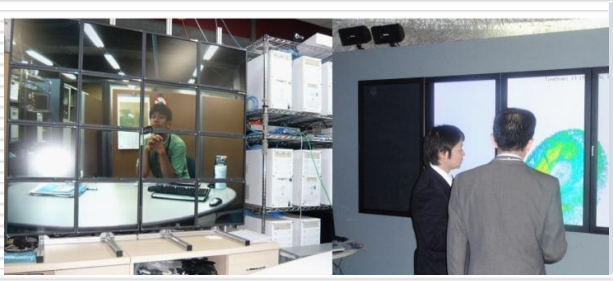
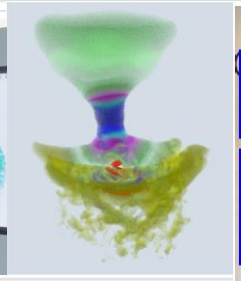
# Video Analysis for Augmented and Mixed Reality

**Kiyoshi Kiyokawa**
**Osaka University**

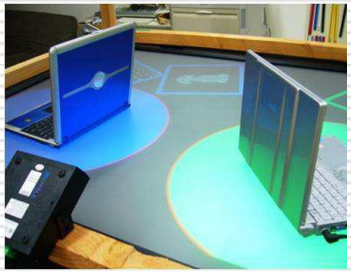
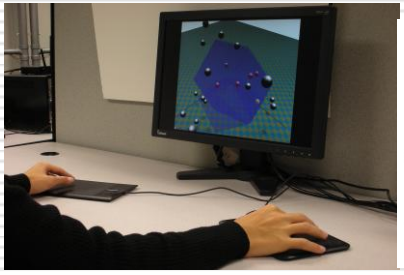
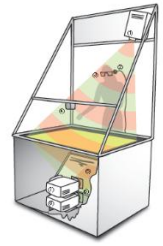
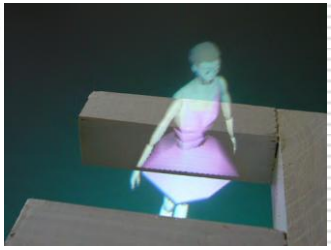
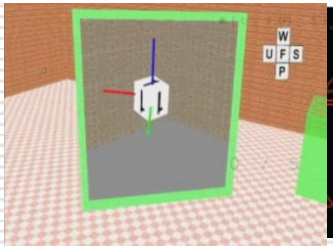
# Who am I ?

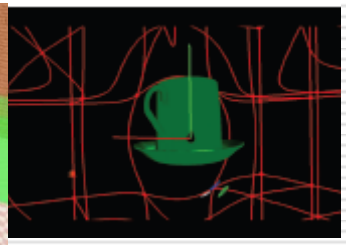
- A researcher on AR / MR / VR / 3DUI / CSCW / Wearable Comp.

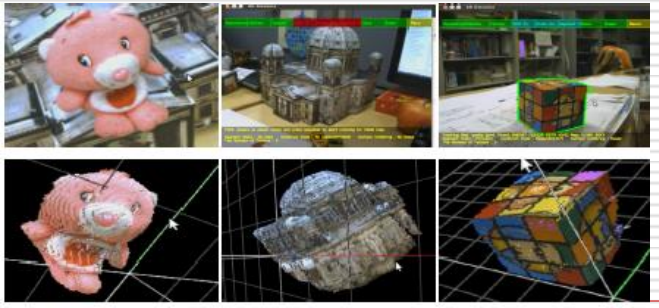
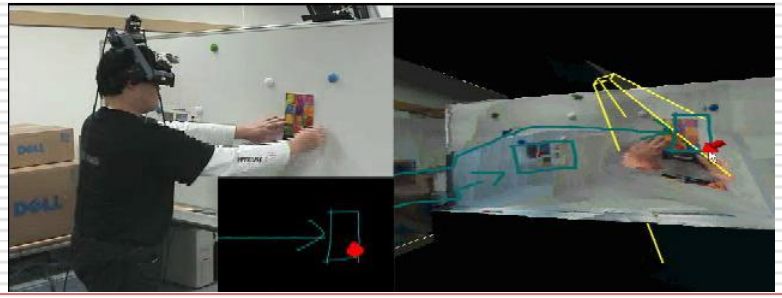
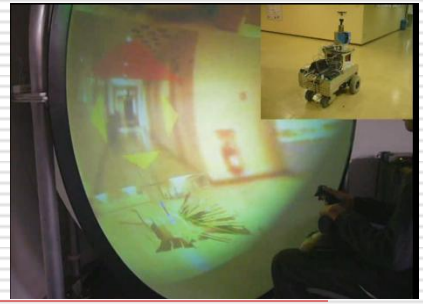



Visualization / Wearable computing / Context-awareness



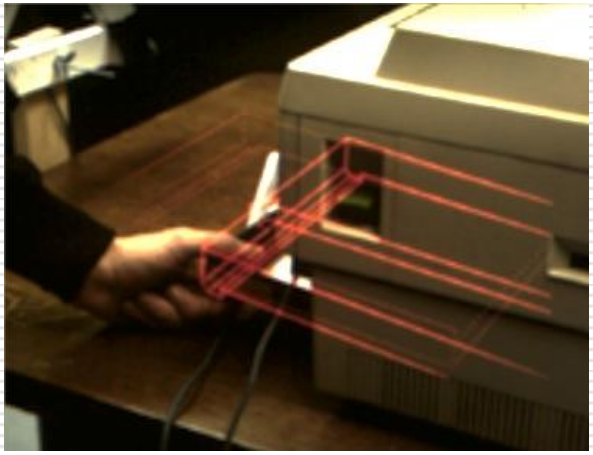
Tabletop display / Multi-touch interaction / New display HW / 3D User Interfaces

3D reconstruction / 3D remote collaboration / Remote robot control

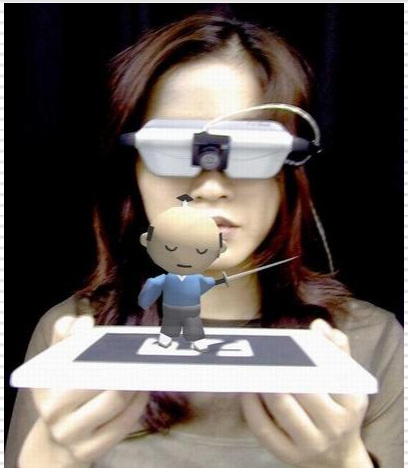
# What is Augmented Reality ?

- Technology that **'augments' the real world** by computer information



Feiner et al., KARMA (1991)

Kiyokawa et al., VLEGO (1997)



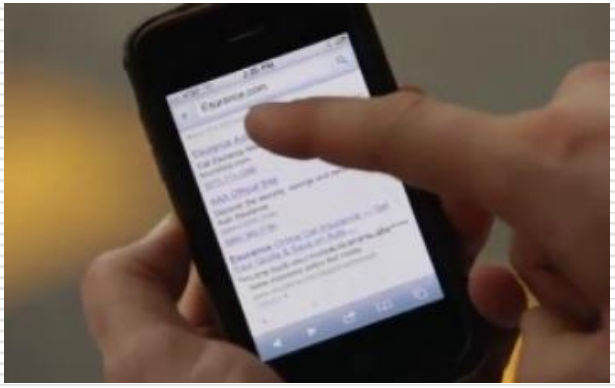
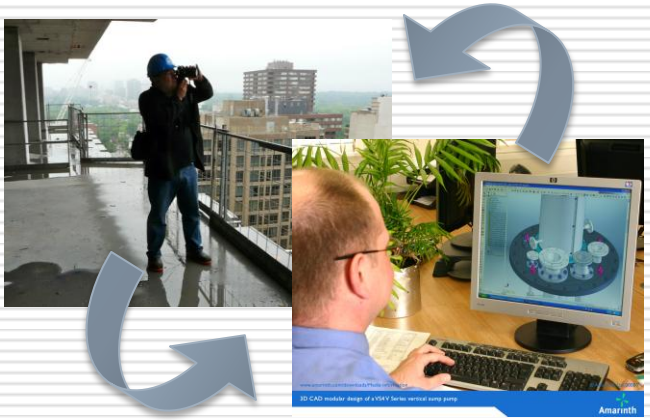
Kato et al., ARToolkit (1999)


SCE, The Eye of Judgment (2007)


Layar (2009)

# Why Augmented Reality ?

- **Direct coupling** of computer intelligence to human behavior provides **intuitiveness** and **efficiency**

| | **Information Retrieval** | **Architectural Review** |
|---|---|---|
| **Conventional User Interface** |  |  |
| | Manual search on the web | Photo-taking / Reviewing on monitor |
| **AR Interface** |   |  |
| | Direct search | Direct assessment |

# Why Augmented Reality ? (cont'd)

- AR-ready devices getting ubiquitous
  - Smartphones / Tablets / Personal Game Consoles
    - Display / Camera / Compass / GPS / Networking

Nintendo 3DS

Google Project Glass

- AR market expected to expand exponentially
  - Global AR Market Forecast by *Research and Market* (2011)
    - Gaming, Automotive, Medical, Defense, E-learning & GPS Applications
    - People use AR
      **0.1%** (2011)→ **1%** (2016)
    - Revenue
      **$181M** (2011) → **$5,156M** (2016)

# What is Mixed Reality ?

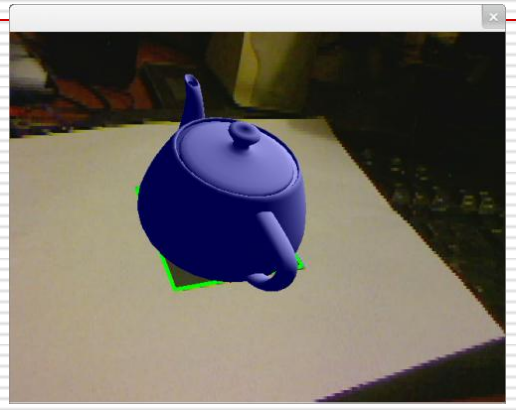
- A superset concept of Augmented Reality
  - A continuum with two extrema, “real” and “virtual”

Real Environment — **Mixed Reality (Milgram et al. ‘94)** — Virtual Environment

- My view
  - AR is a subjective experience and human-oriented
    - **Power-gain** is the key (How useful the UI is)
  - MR is an objective entity and technology-oriented
    - **Consistency** is the key (How well V and R are mixed)

# Video Analysis for AR / MR

- **Geometric Consistency**
  - **Where** should the information be shown ?



- **Photometric Consistency**
  - **How** should the information be rendered ?
  - MR-oriented topic

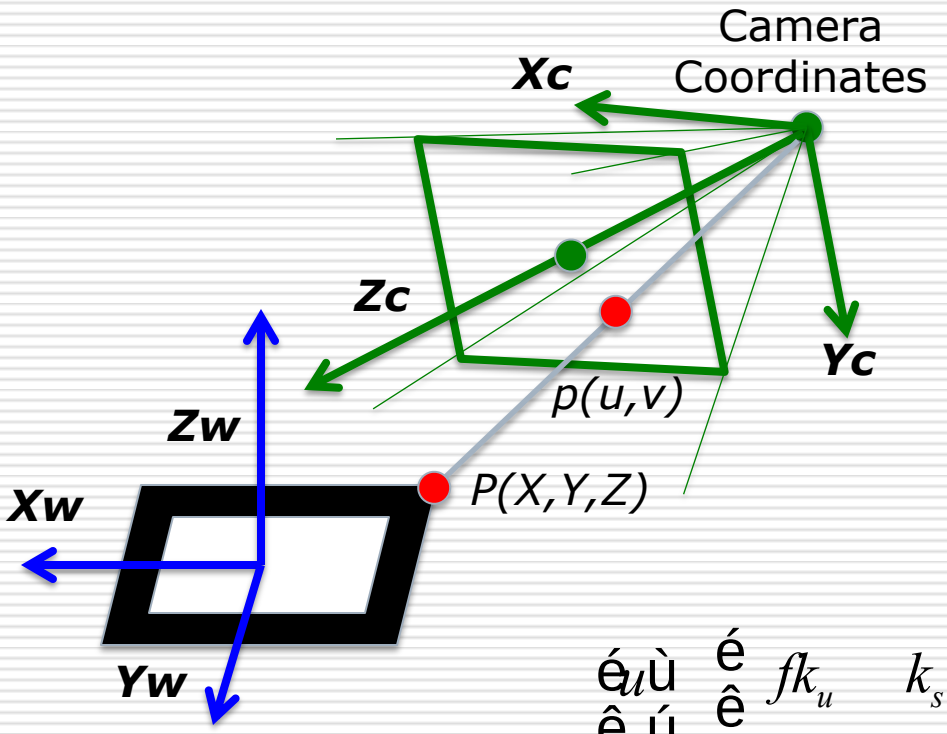
- **Scene Understanding**
  - **What** information should be presented ?
  - AR-oriented topic

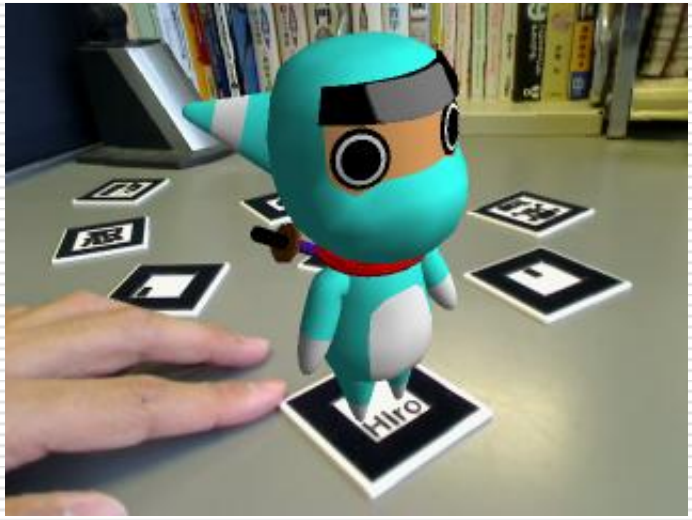
# Geometric Consistency

**Where** should the information be shown ?

# Vision-based Pose Estimation

- Estimates camera pose from captured images



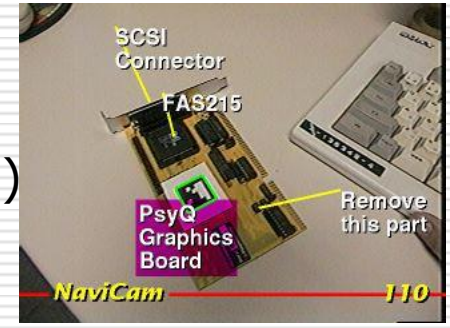

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} fk_u & k_s & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

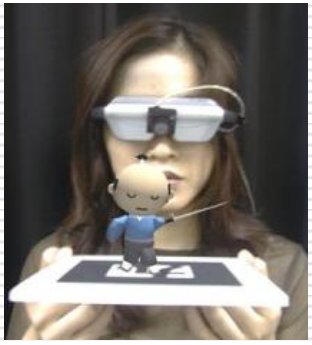
Intrinsic parameters Extrinsic parameters

- But How ?

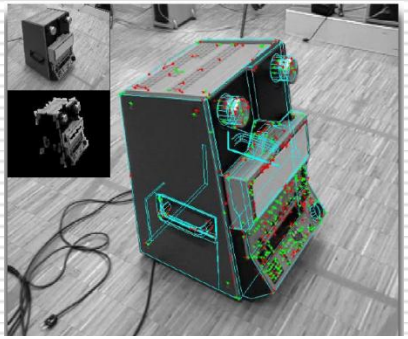
# Vision-based Pose Estimation (cont'd)

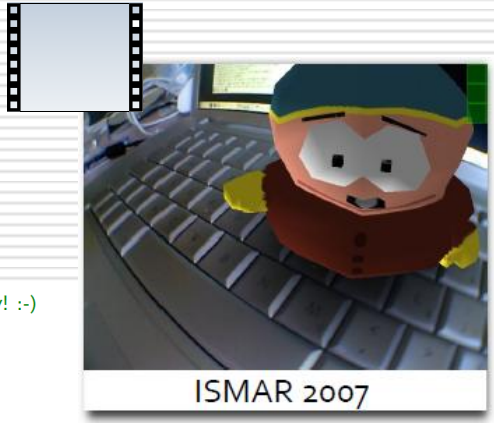
- Marker-based
  - Rekimoto's **CyberCode** ('96)
  - Kato's **ARToolkit** ('99)
  - etc. etc. …



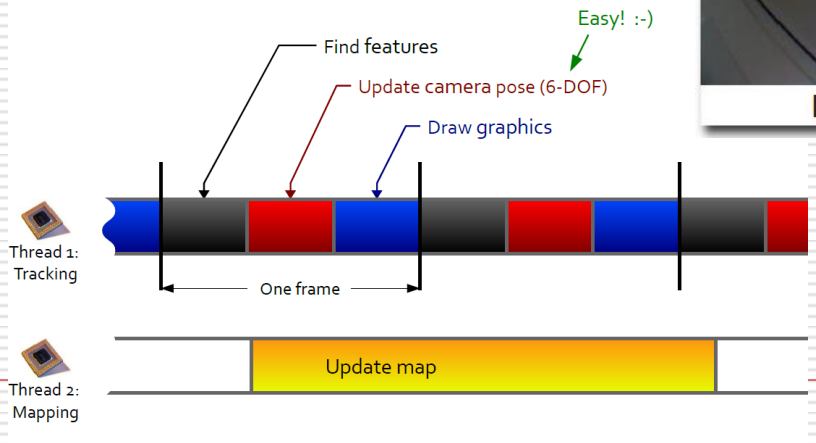




- Model-based
  - Lepetit ('03)
  - etc. etc. …



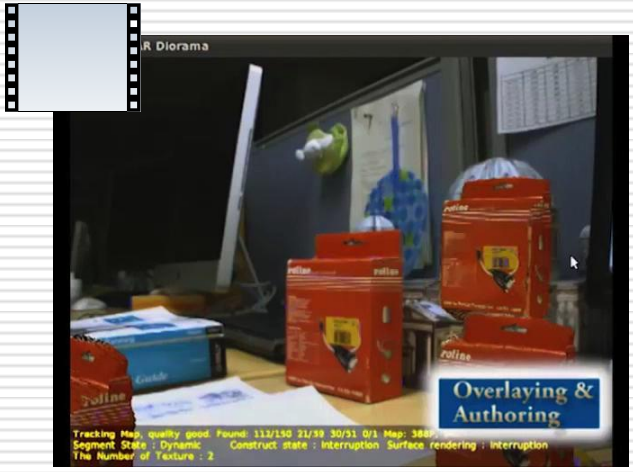



- Feature-based
  - Template-based
  - SLAM
    - Klein's **Parallel Tracking and Mapping** ('07)
  - etc. etc. …

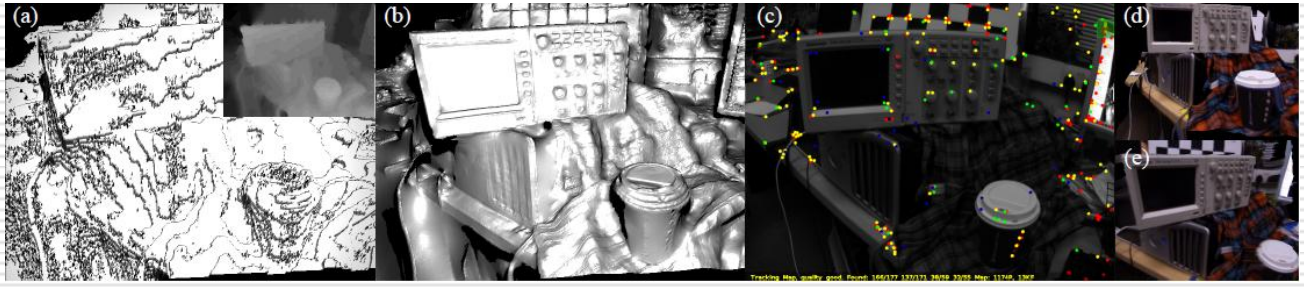


# Visual Tracking and Reconstruction

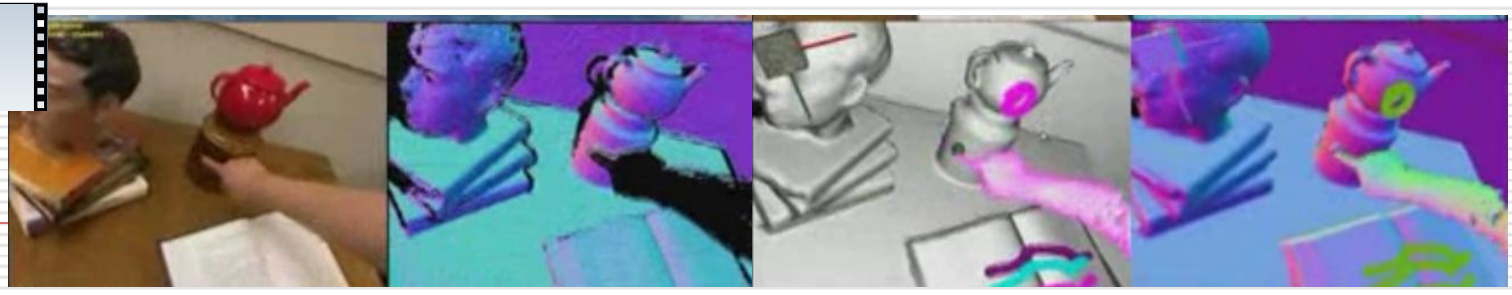
- Region-based
  - Bastian et al.
  - Umakatsu et al.




- Pixel-based
  - **DTAM**




- Real-time RGB-D camera
  - **KinectFusion**

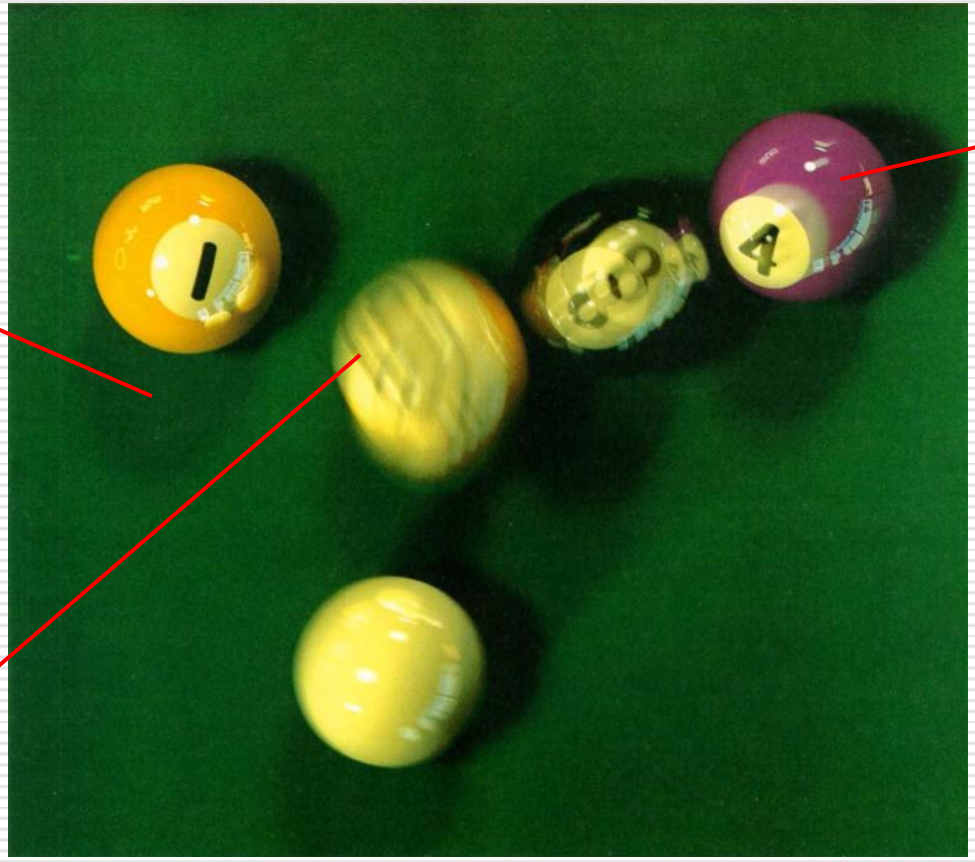


# Photometric Consistency

**How** should the information be rendered ?

# Photorealistic Rendering

- How to acquire the lighting condition of the real environment
- How to render photorealistic objects in real-time




(Pixar, 1984)

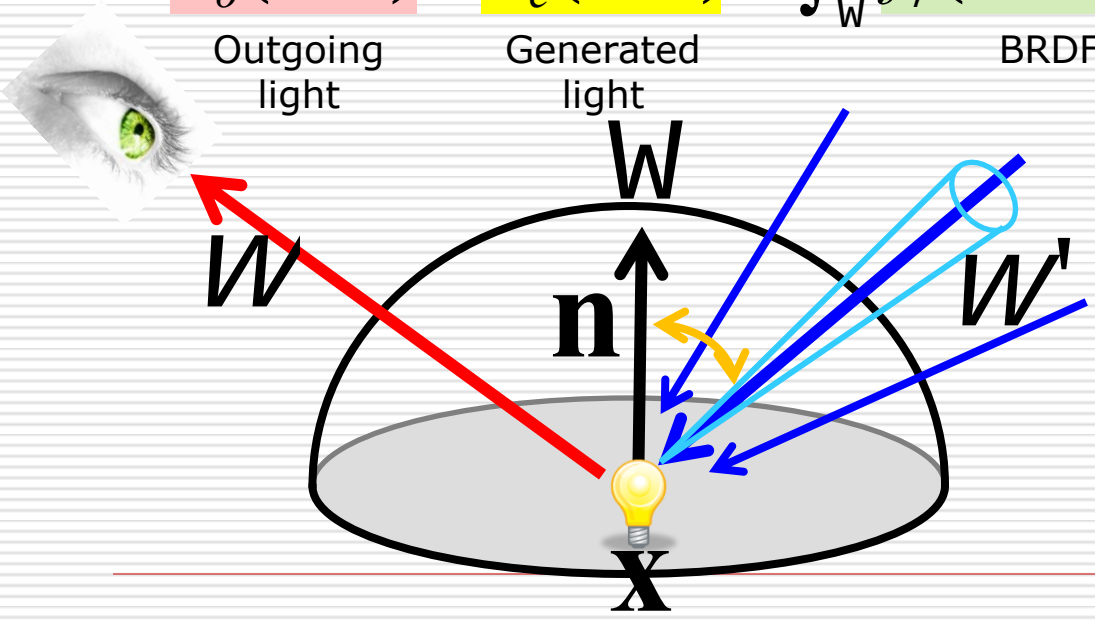
# Global Illumination

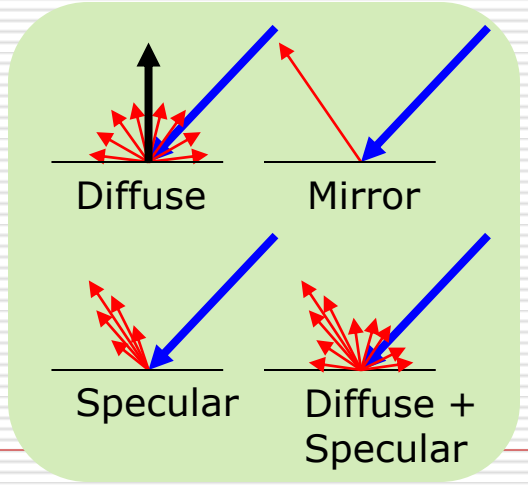
- Evaluating light reflected from a point ***x*** by taking into consideration all illumination that arrives at the point
- Rendering Equation
  - How to acquire a Bidirectional Reflectance Distribution Function (BRDF)
  - How to render objects with a BRDF
    - Use tons of random rays (Monte Carlo methods) e.g. Ray-tracing / Radiosity

Total reflected light

$$L_o(\mathbf{x}, w) = L_e(\mathbf{x}, w) + \int_W f_r(\mathbf{x}, w, w') L_i(\mathbf{x}, w') (w' \mathbf{n}) dw'$$

Outgoing light | Generated light | BRDF | Incoming light | Incident angle

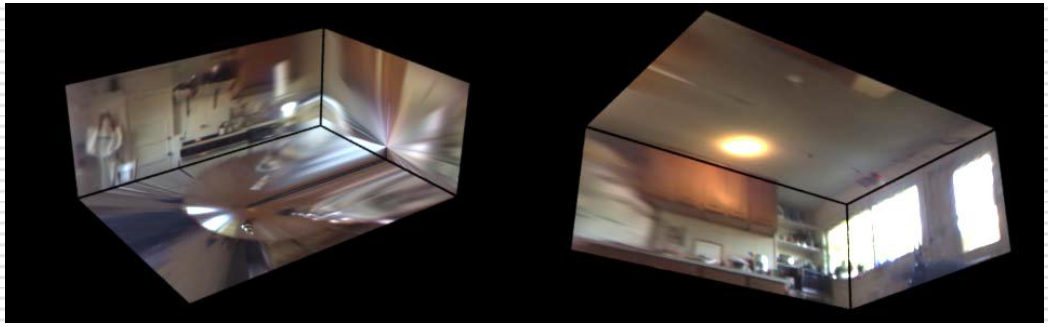


# Image-based Lighting with HDRI

- High Dynamic Range Imagery + Global Illumination (Debevec, '98)
  1) Acquire a radiance map by a light probe

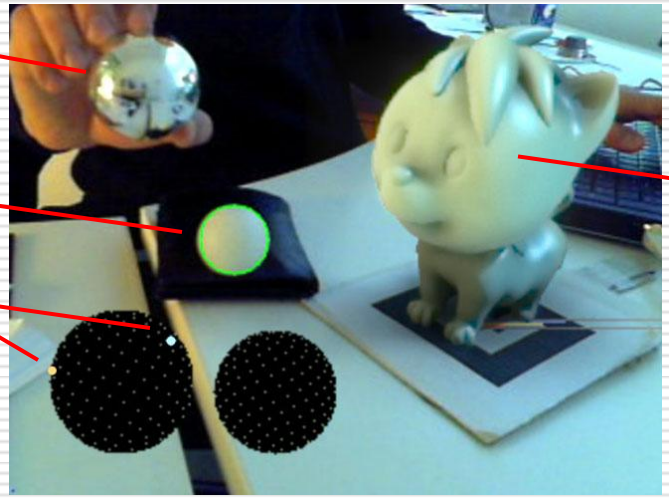
  2) Build a light-based model of the scene

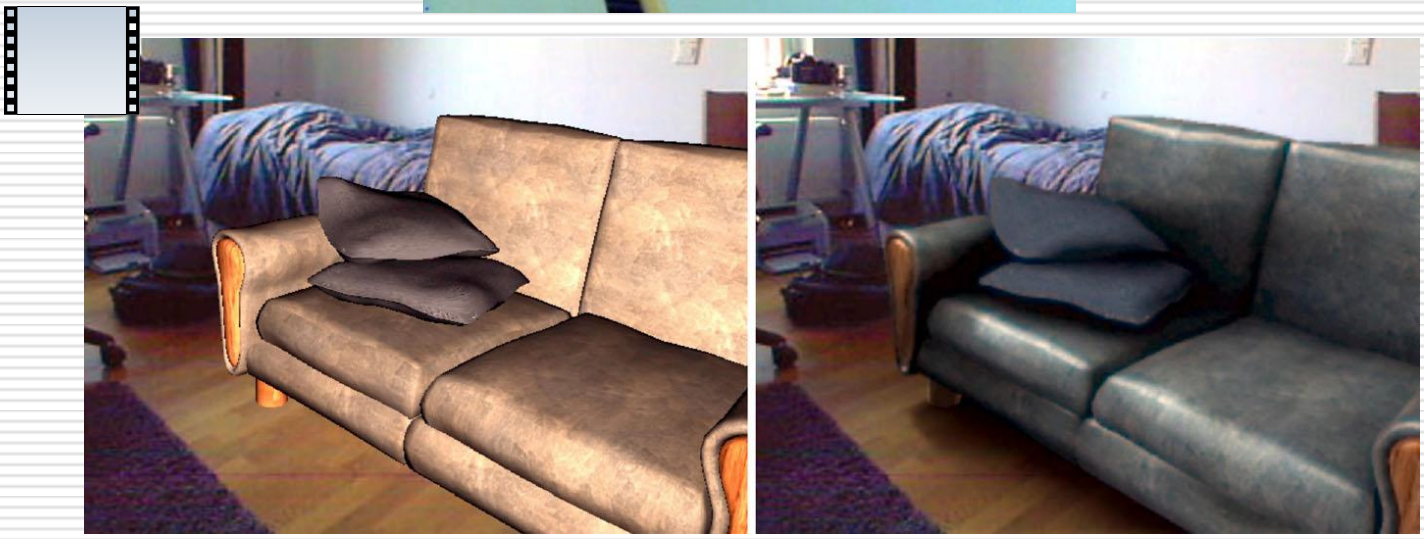


  3) Render virtual objects with global illumination

# Inverse Lighting with LDRI

- Inverse lighting and photorealistic rendering for AR (Aittala, '10)



Basic rendering | Photorealistic rendering

# Advanced Visual Effects for AR / MR

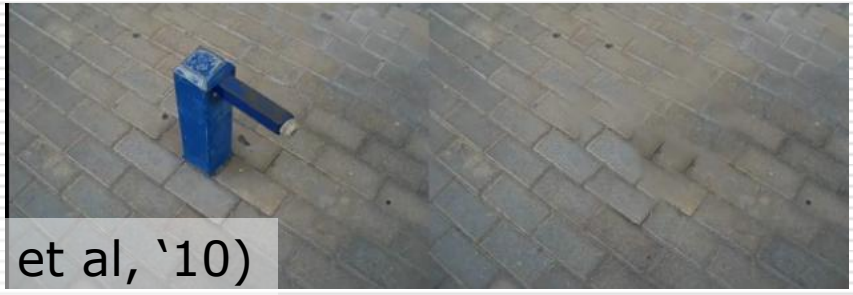
- Motion blur / Depth of field


(Park et al, '09)

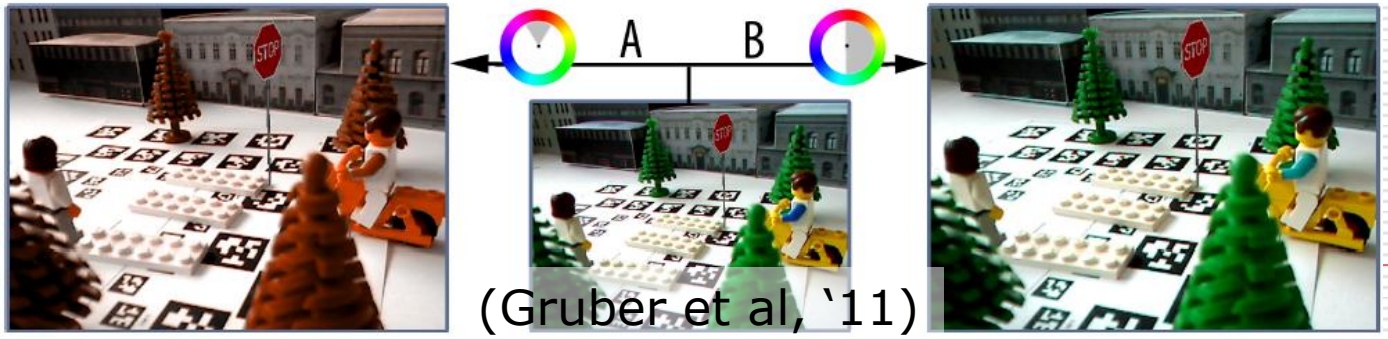

(Okumura et al, '06)

- Diminished Reality


(Herling et al, '10)



- Color harmonization



(Gruber et al, '11)

# Scene Understanding

**What** information should be presented ?

# Sushi Scenario

- How can we annotate each sushi piece ?

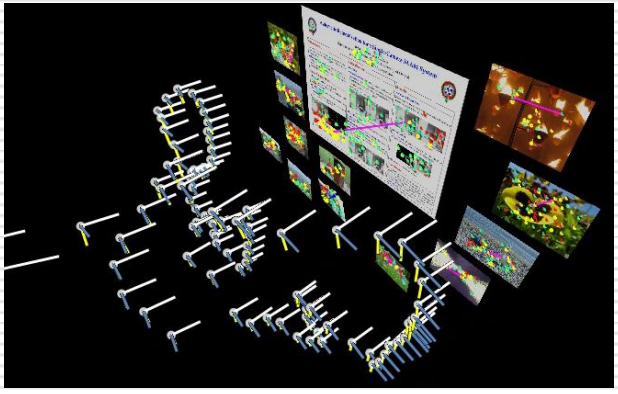
- Powerful object recognition technology is necessary to handle the real world problems
- Wavelet / SIFT / SURF / Haar-like / HOG / Shapelet …

# Real-time Object Recognition for AR

- Simultaneous Recognition, Localization and Mapping (Castle, '10)
  - Use a SIFT-based planar object database

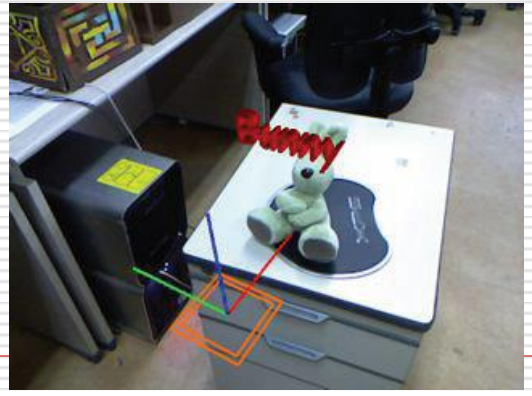

Recognized object names

Key-points



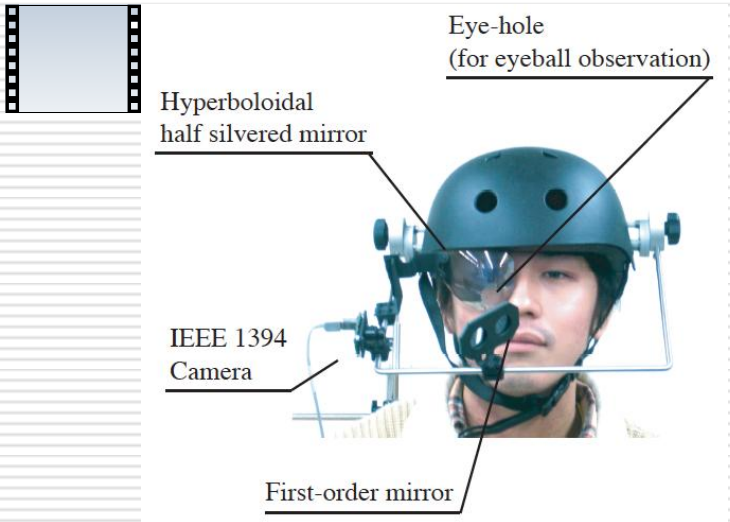
3D view of recovered objects

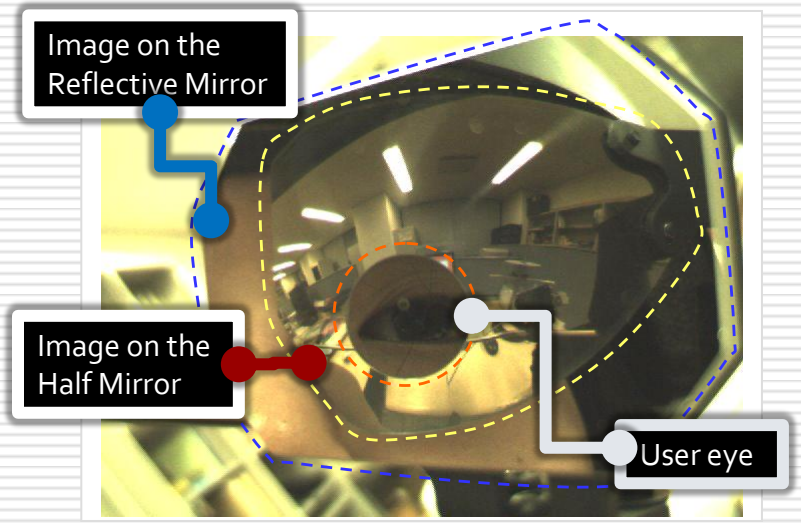
- Depth-assisted Real-time 3D Object Detection for AR (Lee et al., '11)
  - Use RGB+D information in template matching

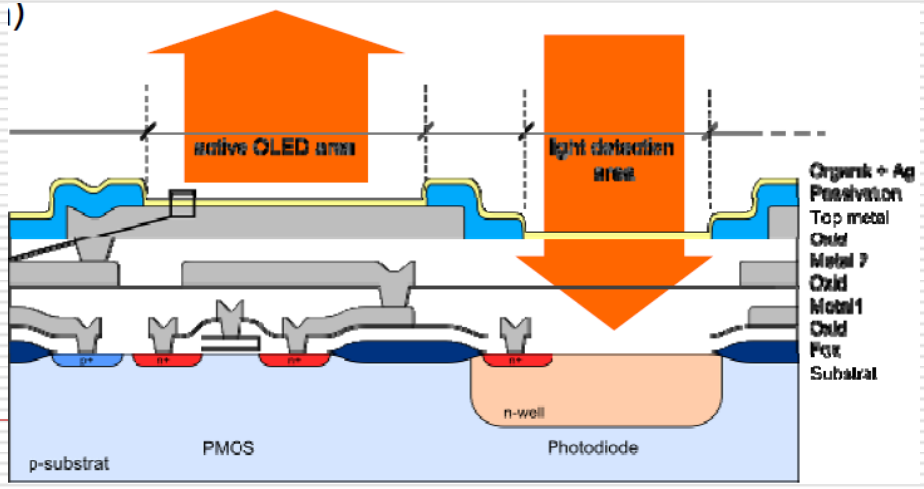

# Eye tracking for Context Awareness

- Parallax-free, super wide-view eye camera (Mori et al., '10)




- Eye tracking OLED display (Fraunhofer IPMS, '11)



# Conclusion

- **"Where" : Geometric Consistency**
  - Tracking and Reconstruction dramatically improved
  - Practical solutions at hand, though some problems remain
- **"How" : Photometric Consistency**
  - Photometric studies getting more and more attention
  - Many open issues remain
- **"What" : Scene Understanding**
  - Scene understanding / object recognition much studied in CV field, but still rare in AR / MR systems
  - Context-aware AR will be the core topic in near future

- Video analysis for AR / MR will be even hotter in future

# Thank you very much!

Any questions ?