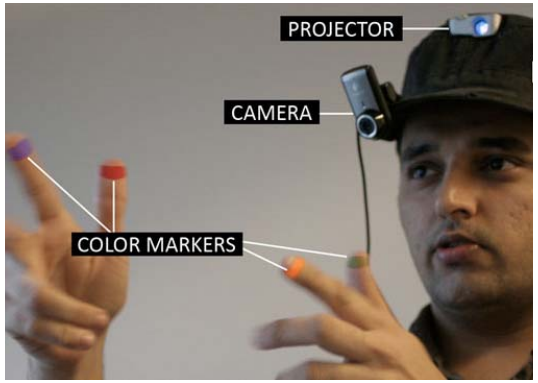
[SmartSkin2002]

Projector-Camera System

- **Mobility**

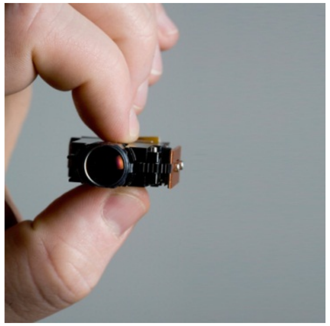
[SidebySide2011]

[Cao2007]



[SixthSense2009]

Pico Projector in Mobile Devices



Pico Projector

Mobile Phone

DC

DV

PROCAMS

Structured Light Sensing

3D Point Clouds

Task in 3D

HCI

Infer User Action

Challenges

- Simultaneous Display and Acquisition (Ch 4)
- 3D Information Interpretation (Ch2, Ch6)
- Segmentation (Ch5)
- Posture Recognition (Ch6)

Contents

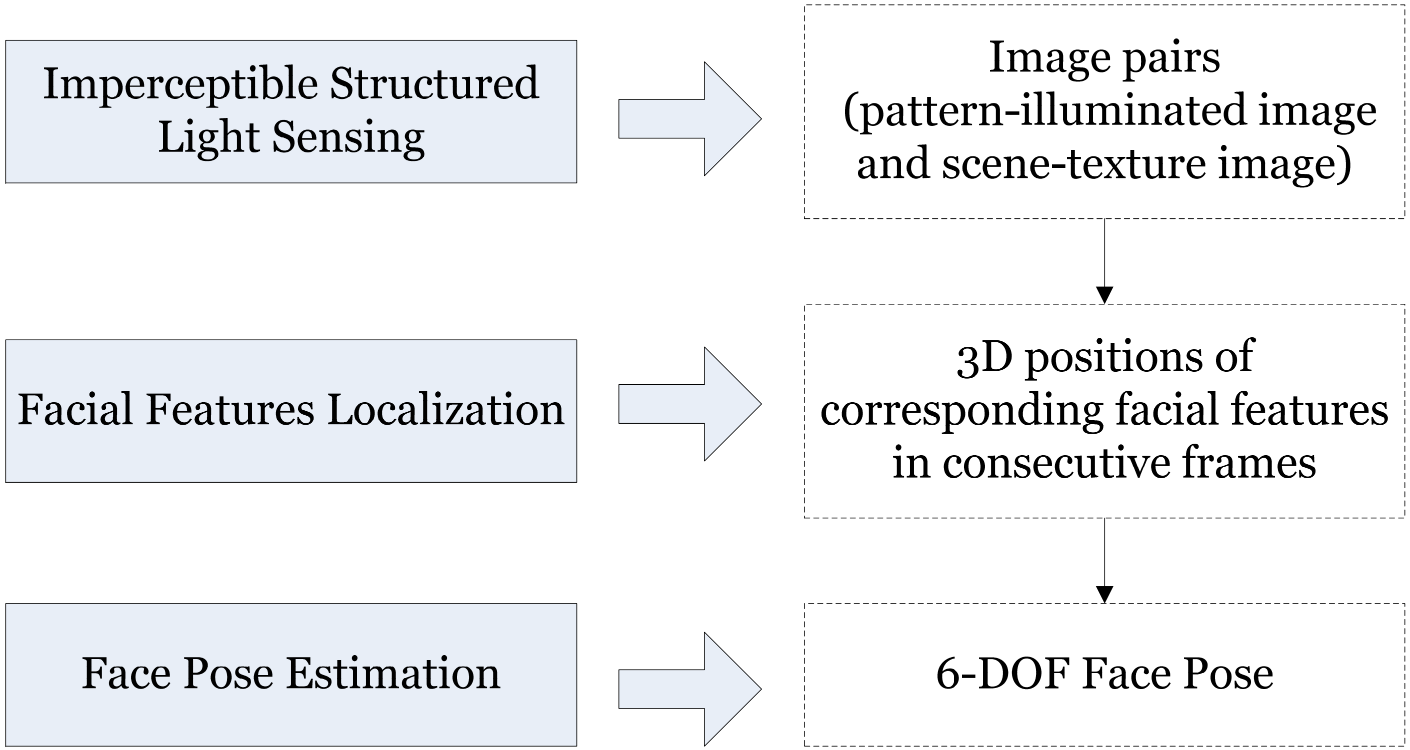
Previous Work

Image- or Video-Based

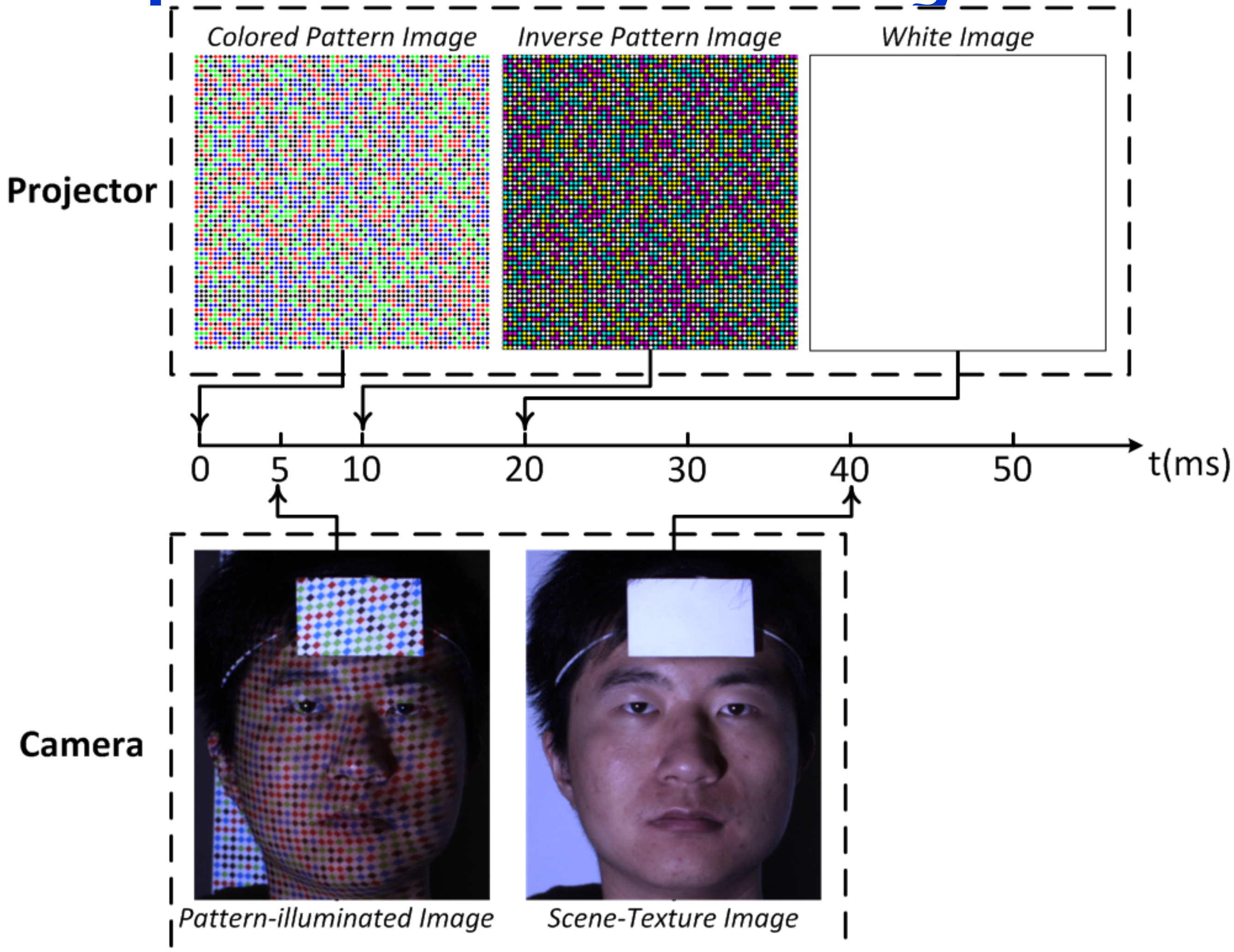
- **Head Pose Estimation**
  - Appearance Template Matching
  - Detector Array
  - Geometric Methods
  - Flexible Models
  - Nonlinear Regression Methods
  - Tracking Methods
  - Hybrid Methods

Head pose is in **3D domain**, so the use of **3D information** is more direct and accurate for pose estimation.

Overview

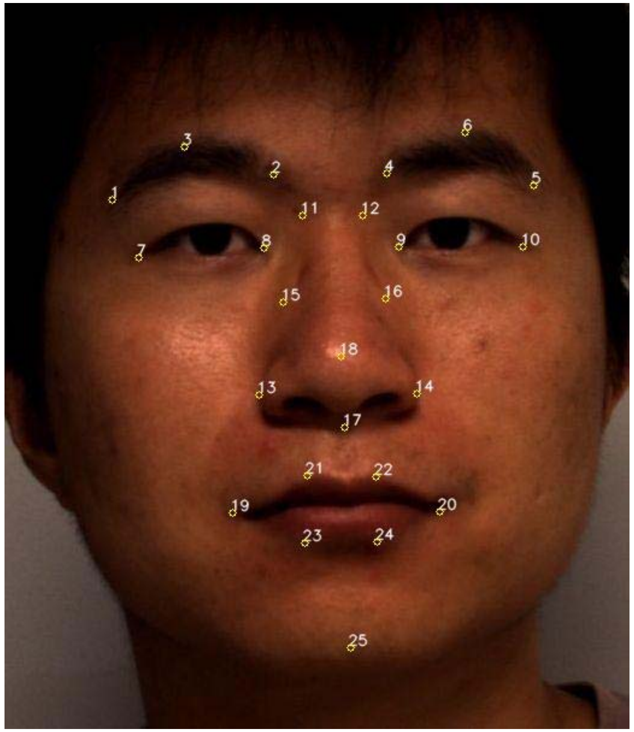

Imperceptible Structured Light Sensing
Image pairs (pattern-illuminated image and scene-texture image)
Facial Features Localization
3D positions of corresponding facial features in consecutive frames
Face Pose Estimation
6-DOF Face Pose

Pattern Projection Strategy for Imperceptible Structured Light Sensing



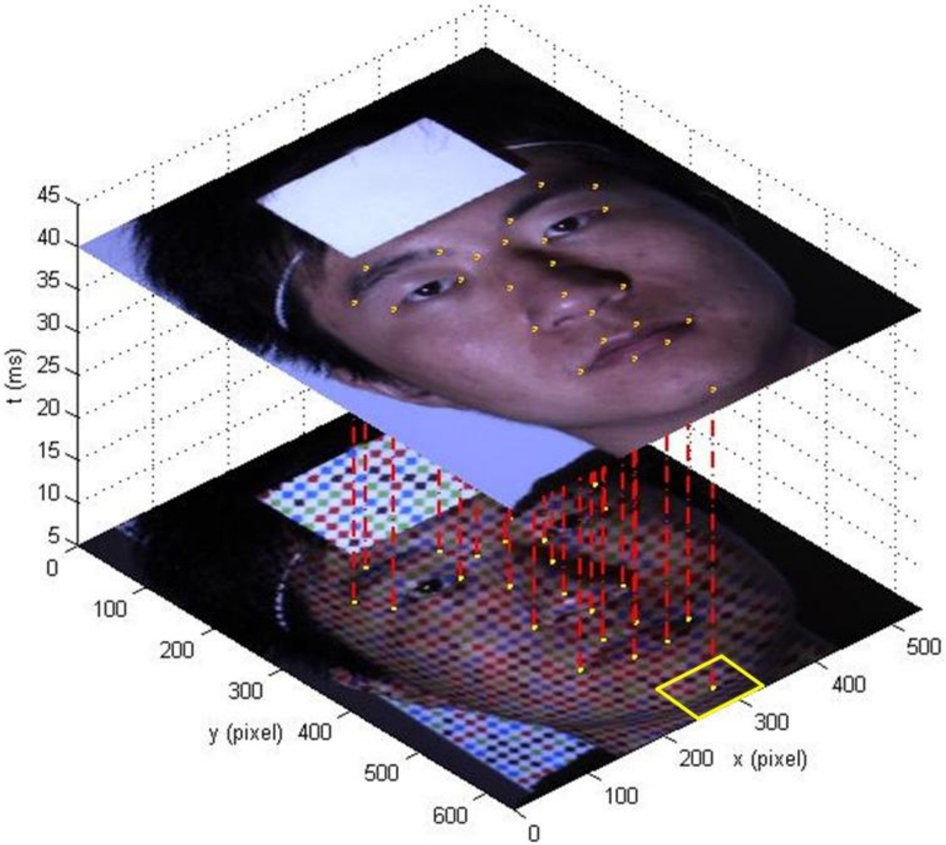
Facial Feature Localization

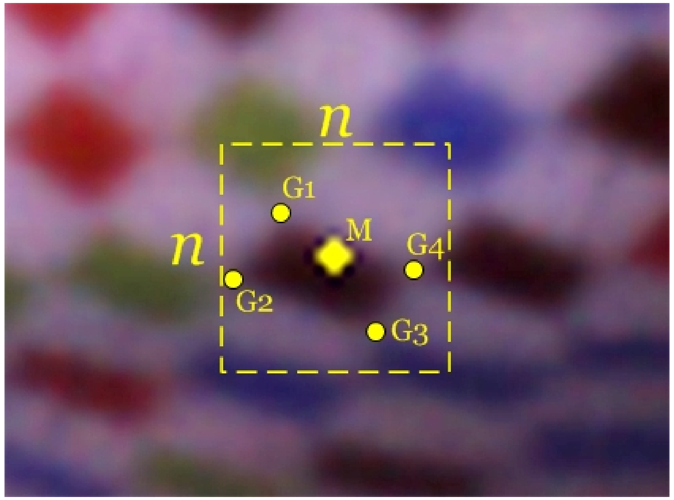
- In Scene-texture Image, localize **2D Positions of Key Facial Feature Points;**
  - Adaboost for Face Detection
  - AAM (Active Appearance Model)
    - 25 points



- In Pattern-illuminated Image, determine **3D Positions of Grid Points;**
  - Traditional structured-light approach

Inferring 3D Positions of Key Facial Features





$$\overline{X} = \sum_{i=1}^{N} \alpha_i X_i,$$

$$\alpha_i = \frac{d_i}{\sum_{j=1}^{N} d_j}.$$

6-DOF Head Pose Estimation

Singular Value Decomposition (SVD) of a correlation matrix composed by corresponding point pairs.

$$H = \sum_{i=1}^{N} p'_{c_i} {p_{c_i}}^T = U \Lambda V^T$$

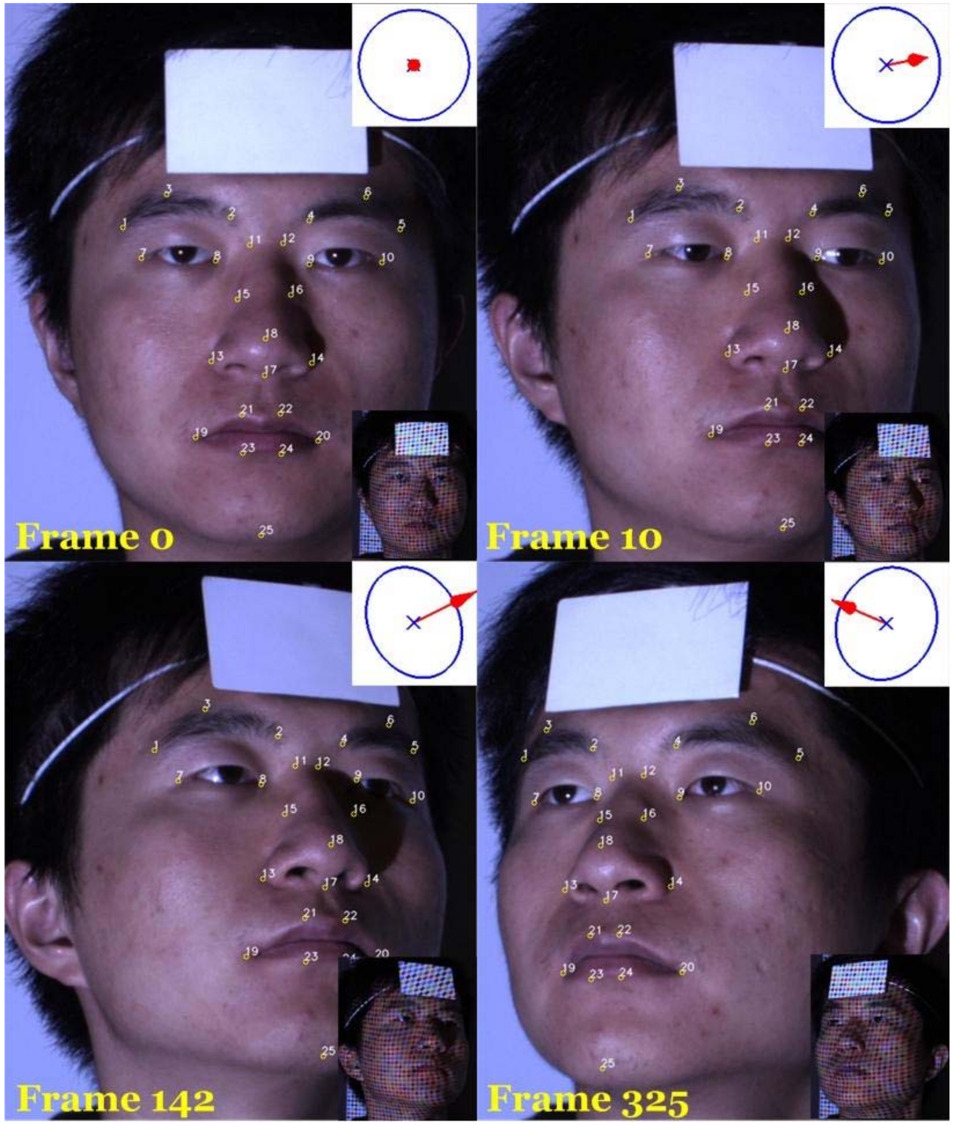
$$p_{c_i} = p_i - \bar{p},\ p'_{c_i} = p'_i - \bar{p}'$$

Rotation $\hat{R} = UV^T,$

Translation $\hat{T} = \bar{p}' - \hat{R}\bar{p}.$

If more than **three non-colinear** corresponding point pairs are known, R & T are determined uniquely.

Experiment Results



Mean Absolute Error:

Yaw – 2.02 degree

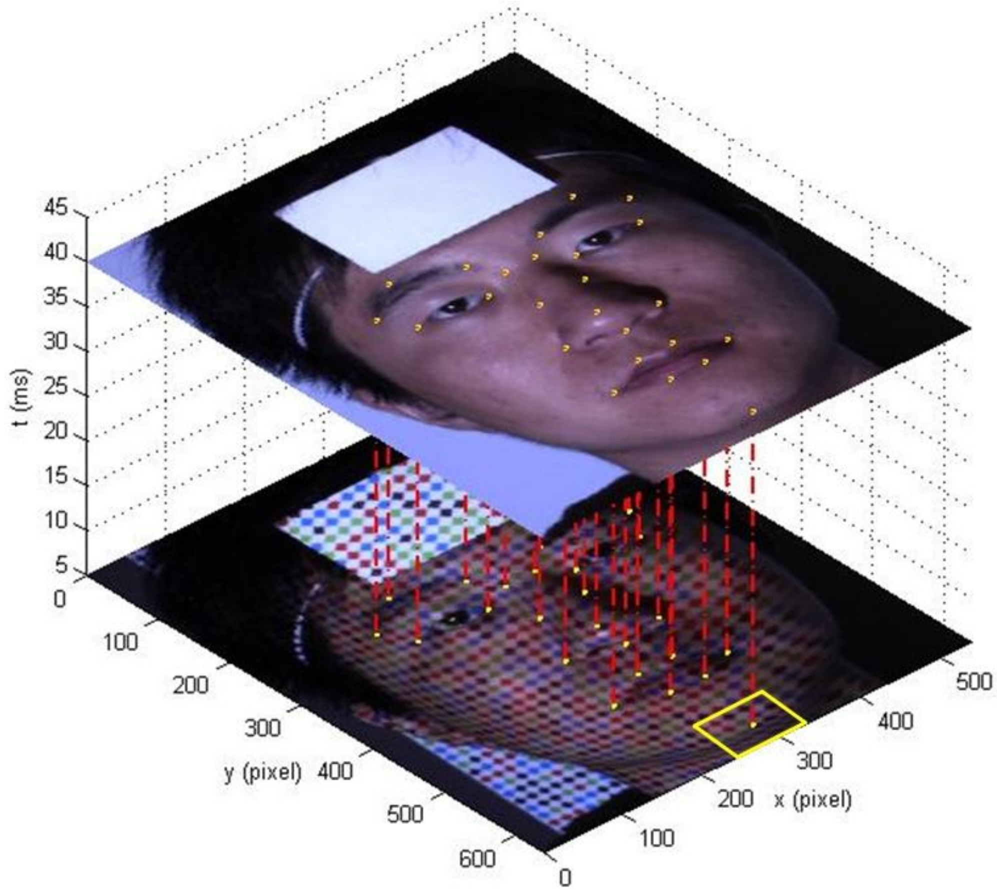
Pitch – 1.18 degree

Roll – 0.75 degree

Similar to systems using **stereo** methods.

Summary

- Pattern-illuminated Image & Scene-texture Image
- 2D Facial Feature Points & 3D Grid Points



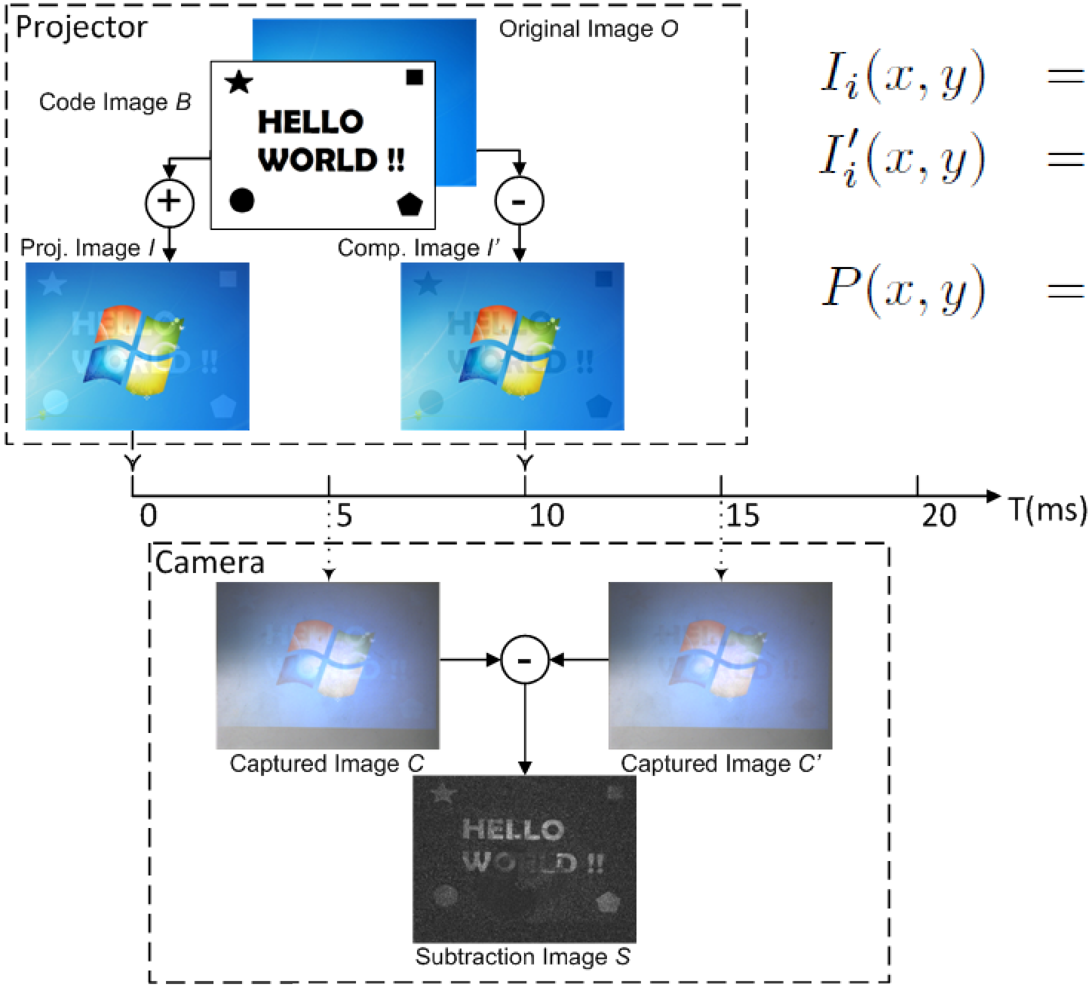
Contents

Previews Works

- **Non-Visible Spectrum (Infrared)**
  - IR Projector + IR Camera (Kinect)
  - Regular Projector and Camera + IR Filters
- **Imperceptible Structured Light (ISL)**
  - [Raskar1998] - fist proof of ISL
  - [Cotting2004] - micro-mirror states in DLP
  - [Park2007] – intensity adaption in YIQ color space
  - [Grundhofer2007] - human contrast sensitivity function
  - [Park2010] - subjective evaluation for ISL

To the best of our knowledge, few works focus on the **decoding method** in imperceptible code embedding configuration.

Principle of Embedding Imperceptible Codes

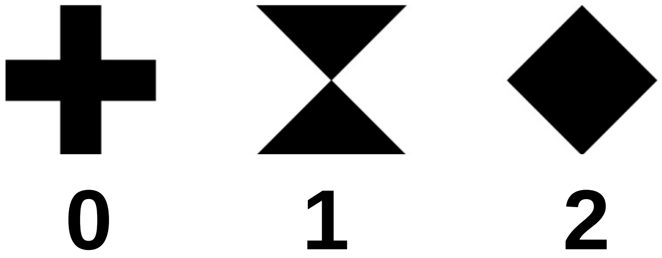


$$
\begin{aligned}
I_i(x,y) &= O_i(x,y) + P(x,y), \\
I'_i(x,y) &= O_i(x,y) - P(x,y), \\
P(x,y) &= \begin{cases} \Delta, & when \quad B(x,y) = 1; \\ 0, & when \quad B(x,y) = 0. \end{cases}
\end{aligned}
$$

$$S(x,y) = \max_i [C_i(x,y) - C'_i(x,y)], i = \{R, G, B\}.$$

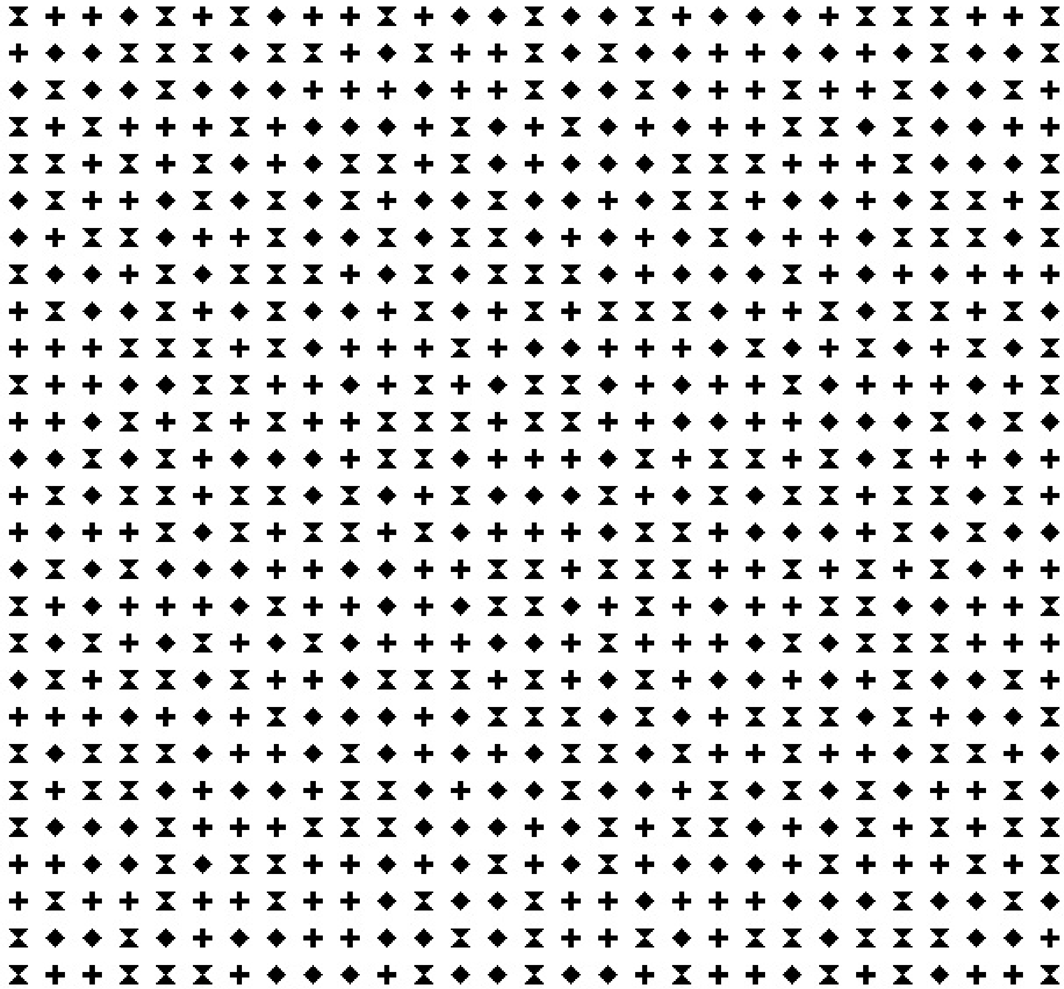
Design of Embedded Pattern

- **Primitive Shapes**
  - Cross
  - Sandglass
  - Rhombus



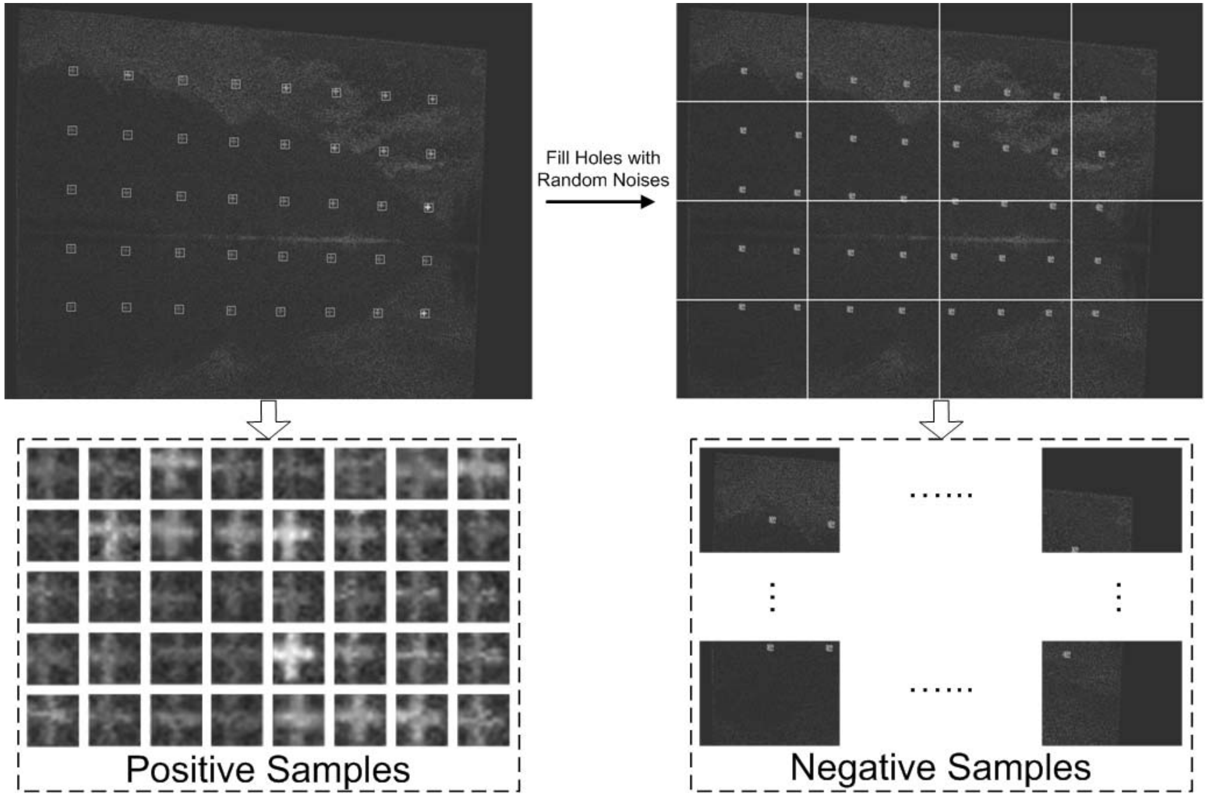
Design of Embedded Pattern

- **Pattern Image**
  - Size: 27 * 29 = 783
  - $\overline{H} = 6.0084$
  - $95.97\% \ (H \geq 3)$

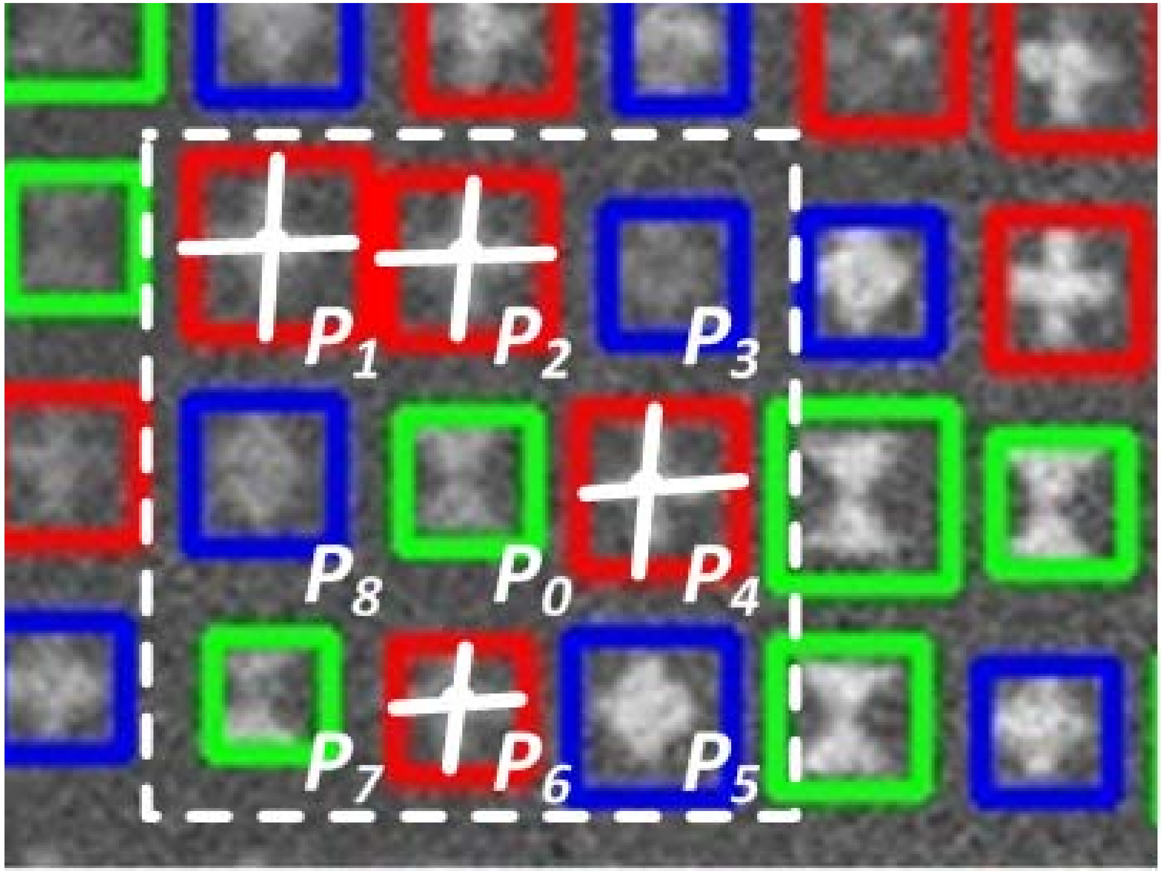


Primitive Shape Identification and Decoding

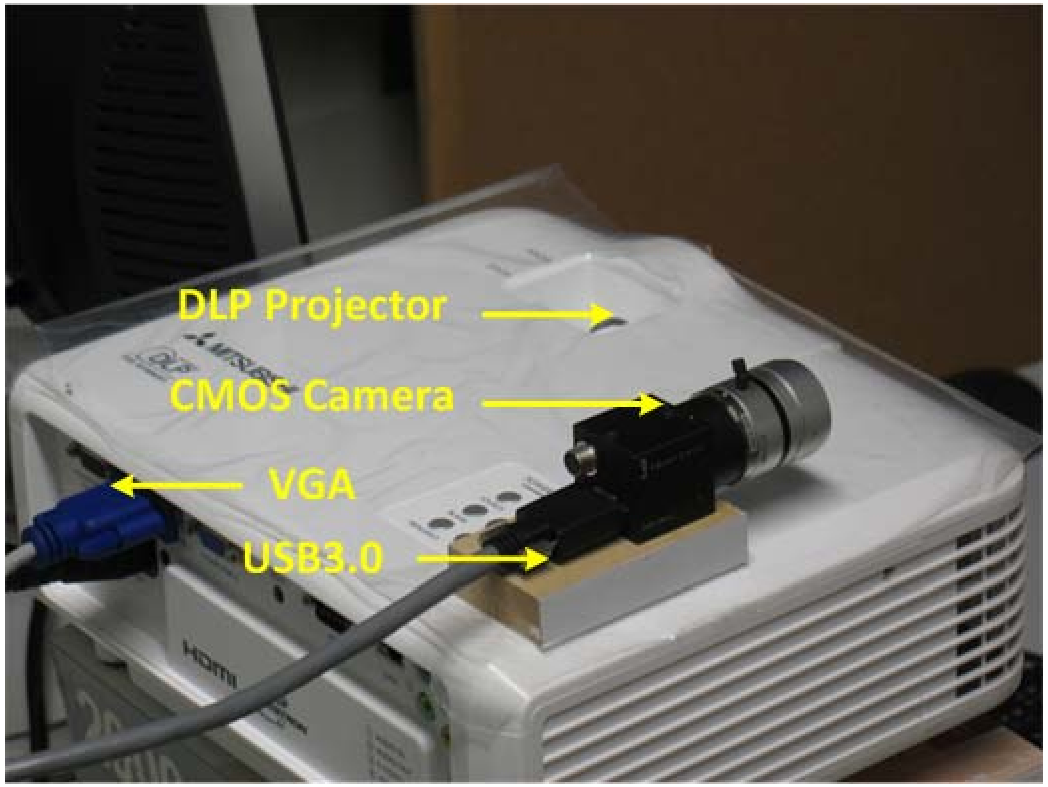
- **Adaboost Training**
  - Harr-Like Features
  - Positive Sample Size
    20 * 20
  - Pos./ Neg. Sample Num.
    7000 / 3000
  - 16-stage cascade classifier



Codeword Retrieval



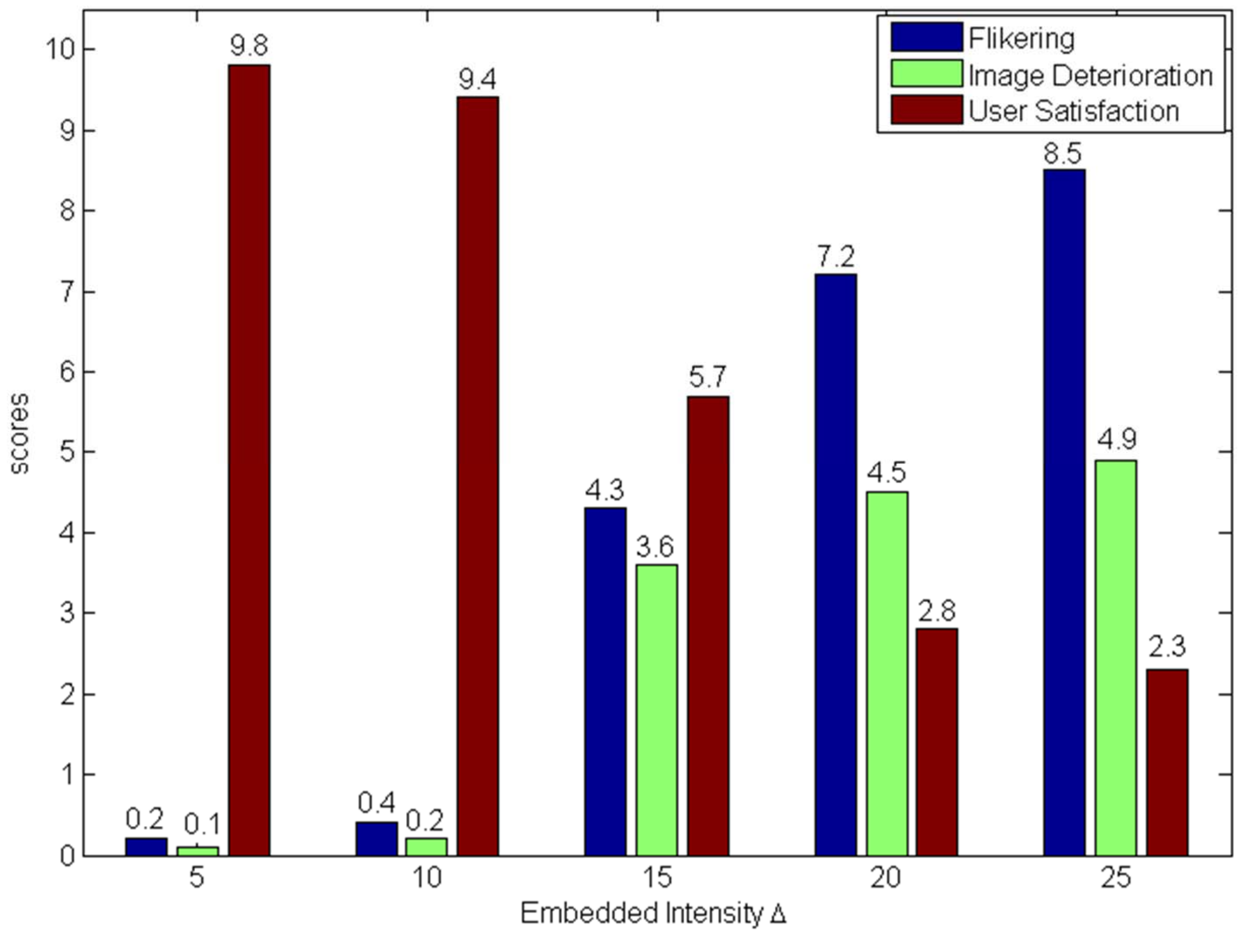
Experiments – System Setup



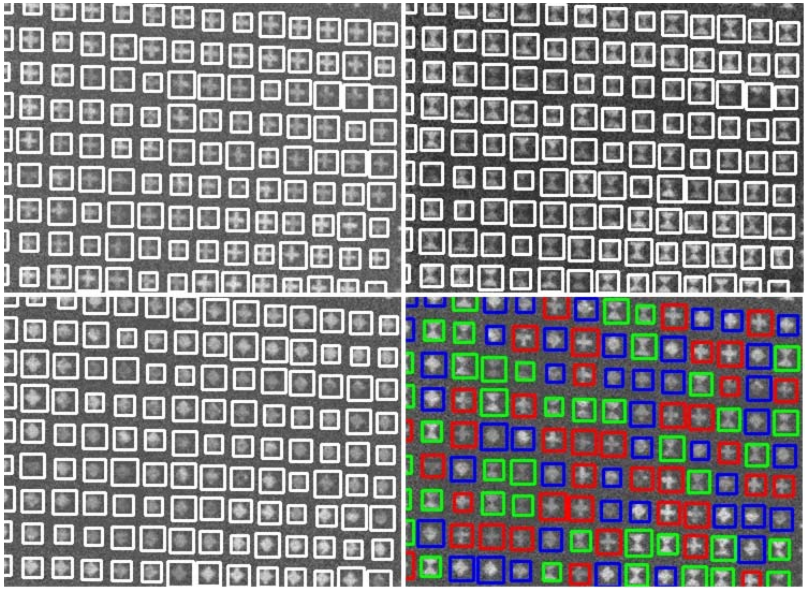
PROCAMS-A

PROCAMS-B

Experiments – Imperceptibility Evaluation

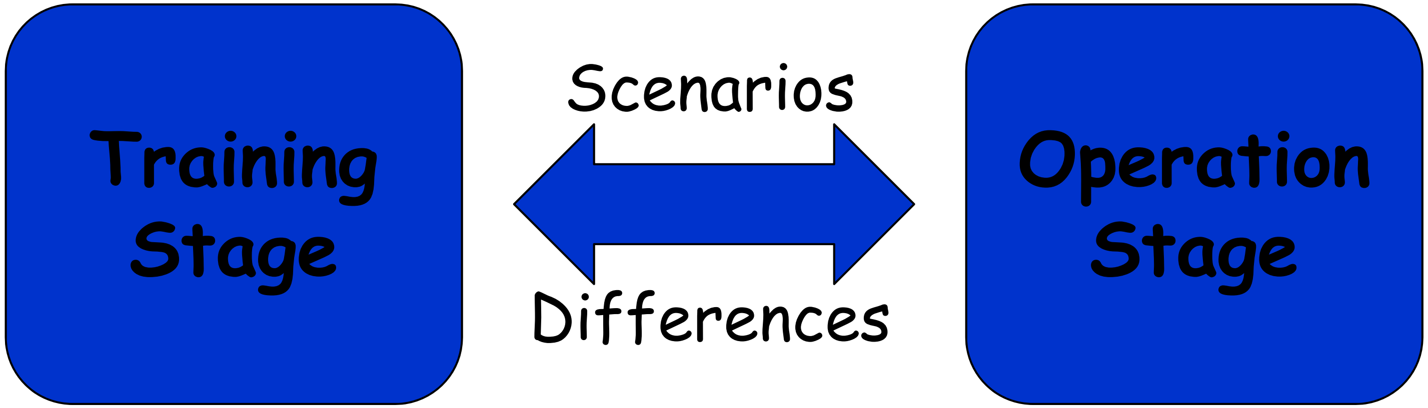


Experiments - Primitive Shape Detection Accuracy Evaluation



	H(%)	M(%)	F(%)	E_d(pixel)	Corr. Acc.(%)
Cross	94.53	3.95	1.52	1.632	—
Rhombus	95.21	3.59	1.20	1.833	—
Sandglass	95.50	3.63	0.87	1.542	—
Whole Pattern	92.11	11.06	5.28	2.013	95.74

Sensitivity Evaluation

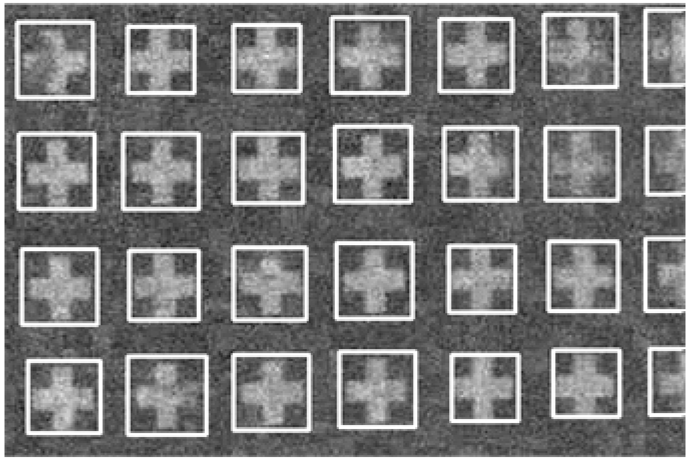


- Sensor-Object Localization
- Projection Surfaces
- Surrounding Environment
- Hardware Platforms

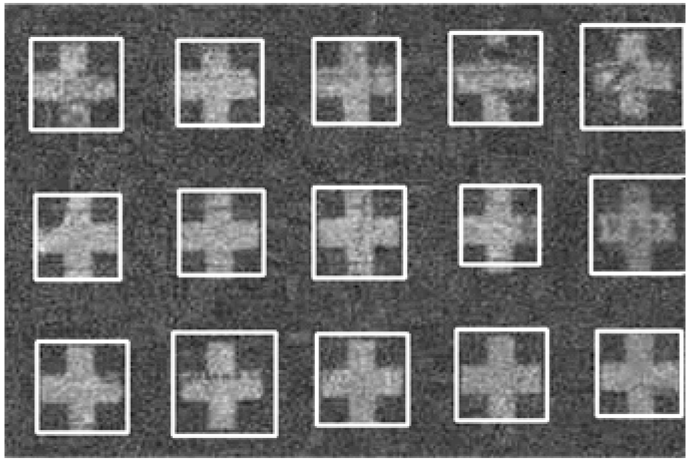
Sensitivity Evaluation: Working Distance

In training data collection:
Working distance: 800mm

500mm



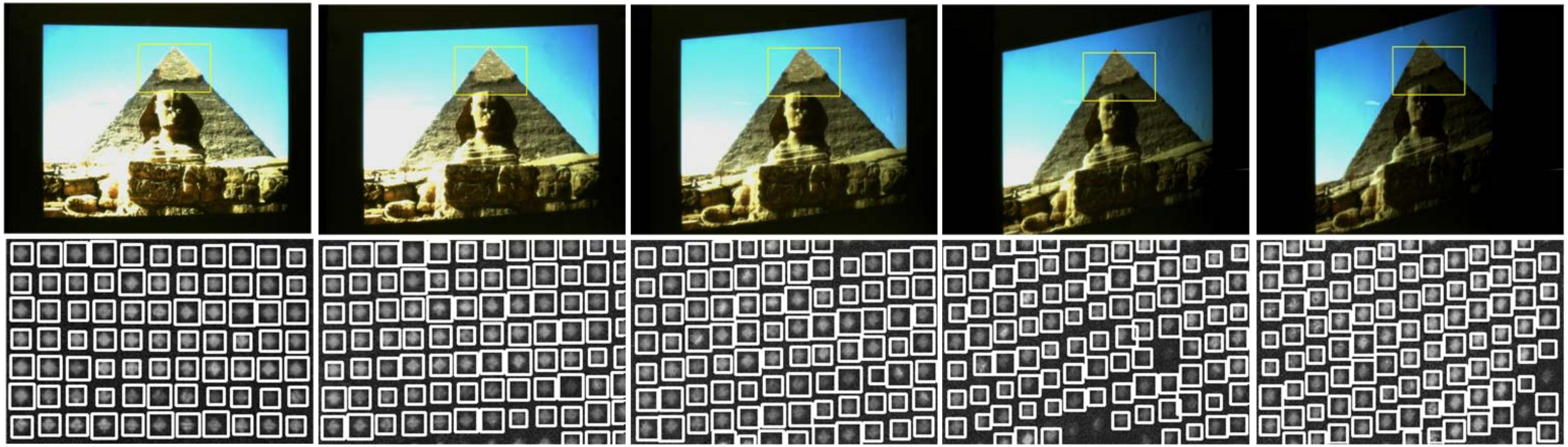
1200mm



1600mm

Sensitivity Evaluation: Projection Surface Orientation

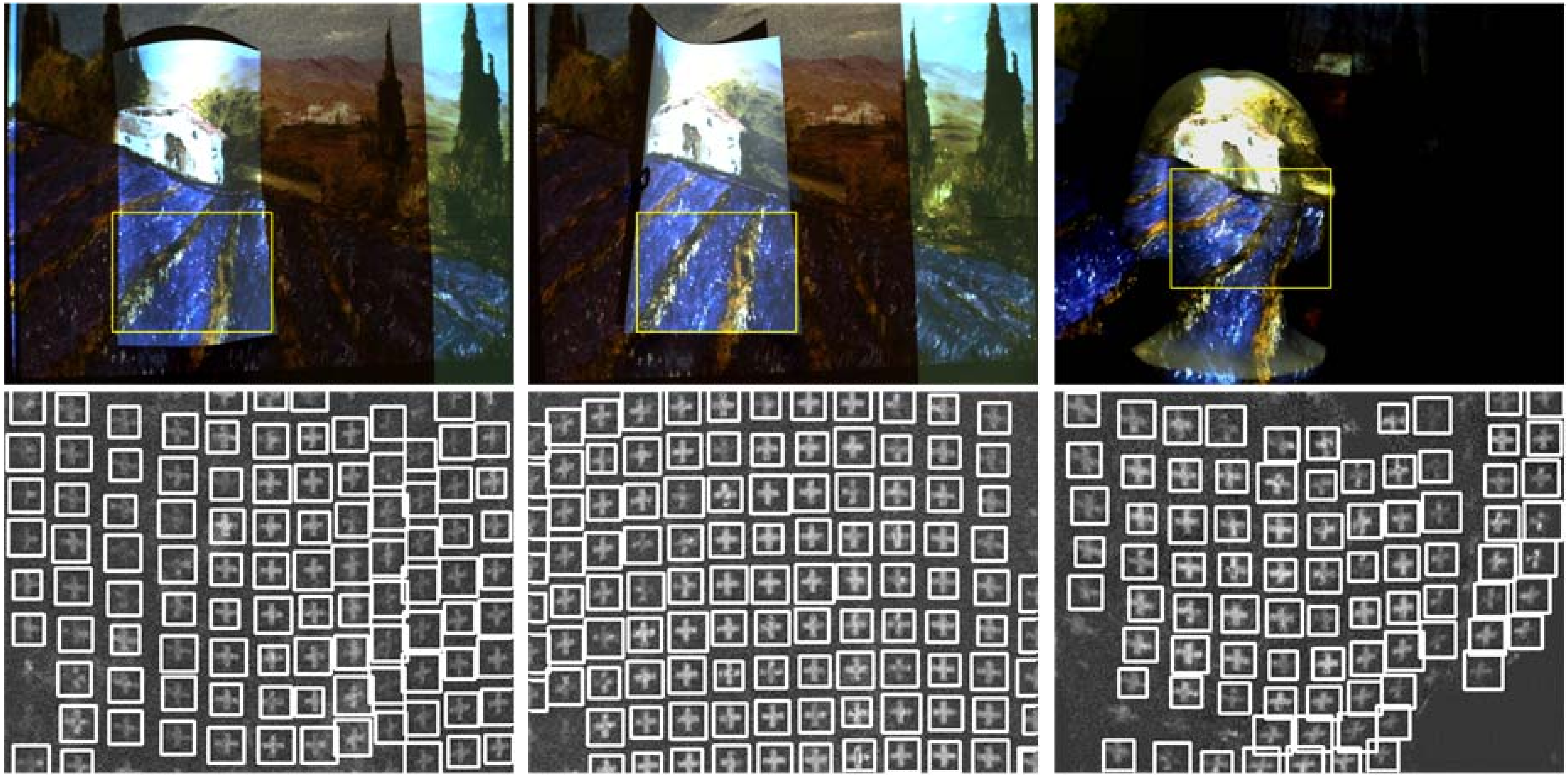
In training data collection:
Surface Orientation: $\theta = 0^{\circ}$



$\boldsymbol{\theta} = \mathbf{10}^{\circ}$ $\boldsymbol{\theta} = \mathbf{20}^{\circ}$ $\boldsymbol{\theta} = \mathbf{30}^{\circ}$ $\boldsymbol{\theta} = \mathbf{40}^{\circ}$ $\boldsymbol{\theta} = \mathbf{50}^{\circ}$

Sensitivity Evaluation: Projection Surface Shape

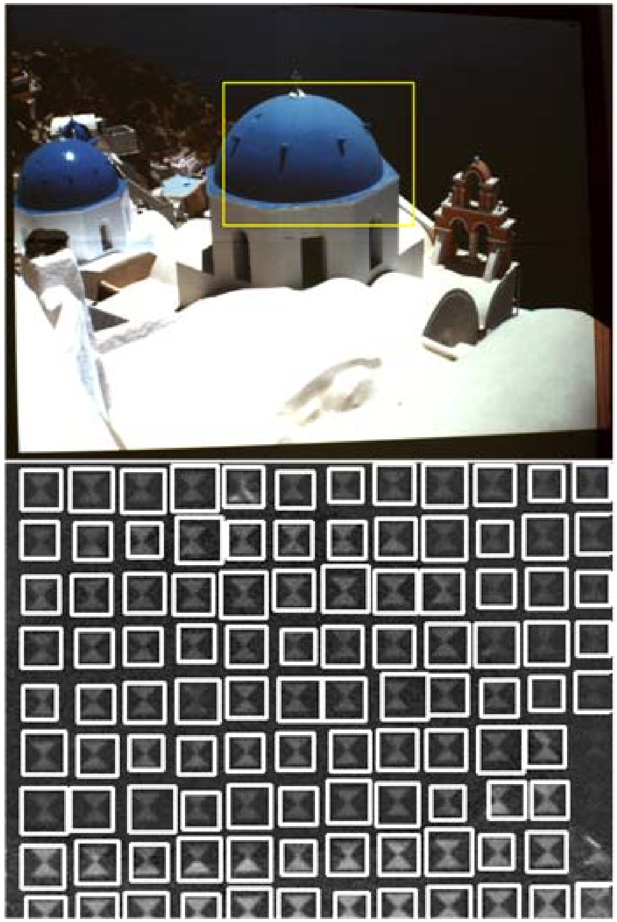
In training data collection:
Projection Surface: Planar



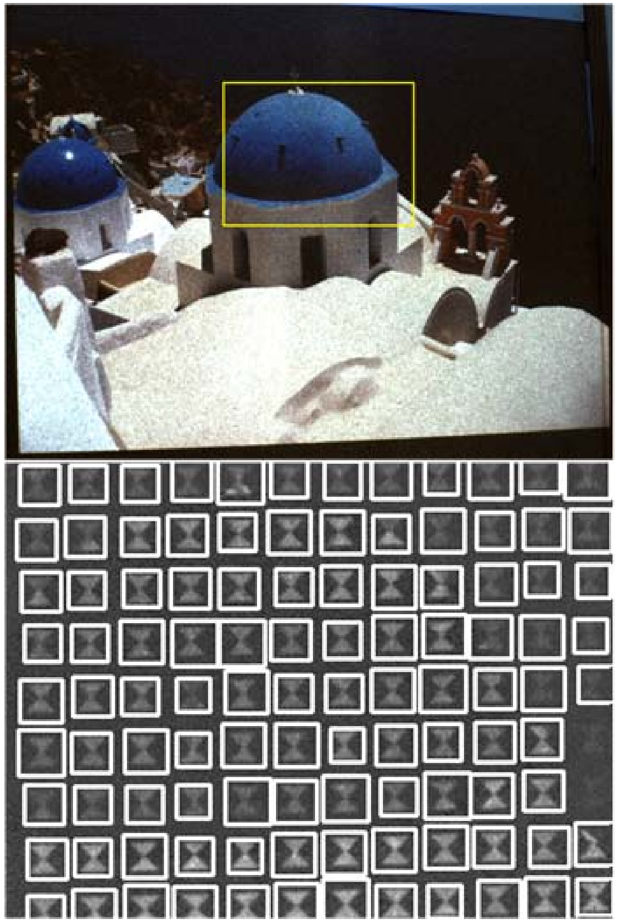
Convex Surface | Concave Surface | Plaster Statue

Sensitivity Evaluation: Projection Surface Texture

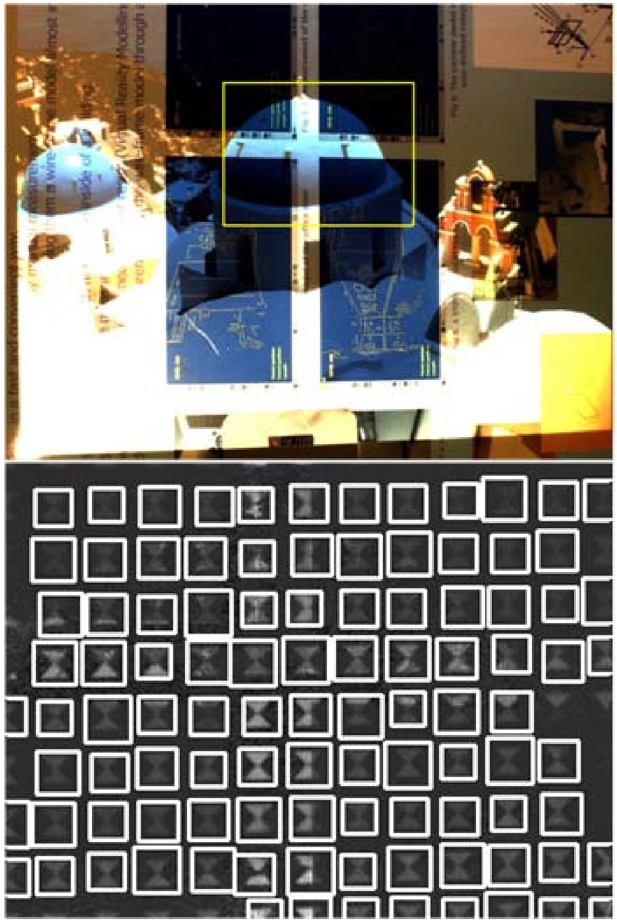
In training data collection:
Working distance: 800mm



Green Paper



Cork Board

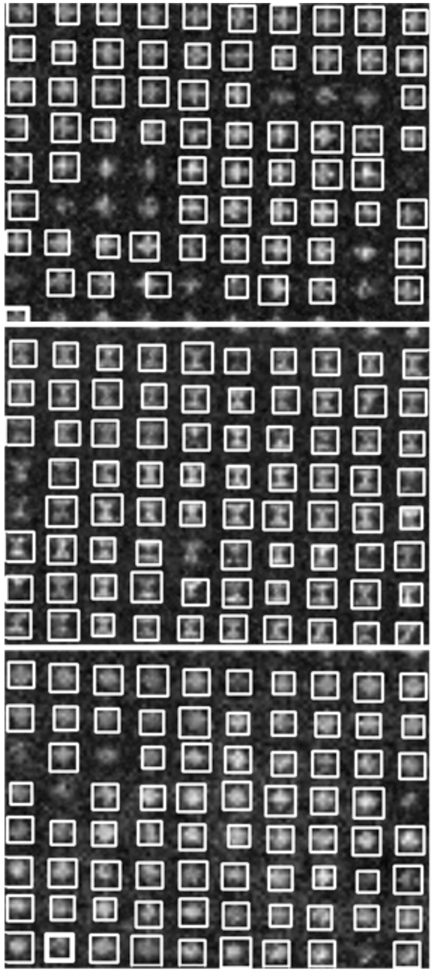


Poster

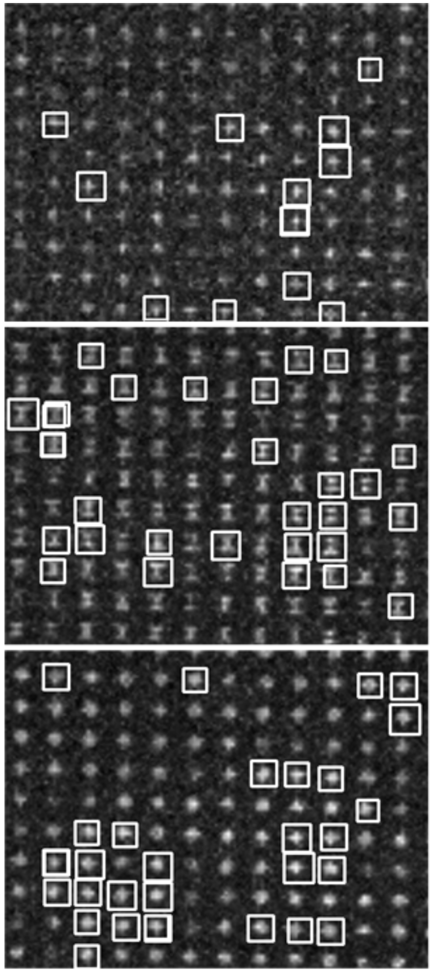
Sensitivity Evaluation: PROCAMS

In training data collection:
PROCAMS: PROCAMS-A

Captured Image



Cropped Patt.



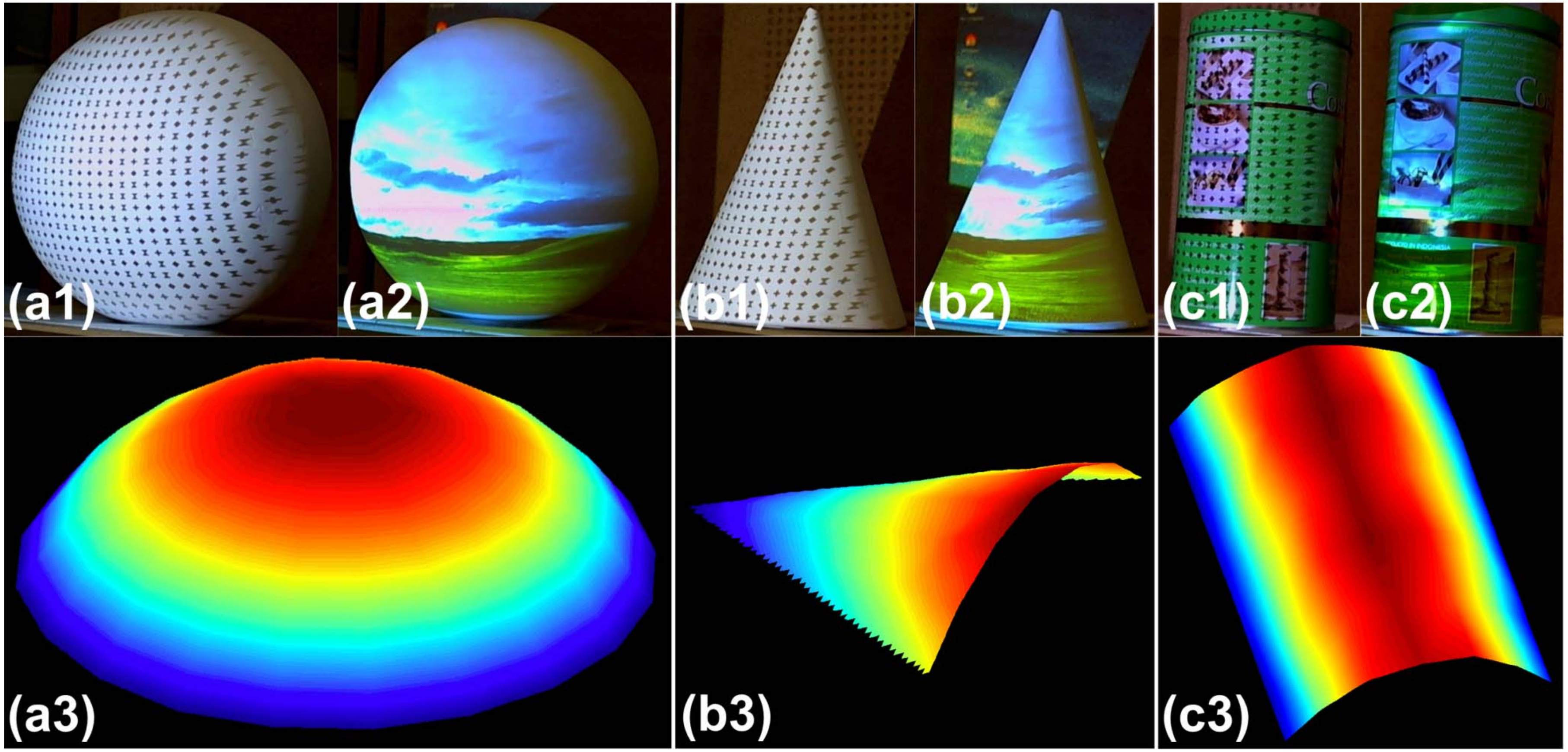
Resized Patt.

Sensitivity Evaluation: Conclusion

Condition	Hits (%)	Missed (%)	False (%)	Ed (pixel)
Benchmark	94.53	3.95	1.52	1.632
Distance (500mm)	86.21	11.63	2.16	1.814
Orientation (50 degree)	85.91	12.03	2.06	2.728
Surface (Plaster Statue)	84.81	13.33	1.86	2.028
Texture (Poster)	91.74	6.63	1.63	2.024
PROCAMS (Cropped Pattern)	80.23	14.43	5.34	3.028

For more detailed sensitivity evaluation results, please refer to the chapter 4.5 in thesis

Applications: 3D Reconstruction

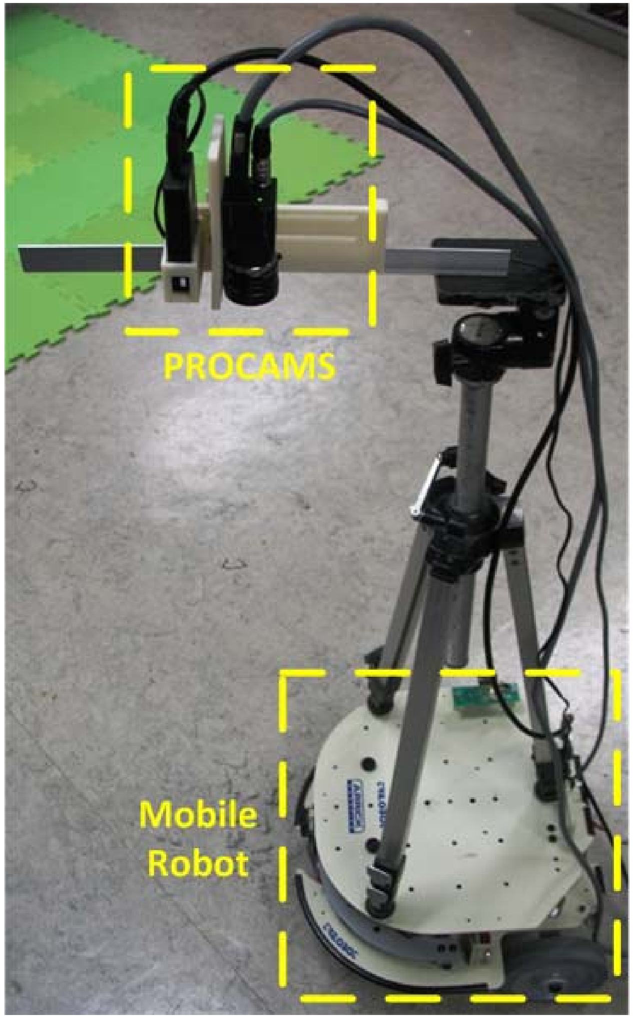


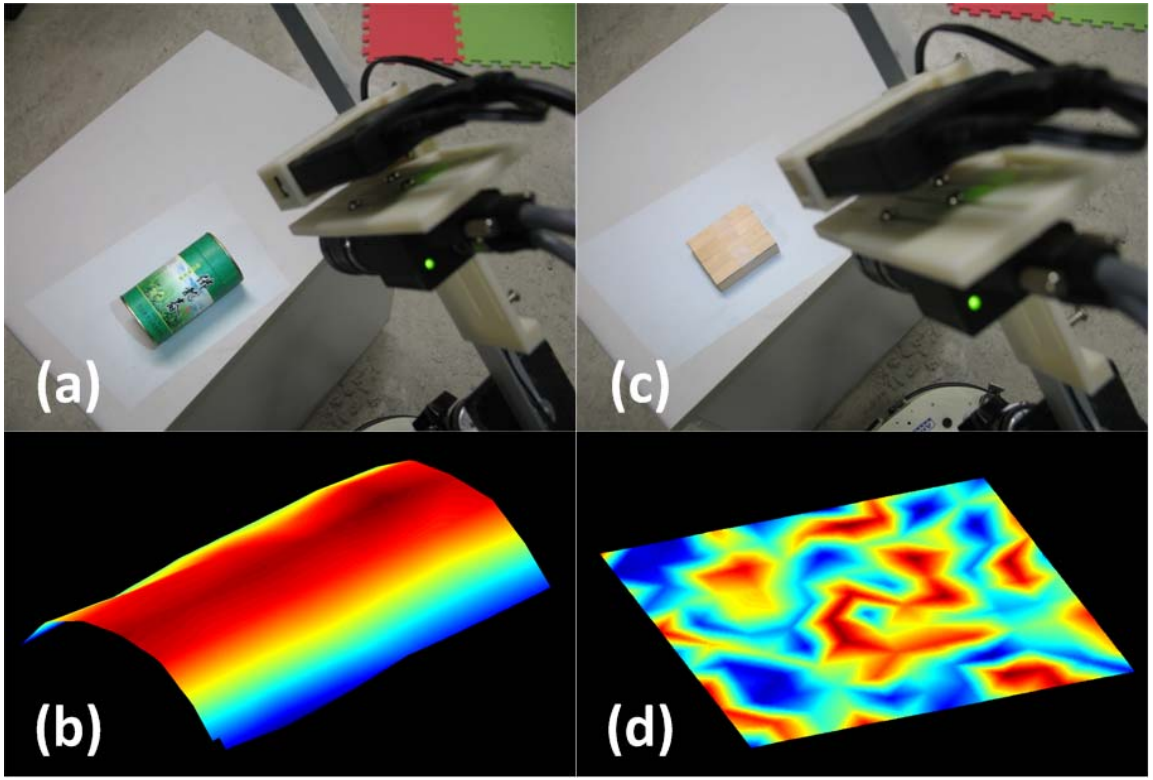
3D Reconstruction with Regular Video Projection

3D reconstruction accuracies on different objects

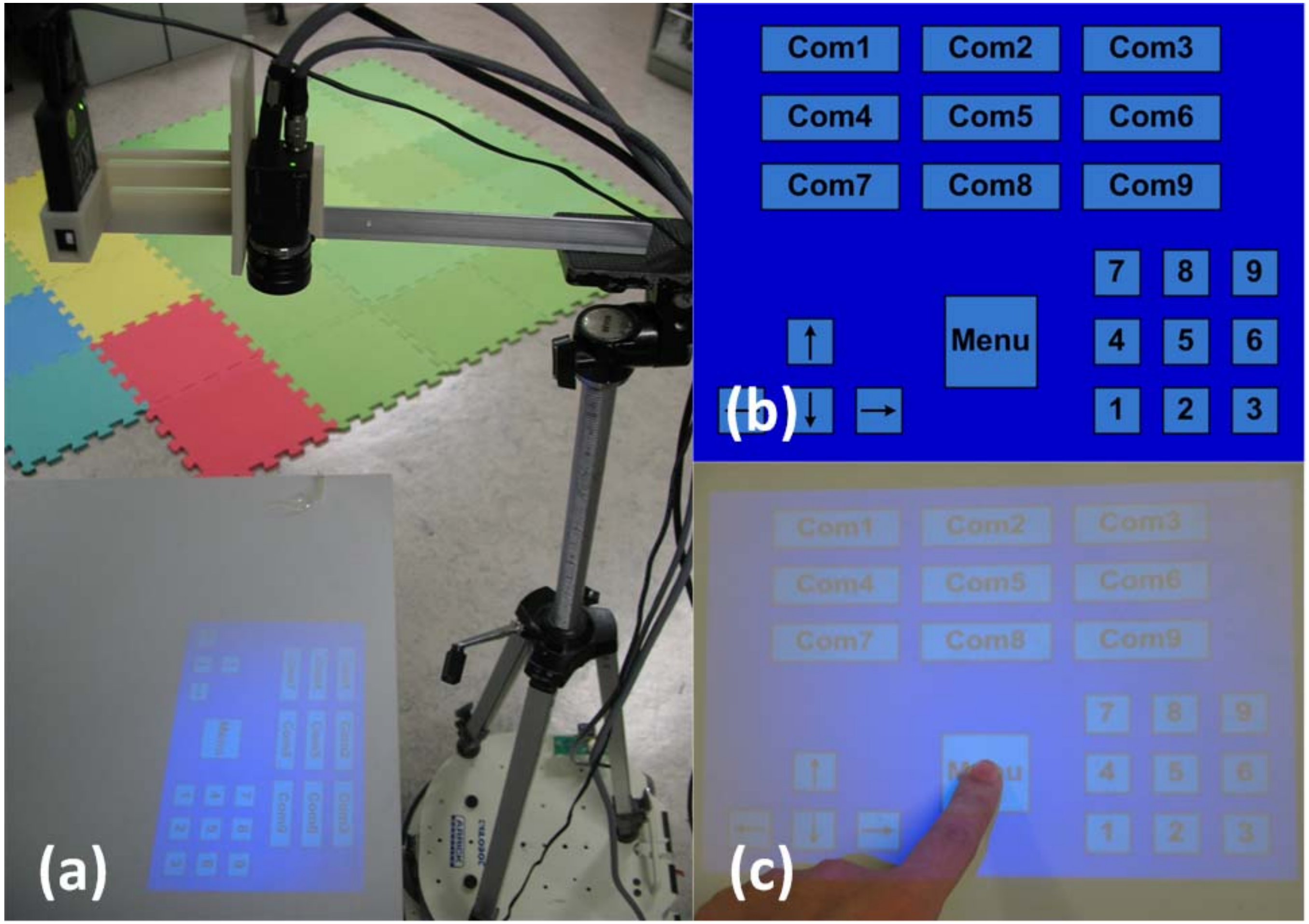
Object	General SL [10]		Our Method	
	$E_\mu(mm)$	$E_\sigma(mm)$	$E_\mu(mm)$	$E_\sigma(mm)$
Sphere	1.502	0.576	1.410	0.587
Cylinder	2.054	0.824	1.939	0.762
Cone	1.383	0.557	1.391	0.564

Applications: Mobile Robot Platform





Applications: Mobile Robot Platform



Summary

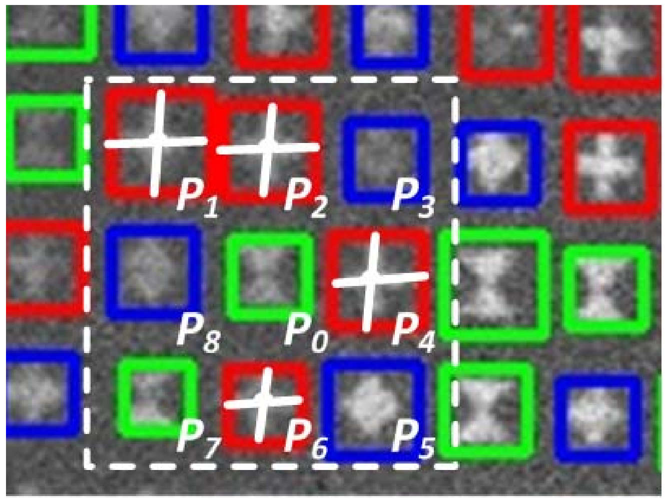
Noise-Tolerance Scheme

Coding

- ❑ specifically designed shapes
- ❑ large hamming distance

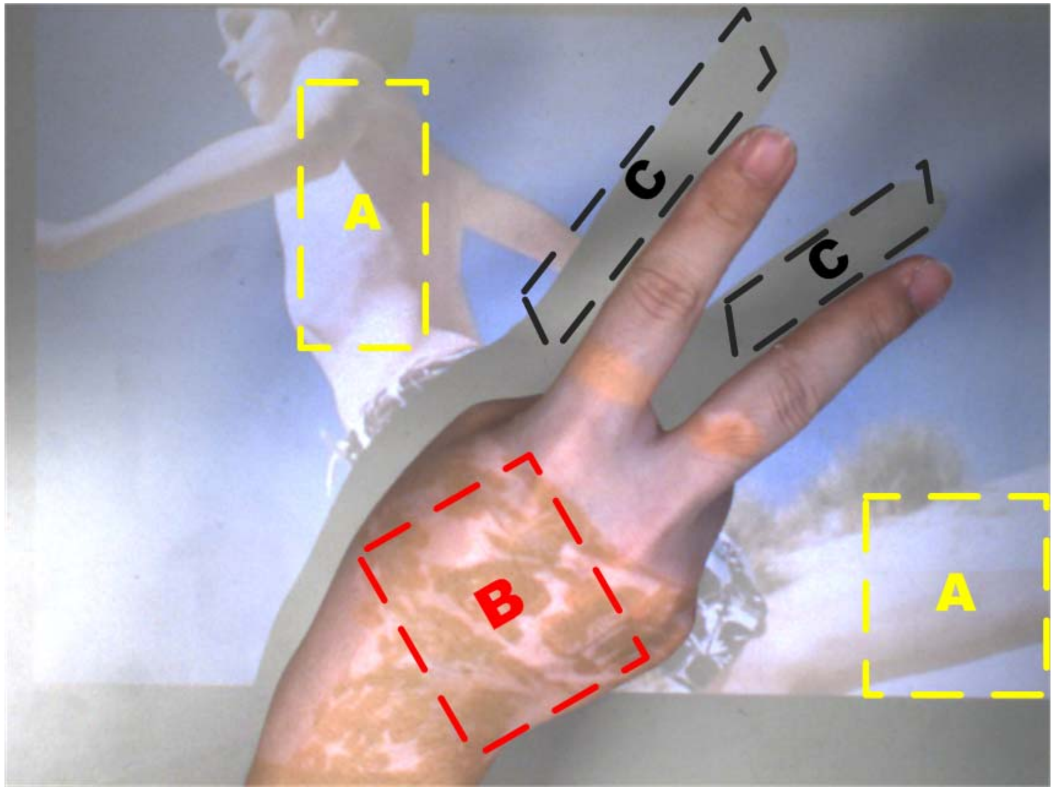
Decoding

- ❑ Pre-trained shape detector



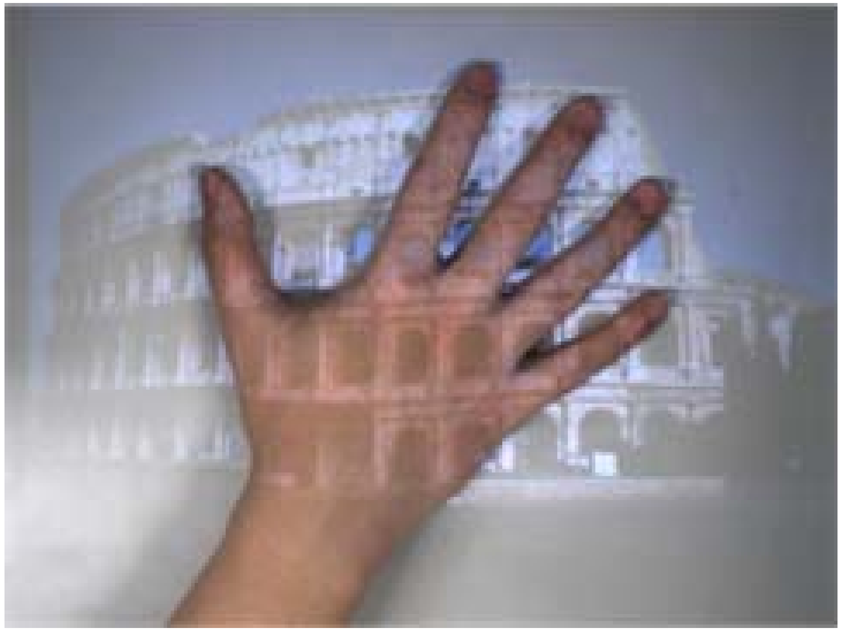
Contents

Background



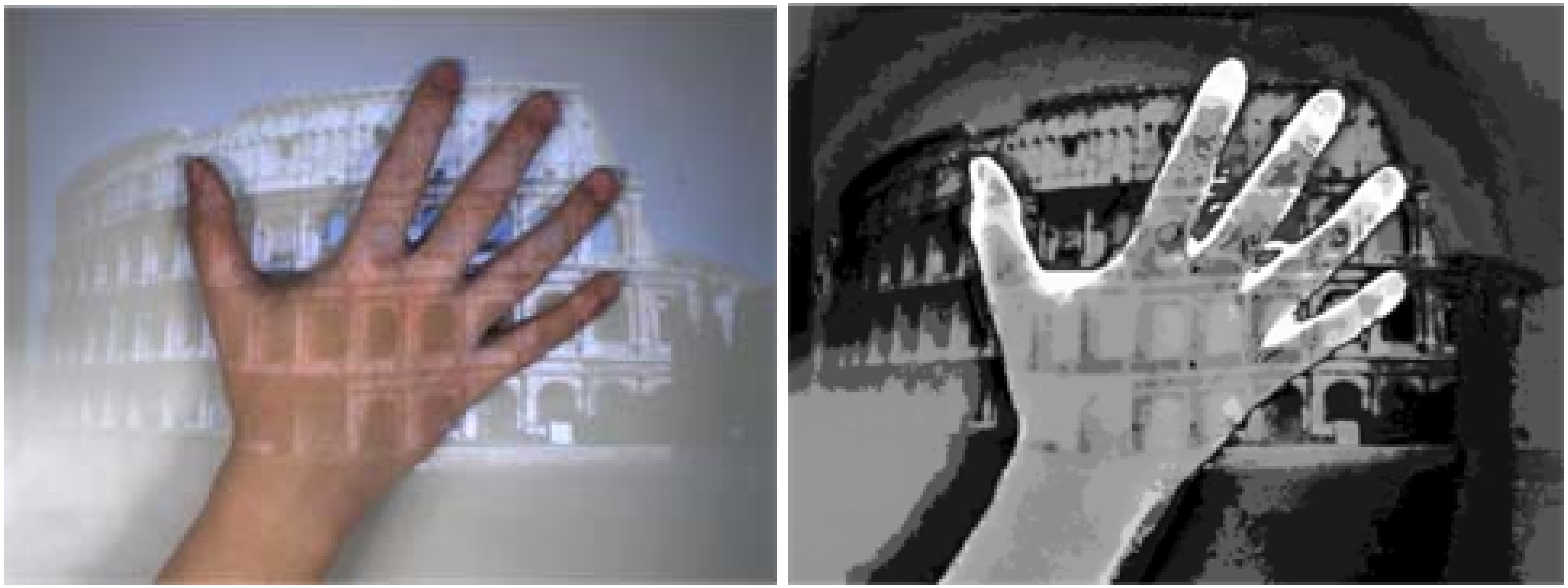
- Skin-color
- Background Subtraction
- Graph-based Approaches
- Additional Sensors (inferred camera, stereo camera, depth camera).

Saliency Detection



- Saliency Detection
  - 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



- Saliency = color contrast to all other pixels

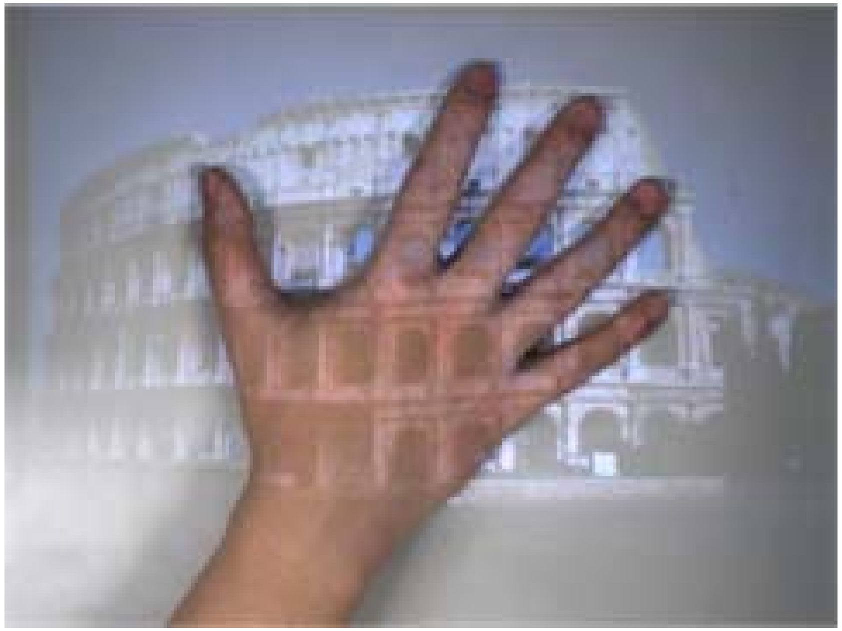
$$S(I_k) = \sum_{i=1}^{N} D(I_k, I_i)$$

Note: Some implementation issues are detailed in [Cheng2011].

Mean-Shift Region Smoothing

- Mean-Shift based smoothing in the regions that are highlighted.
- The image is divided into several **candidate partitions**, while the **boundary of the hand** is preserved well.

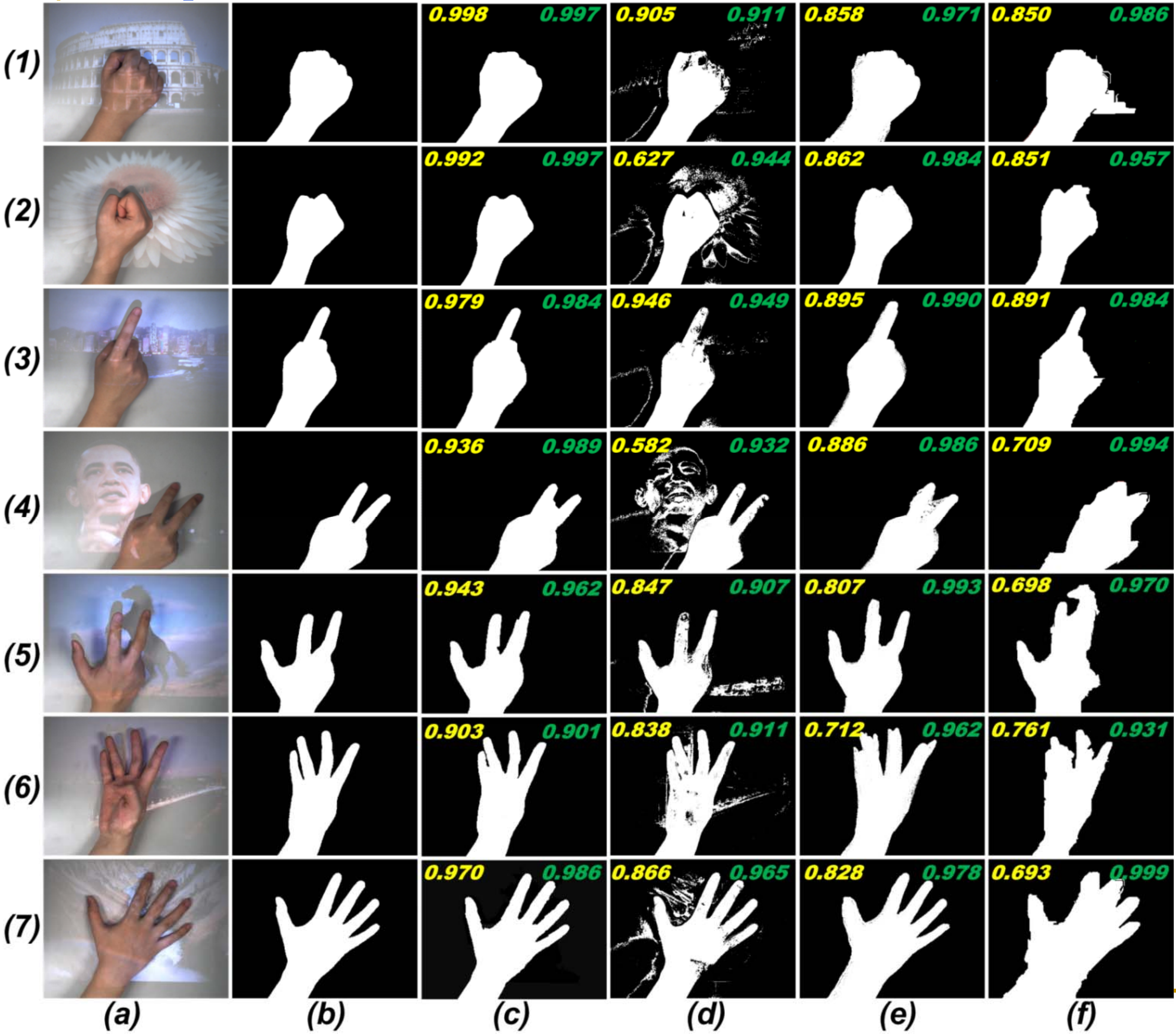
Precise Segmentation by Fusing



Confidence Function

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

Experiments



(a) Original Image

(b) Ground-truth **(Interactive GrabCut)**

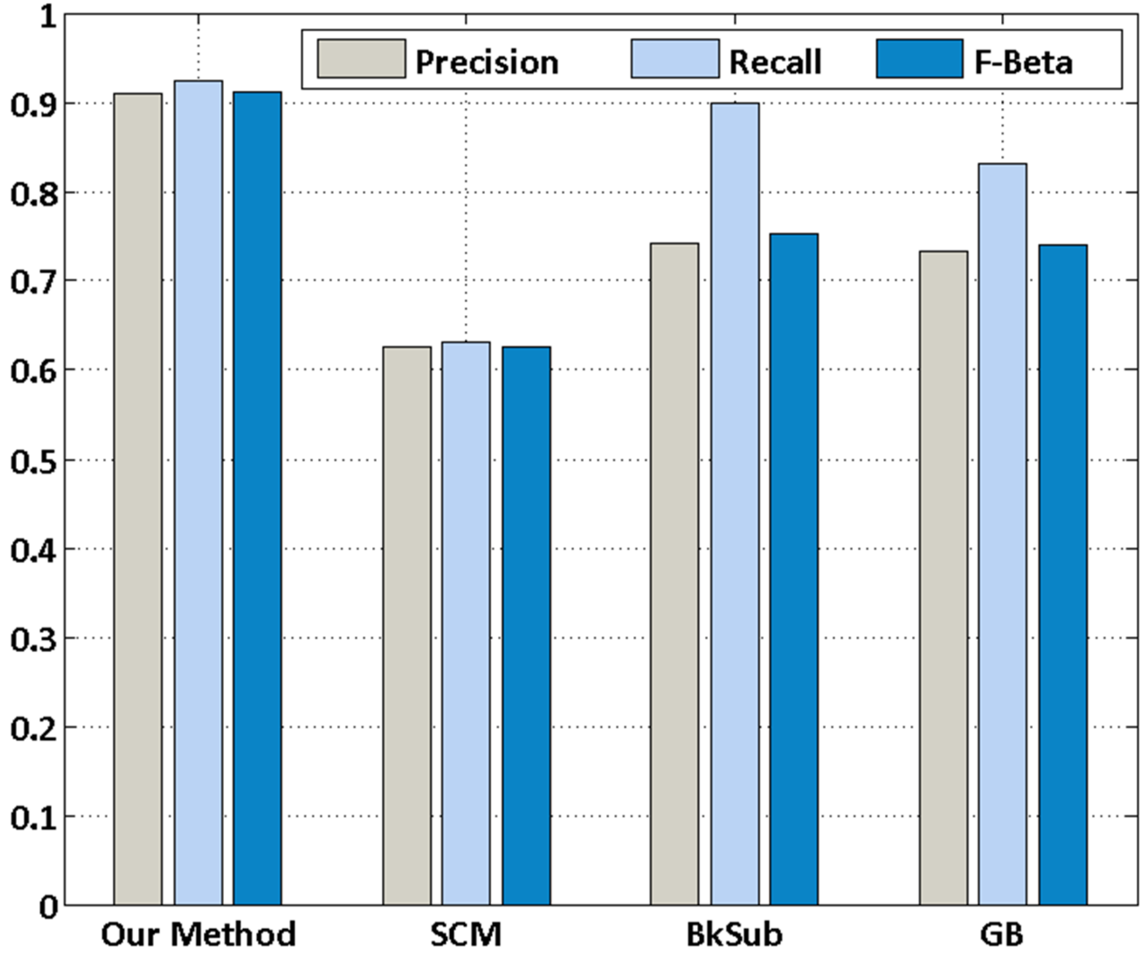
(c) Our method

(d) Statistical color-model **(most cited)**

(e) Background Subtraction **(more accurate)**

(f) Graph-based **(Recent approach)**

Results



Precision-Recall bars for hand segmentation using different methods. Our method shows **high precision, recall and F-Bata values.**

Summary

- Combine **Contrast Saliency** and **Region Discontinuity** for Precise Hand Segmentation in PROCAMS

Contents

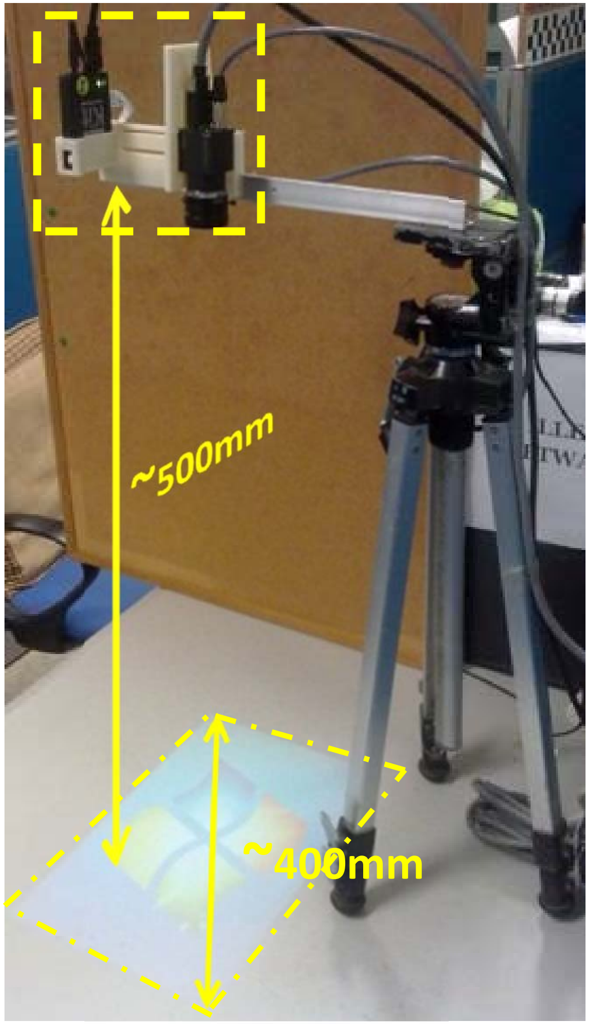
- Motivation & Challenges
- Head Pose Estimation by ISL
- Embedding Invisible Codes into Regular Video Projection
- Hand Segmentation in ProCams
- Touch-Sensitive Display in Arbitrary Planar Surface
- Conclusion and Future Work

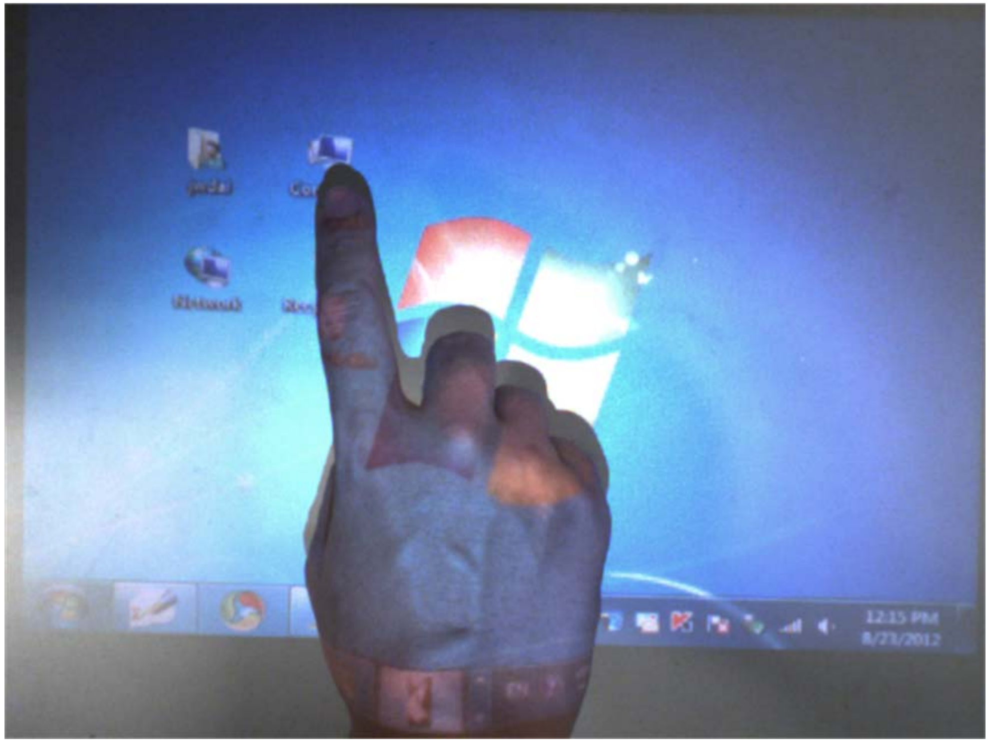
Motivation

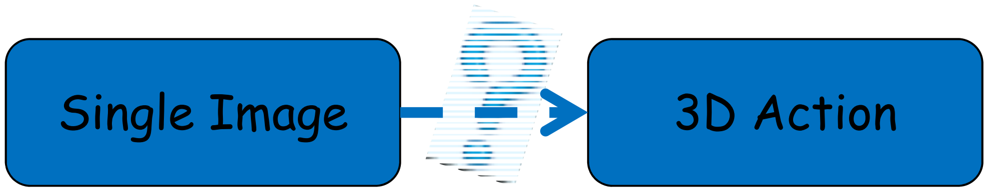
VS.

Bigger Display

Portability


~500mm
~400mm

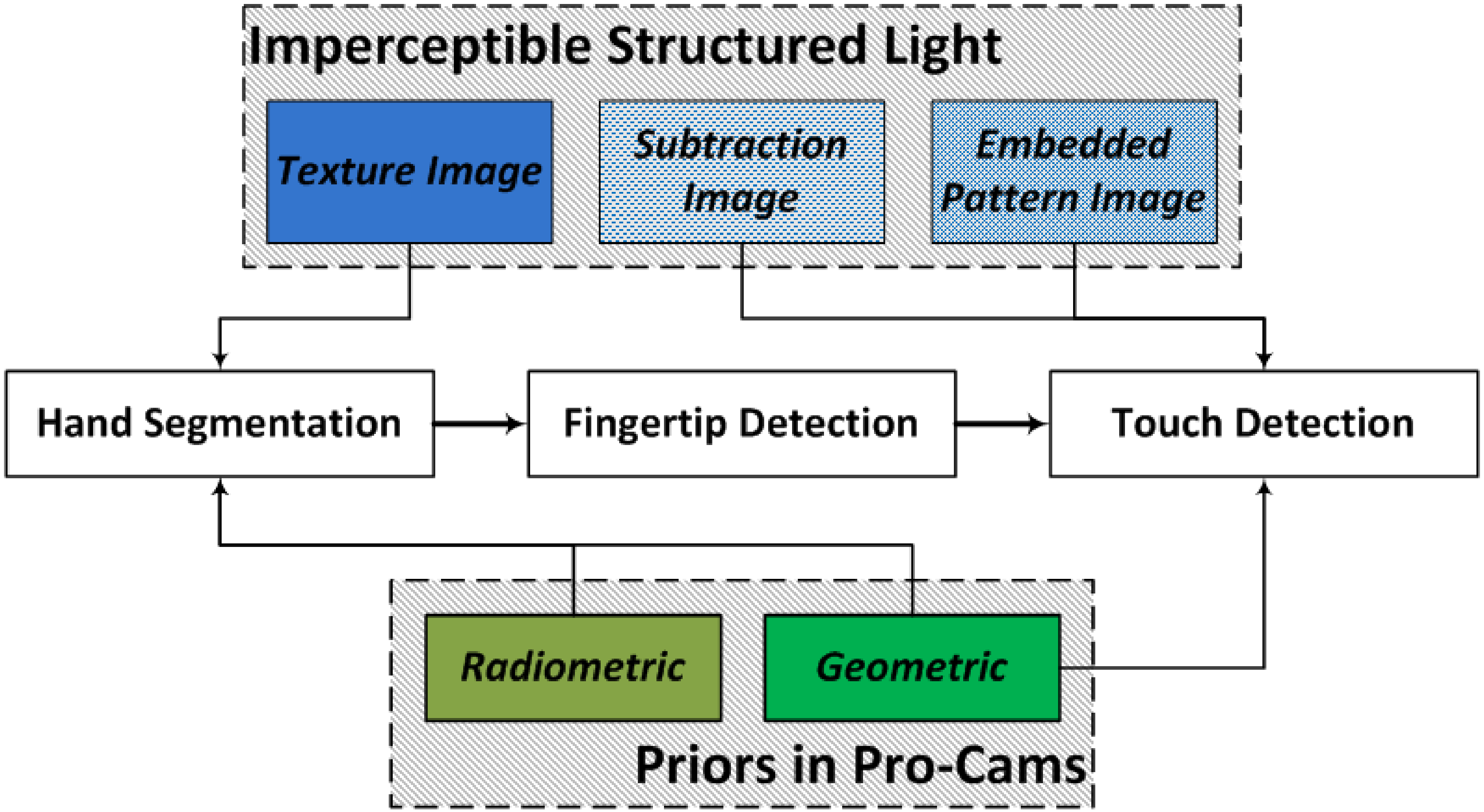



Single Image
?
3D Action

Previews Works

- **Additional Sensors**
  - **Light Touch** (IR optical sensors)
  - **Diamondtouch** (capacitive sensor array)
  - **Smartskin** (mesh-shaped antenna)
  - **Skinput** (bio-acoustic sensing array)
  - **LightSpace**, **Omnitouch** (Kinect)
- **Computer Vision**
  - **[Letessier2004]** -- Fingertip tracking, not touching detection
  - **[Kjeldsen2002, Hardenberg2001]** -- Delay-based scheme
  - **[Marshall2008]** – Color change of the fingernail
  - **[Song2007, PlayAnywhere2005]** -- Shadow casted by finger
  - **[Fitriani2007]** -- Deformation on soft surface

Overview


Imperceptible Structured Light
Texture Image
Subtraction Image
Embedded Pattern Image
Hand Segmentation
Fingertip Detection
Touch Detection
Radiometric
Geometric
Priors in Pro-Cams

Priors in Projector-Camera System

- **Geometric (Homography)**

Camera's image plane — Projector's projection panel

$$
\begin{tikzcd}
\Pi_C \arrow[rr, "H_{CP}"'] \arrow[dr, "H_{CT}"'] & & \Pi_P \arrow[dl, "H_{PT}"] \\
& \Pi_T &
\end{tikzcd}
$$

Table surface

This geometric priors can be derived through **2 projection-capture cycle** in initialization stage.

Embedded Pattern Design Strategy

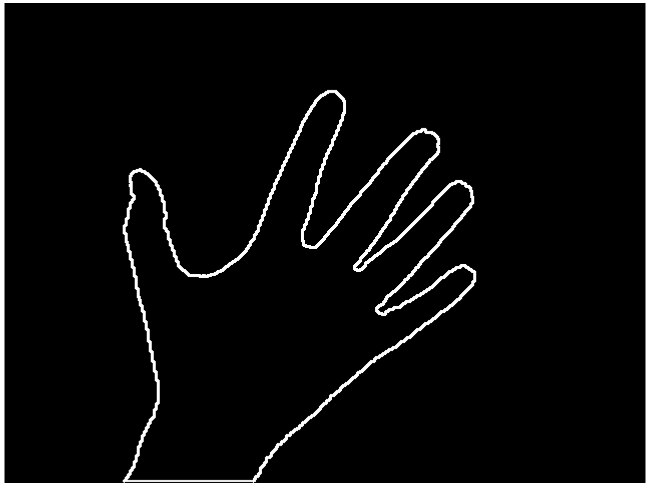
Method	Array Size	Win. Size	Alph. Length
[Morita 1988]	24 * 24	3 * 4	2
[Kiyasu 1995]	18 * 18	4 * 2	2
[Salvi 1998]	29 * 29	3 * 3	3
[Spoelder 2000]	65 * 63	2 * 3	2
[Dai 2012]	27 * 29	3 * 3	3
[Desjardins 2007]	53 * 38	3 * 3	3
[Chen 2008]	82 * 82	3 * 3	7

Summary of typical spatial coding methods

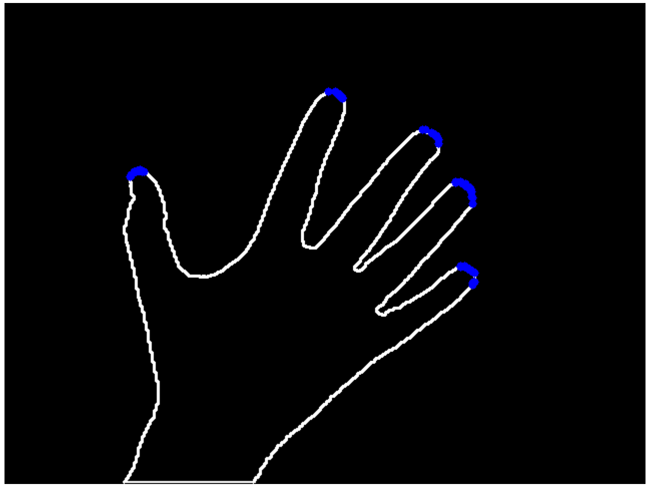
- Constraints of Pattern Generation
  - Code Uniqueness
  - Large Hamming Distance

Hand Segmentation & Fingertip Detection

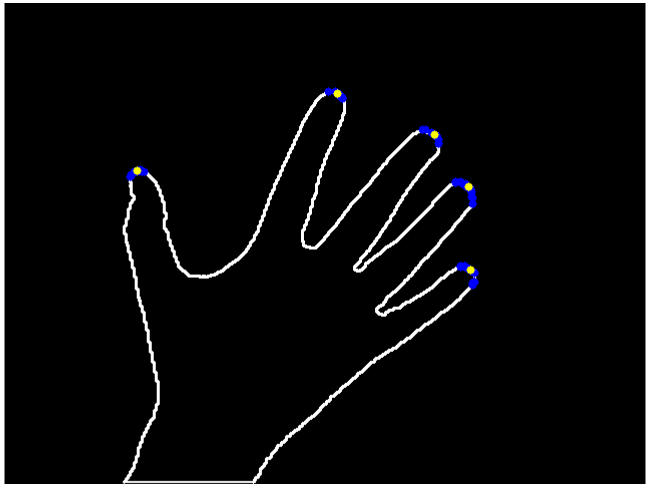
Binary Hand Image



Hand Contour

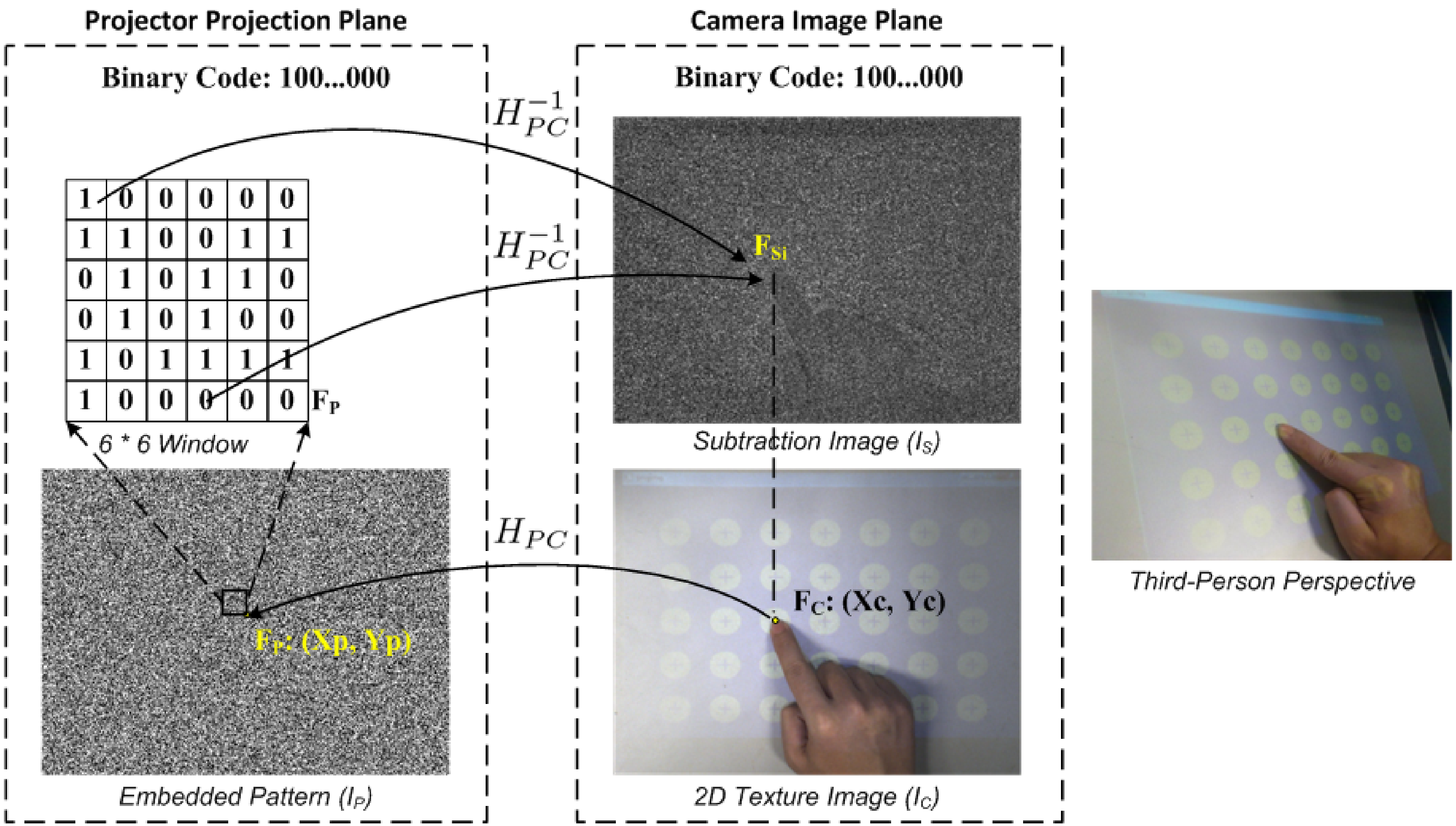


Fingertip Candidates



Detected Fingertips

Touch Detection Through Homography and Embedded Codes



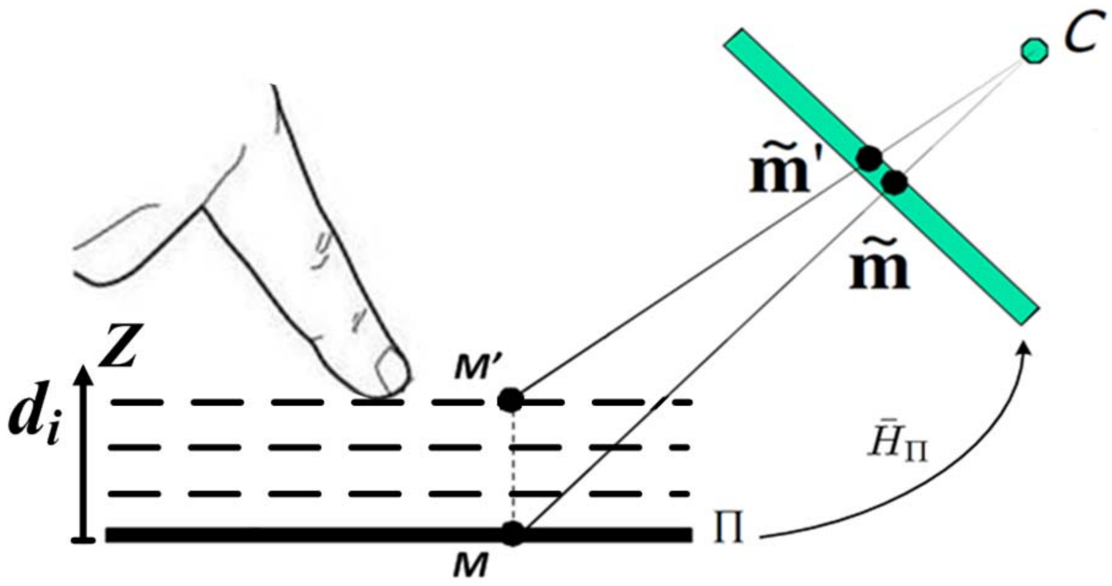
Embedded Pattern (I_P)

Subtraction Image (I_S)

2D Texture Image (I_C)

Third-Person Perspective

From Resistive Touch to Capacitive Touch or Floating Touch



Experiments
Touch Accuracy Evaluation

Experiments
Touch Accuracy Evaluation

Surface	Illumination			
	Dark		Normal	
	ϵ(px)	FRR/FAR(%)	ϵ(px)	FRR/FAR(%)
Gray	2.98	1.12/0.45	3.05	1.32/0.48
Yellow	3.04	1.23/0.57	3.12	1.54/0.61
Artifact	3.12	1.77/0.67	3.20	1.76/0.63

Comparison with recent depth-camera sensing based methods

In [Wilson 2010], the informal observed spatial error of finger detection on planar surface was **between 3-6 pixels**,

In [Omni-Touch 2011], the FRR and FAR of finger click detection on four different surfaces were **0.8% and 3.3%**.

Experiments
Trajectory Tracking

- Video

Experiments
Multiple-Touch

Summary

- Using only **off-the-shelf devices**
- Achieving 3D sensing **without explicit 3D reconstruction**
- Use of **prior knowledge** to pixel-wise coding

Contents

Conclusion and Contribution

- **A novel 6-DOF head pose estimation approach by imperceptible structured light sensing.**

  Combine 2D & 3D information to achieve continuous, accurate and real-time head pose estimation.
  (ICRA2011)

- **Embedding invisible patterns into regular video projection to make projector both a display device and a 3D sensor.**

  Robust coding scheme; accuracy decoding method through pre-trained primitive shape detector.
  (ISVC2012, IROS2012, WoRV2013*, IEEE TCSVT*)

Conclusion and Contribution

- **A novel coarse-to-fine hand segmentation method in projector-camera system.**
  Combine contrast saliency and region discontinuity to segment the hand under projector's illumination.
  (ICRP2012, IEEE TPAMI*)

- **A touch-sensitive display on arbitrary planar surface.**
  Just by use of mere a projector and a camera.
  (PROCAMS2012, IEEE TPAMI*)

Note: (*) indicates the papers are under review or prepared for submission.

Future Work

- **Motion compensation** for the displacement between successive images resulting in **blur or destruction** of the embedded codes in the difference image.

- **Image enhancement approach** to increase the low signal-to-noise ratio of subtraction image.

- Extension to **multi-hand supporting** and **advanced touch gestures recognition** in the touch-sensitive interface.

Related Publications

Conference paper

[1] J. Dai and R. Chung, Head Pose Estimation by Imperceptible Structured Light Sensing, In *Proc. of IEEE International Conference on Robotics and Automation* ***(ICRA'11)***, pages 1646-1651, May 2011.

[2] J. Dai and R. Chung, Making Any Planar Surface into a Touch-sensitive Display by a Mere Projector and Camera, In *Proc. of 9th IEEE International Workshop on Projector–Camera Systems* ***(PROCAMS2012)***, June 2012.

[3] J. Dai and R. Chung, On Making Projector both a Display Device and a 3D Sensor, In *Proc. of The 8th International Symposium on Visual Computing* ***(ISVC'12)***, July 2012.

[4] J. Dai and R. Chung, Embedding Imperceptible Codes into Video Projection and Applications in Robotics, To Appear in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems* ***(IROS'12)***, October 2012.

[5] J. Dai and R. Chung, Combining Contrast Saliency and Region Discontinuity for Precise Hand Segmentation in Projector-Camera System, To Appear in *Proc. of The 21st International Conference on Pattern Recognition* ***(ICPR'12)***, November 2012.

[6] J. Dai and R. Chung, Sensitivity Evaluation of Embedded Code Detection in Imperceptible Structured Light Sensing, *Submitted to IEEE Workshop on Robot Vision* ***(WoRV'13)***, January 2013.

Journal paper

[7] J. Dai and R. Chung, Embedding Invisible Codes into Normal Video Projection: Principle, Evaluation and Applications. *Submitted to IEEE Trans. on Circuit System and Video Technology* ***(TCSVT)***.

[8] J. Dai and R. Chung, Touch-sensitive Display on Arbitrary Planar Surface by a mere Projector and Camera, *Preparated to submit to IEEE Trans. on Pattern Analysis and Machine Intelligence* ***(TPAMI)***.

THANKS
QA