Visual Interaction for Intelligent System

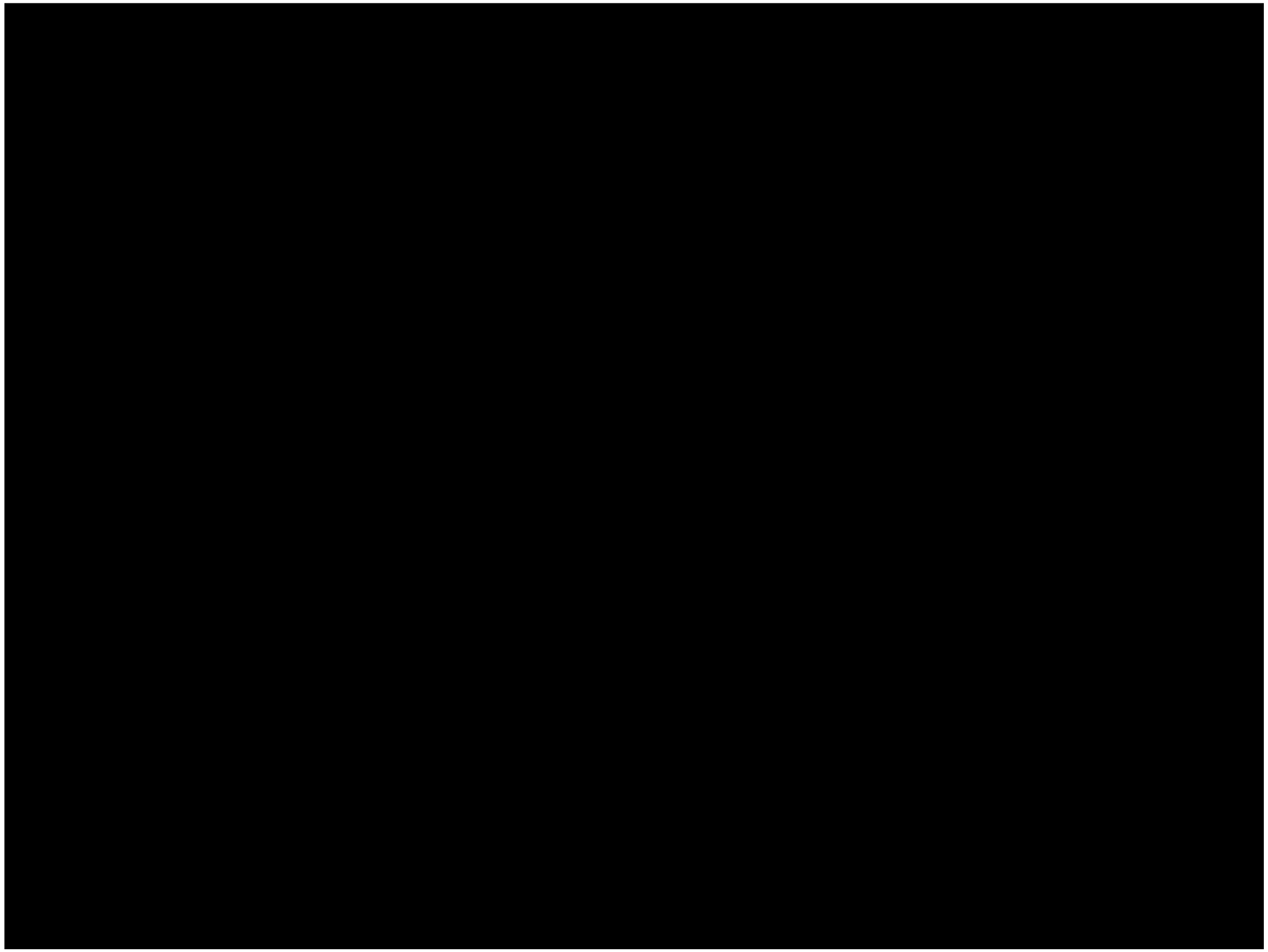
Lee, Chilwoo
School of Electronics and Computer Engineering,
Chonnam National University, Kwangju, Korea

Chonnam National University in Gwangju

CNU: Chonnam National University

- Located in Kwangju(光州)
- Kwangju is the city of Light & Art
  - 200 miles south from Seoul
  - http://image.jnu.ac.kr
  - http://mdirc.jnu.ac.kr
  - http://ct.jnu.ac.kr

What I did..?



Interaction and Real World

The Real World

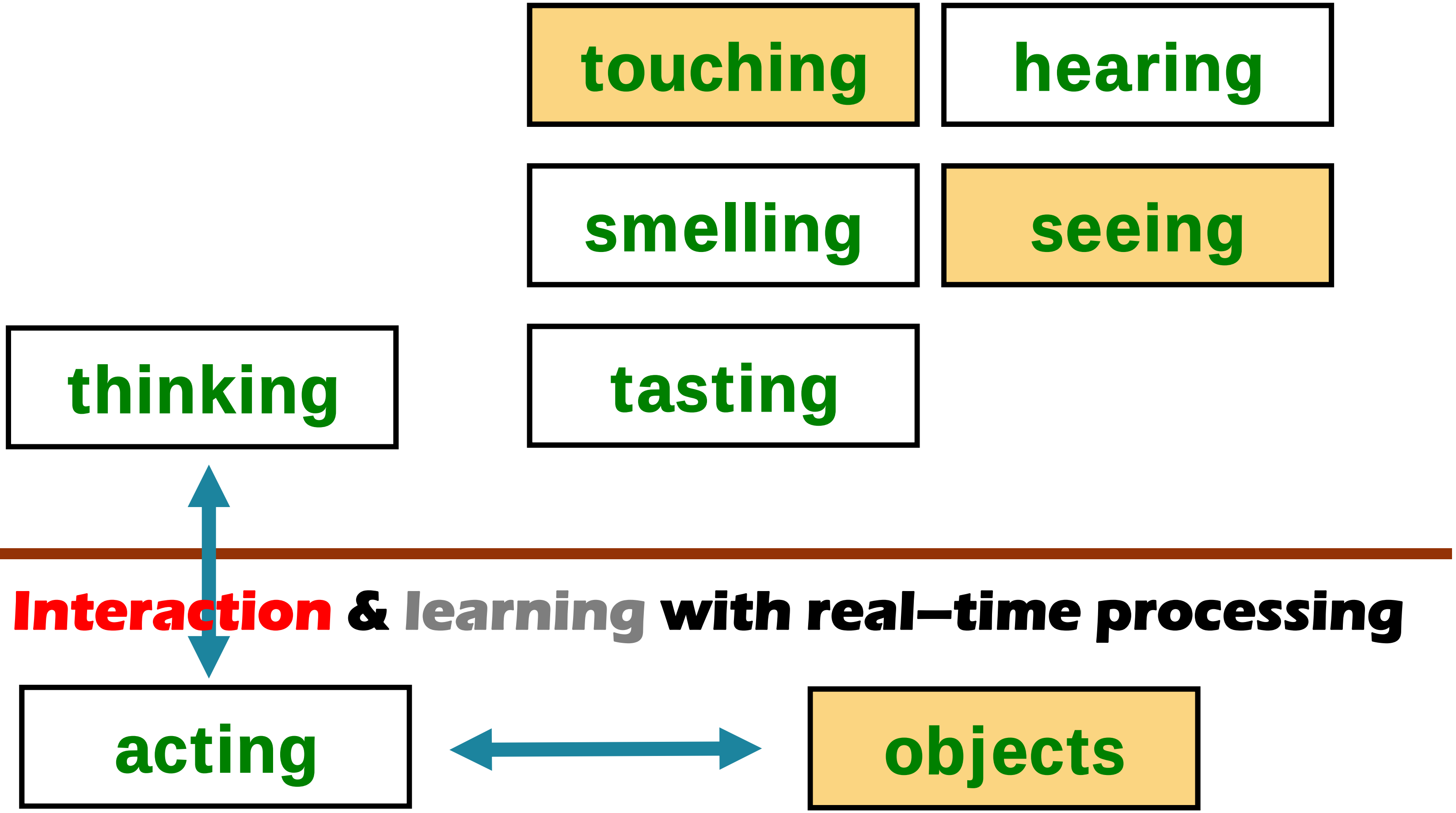
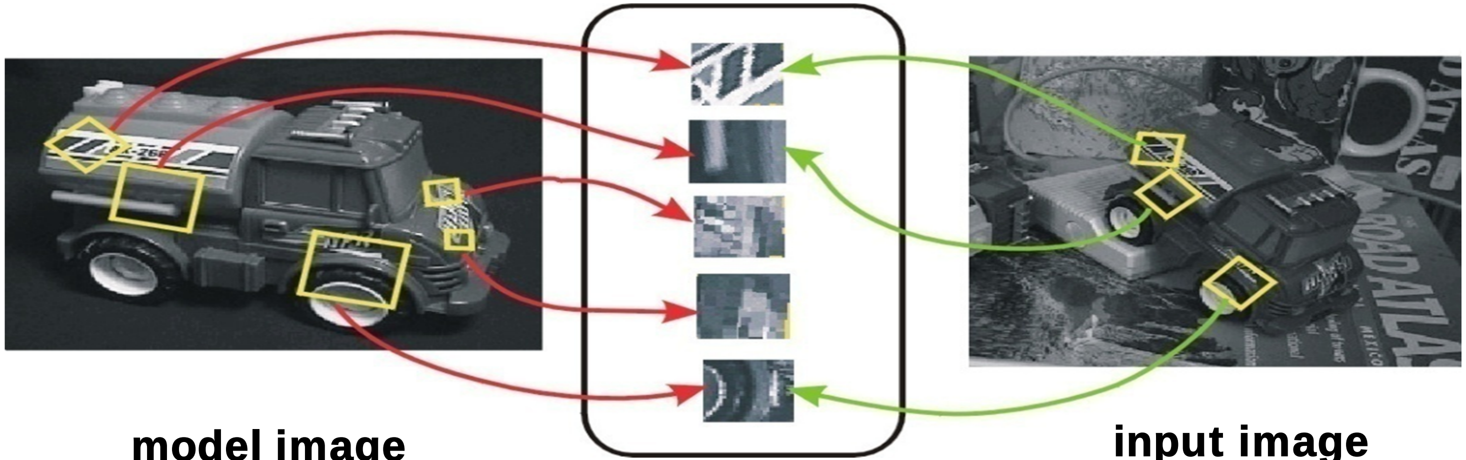


Image Recognition: feature based



model image
← training

feature space

input image
→matching

1988	**Feature Detection**	Harris, Stephens: Harris Corner Detector
1994	Scale Invariant	Lindeberg: Scale-space theory
1997	Rotation Invariant	Schmid, Mohr: Local grayvalue invariant
1999		Lowe: SIFT
2004	Robustness Reinforcement	Ke, Sukthankar: PCA-SIFT
2005	Background Deletion	Stein, Herbert: BSIFT


gachi

Problems

- Human body is a very Complex 3D articulated object with High DOF
  - non-dominant features
  - occlusion among limbs

- Human body is a freely moving object with various clothes
  - high speedy motion
  - segmentation difficulty

How can we extract the features?

Content

- Gesture recognition algorithms and systems
  - Definition for related terminologies
  - Template based gesture recognition: Interactive Gesture Game
  - APM(Active Plane Model) pose recognition
  - Articulated Pose Estimation with Pictorial Structure
- Tabletop Display and Interactive Content
- Conclusion

Interaction: definition

- A kind of action that occurs as two or more objects have an effect upon one another. The idea of a two-way effect is essential in the concept of interaction, as opposed to a one-way causal effect.

- Control
- Operation/Manipulation
- Synchronization
- Indication
- Selection/Mining/Retrieval
- Stimulation/Encourage
- Communication

Gesture-like Key Terms: comparison

- Behavior/Behavior(행위/행동/습성)
  - unconscious/conscious actions or reactions of an object or organism, usually in relation to the environment
- Conduct(행위/품행)
  - personal behavior; a way of costumed acting
  - using hand gesture to direct
- Action(행동/동작)
  - Human bodily movement in intention
- Motion(동작/운동/몸짓/움직임)
- Gesture(제스처)
  - bodily motion which has some meanings for non-verbal communication: body language

Gesture Recognition

- **Gesture:**
  expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body with intent of
  - conveying meaningful information
  - interacting with the environment

- **Importance of gesture recognition**
  - non-linguistic information with gesture for human-like communication
  - more natural human-computer interaction
  - gesture recognition is a core technology for various applications

Template Based Gesture Recognition: feature & modeling

(a) Left arm maskable template models.

(b) Detailed left arm maskable template models.

gesture	pose		code	operation
	Arms(L)	Legs(L)		
	4<N<6	–	10	punch
	–	3≤N≤5	11	kick
	5.5≤N≤8 2≤N≤4.5	–	1	special(1)
	8<N≤9	–	3	special(2)

Table 1: Sample of operation code table for fighting game

Game Action : recognition

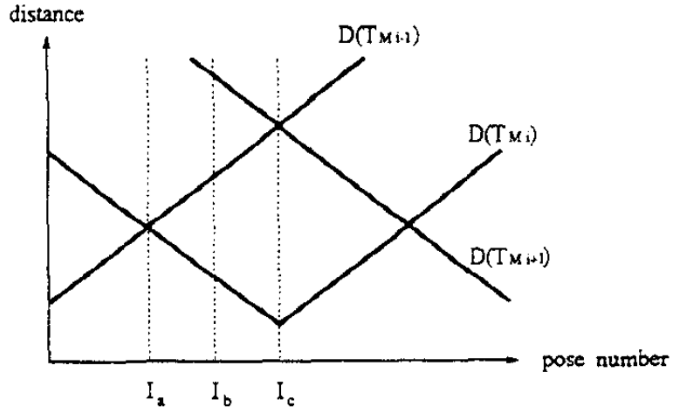
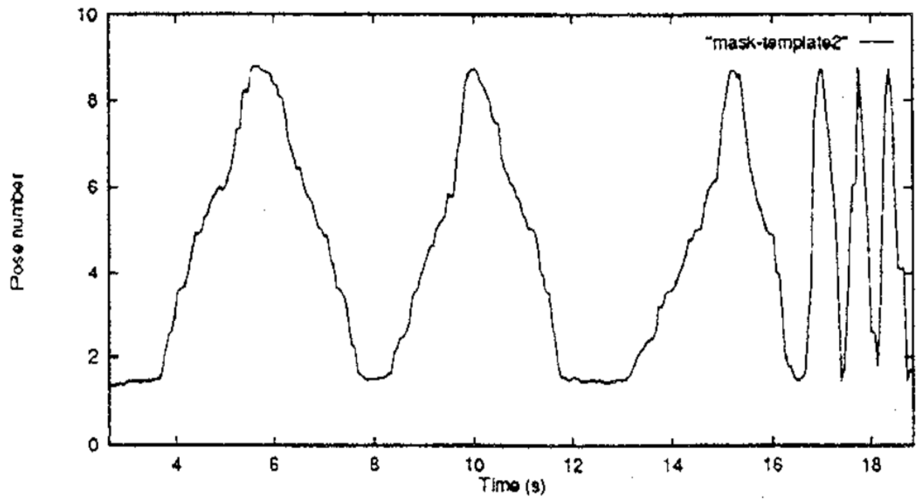
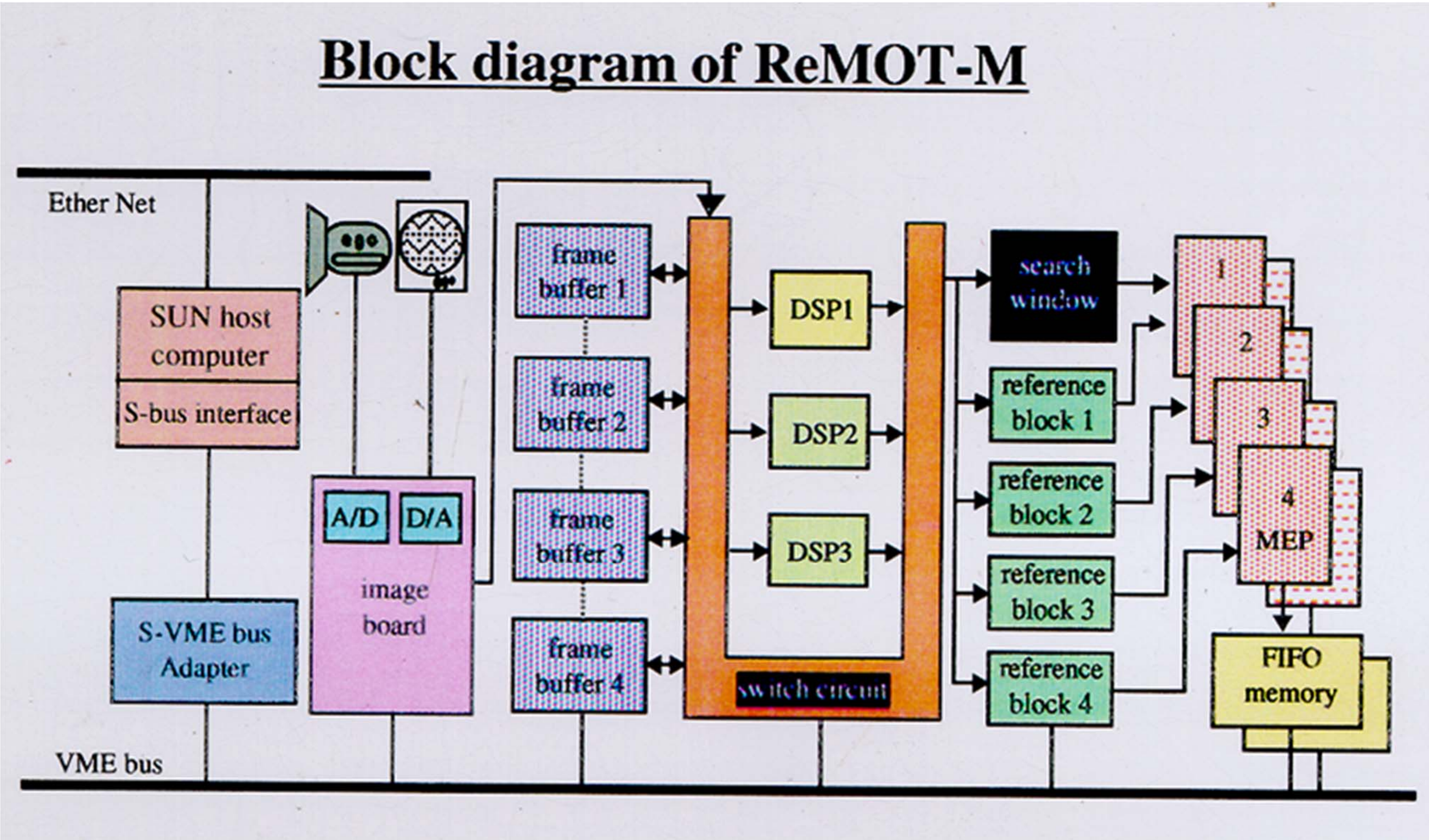


Fig.4: Pose interpolation. Estimation of pose of a
is done by comparing distance between a minimum
tance model and two neighboring models



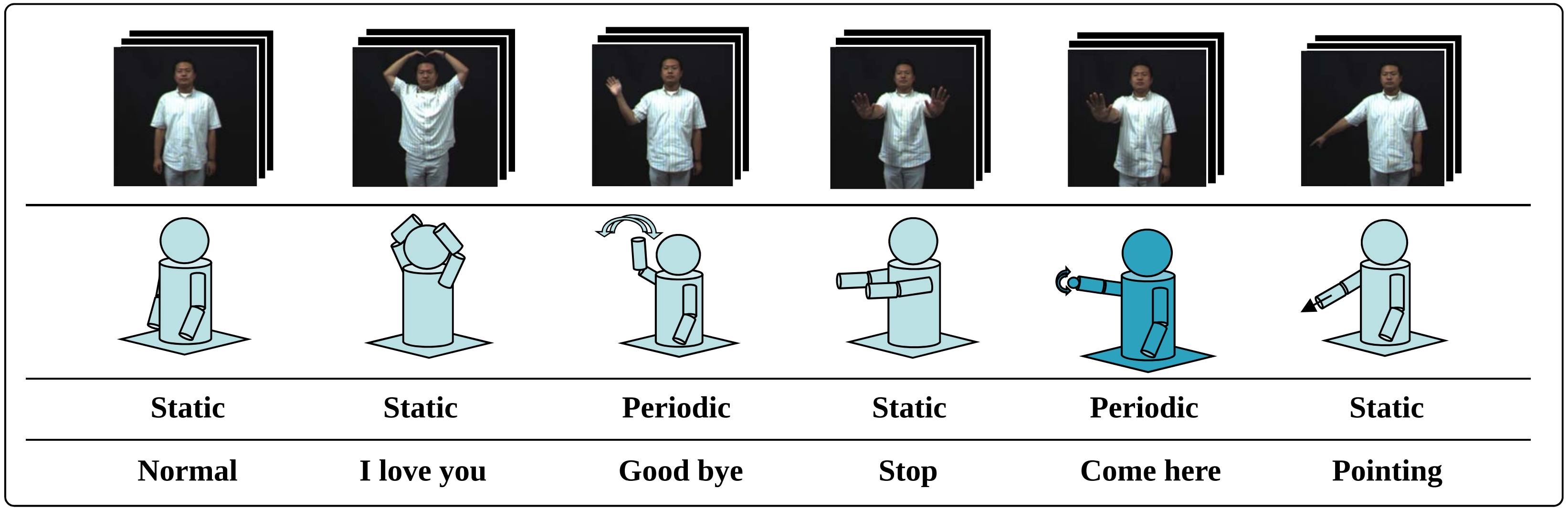
(c) Result of arm pose estimation using template(b) in Fig.3

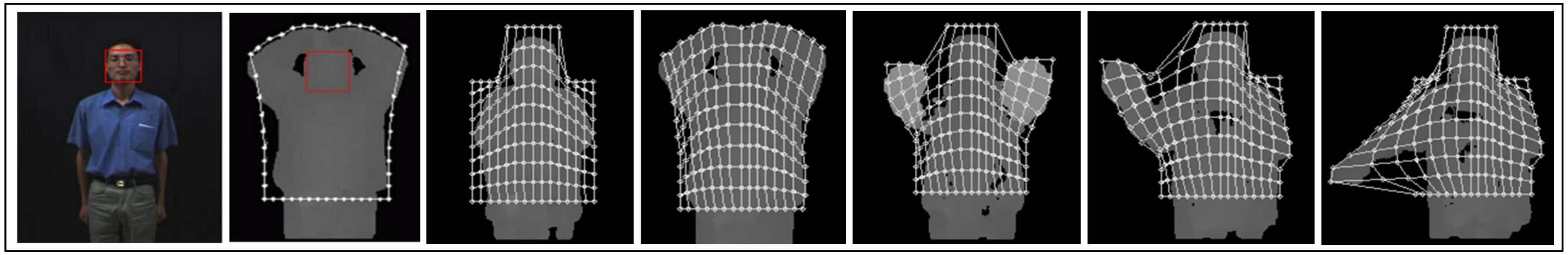
Real-time H/w System



Gesture Game : demo

Humanoid Control: feature & modeling

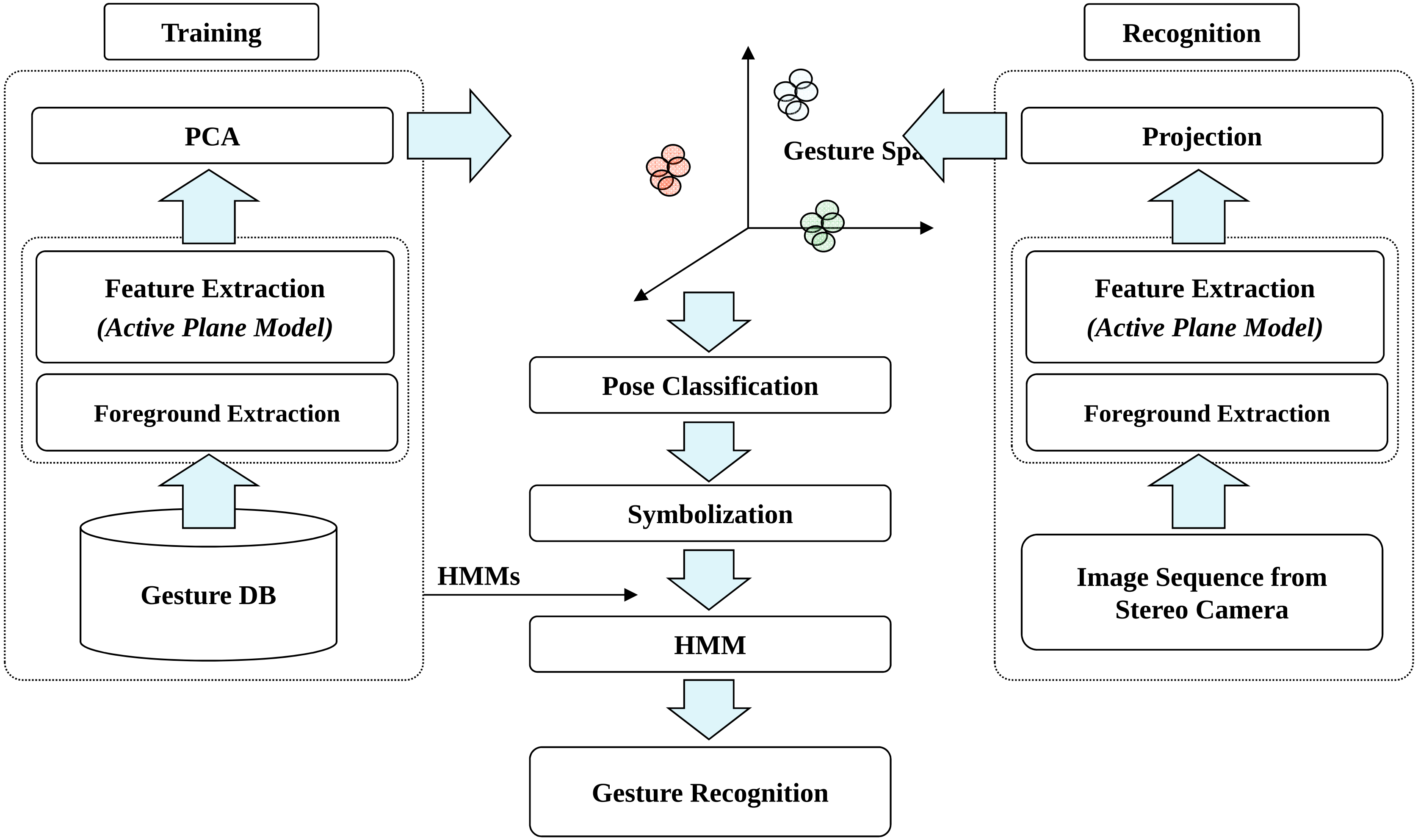




Deformed APM(Active Plane Model) Examples for several gestures

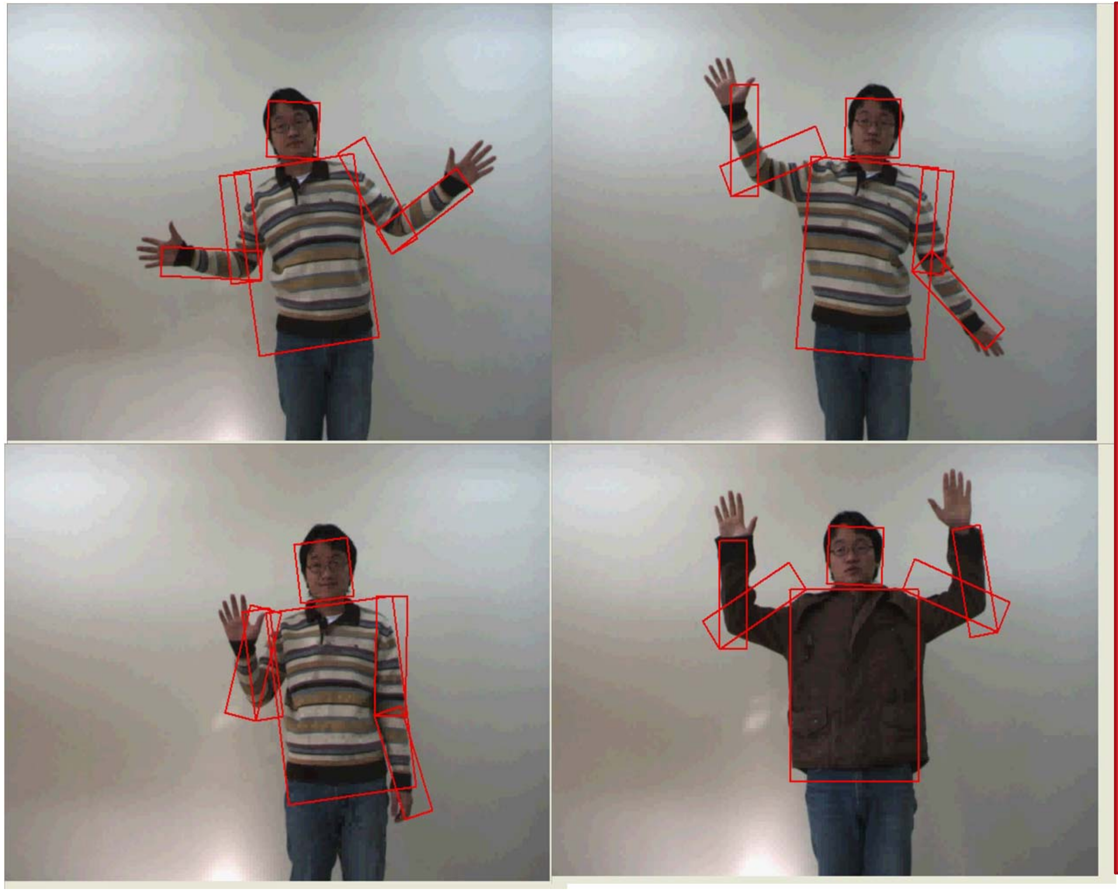
Humanoid Control: process

DEMO




Arirang
KOREA
Gesture Recognition By Attention Control Method
for Intelligent Humanoid Robot
Chonnam National University
Intelligent Image Media Lab.

Upper Body Pose Estimation



Human Computer Interaction

Human Robot Interaction

Problems

- Articulated Pose Estimation is one of the difficult issues in Computer Vision
  - Many problems arise due to the human body **deformability**.
  - The possibility of **Occlusion** among limbs.
  - **Illumination change**
  - **Computational cost** is often too expensive.
  - **Automatic initialization** is dependent on environment.

Deformation

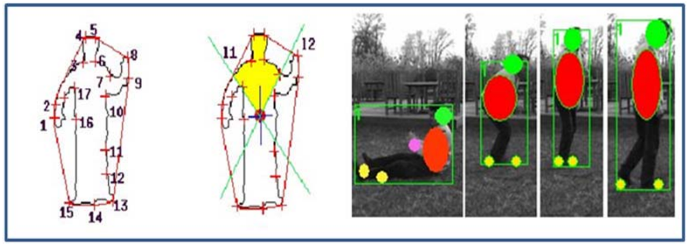
Occultation

$$\mathrm{nHypotheses} = \mathrm{l(w)l(h)} \prod \mathrm{l(p_i)}$$

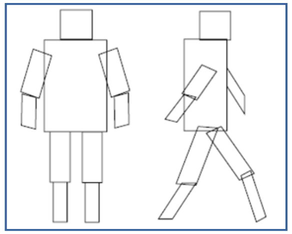
1.5 million (w320,h240)

Computational Cost

Related Works



- Pose Modeling
  - Appearance Model
    - Heuristic model [Haritaoglu,ICPR98]
      It is real-time but needs a precise background subtraction.



  - Geometric(Graphical) Model
    - 2D Cardboard model [Ju,ICAFGR96]
      Easy to model but can't express whole pose space.

    - 3D Volumetric model [Demirdjian,ICCV05]
      Express all poses but no real-time and difficult to be initialized and manipulated

Related Works

- Pictorial Structures

  [Felzenszwalb, Pictorial structures for object recognition, IJCV05]

- It is a kind of 2D cardboard model
- It is a framework how to model deformable object and to start tracking objects effectively

M. Andriluka et al. "Pictorial Structures Revisited: People detection and Articulated Pose Estimation" CVPR09

- A generic approach based on the pictorial structures framework
- The Adaboost classifiers are trained by appearance of body parts
- BP(Belief Propagation) to infer the posterior of each part
  - Limitation
    - BP is very slow
    - Pose estimation is dependent on part detection

Related Works

- Pose Estimation
  - **Sampling : Particle Filtering** [Deutscher, CVPR00]

    It is robust in high speed of motion.

    But usually needs large number of particles.

  - **Belief propagation by Brutal force search** [Pearl, CVPR03]

    Very powerful algorithm.

    But it needs huge computational cost of CPU time.

  - **Non-parametric belief propagation** [Sigal, CVPR04]

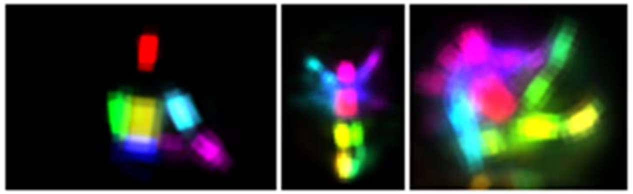
    Improved version of BP.

    But it needs very slow Gibbs sampler.

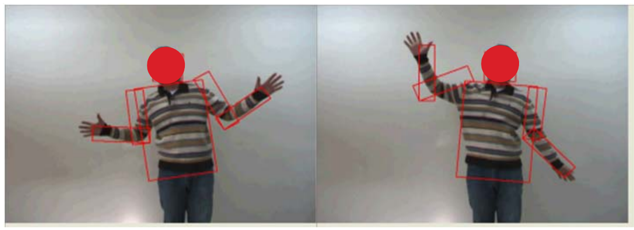
Related Works

- Initial Positioning
  - Automatic Initialization for starting pose estimation
    - Pedestrian Detector, Part Detector or Head(Face) Detector is used to set the initial configuration of a defined pose model.

Pedestrian Detector



Part Detector



Face Detector

Our Approaches

- Pose Modeling with Pictorial Structures
  - **Easy to use**
    Heuristic knowledge for hierarchical appearance of upper body is modeled by Pictorial Structure
  - **Real-time**
    It could be used in real-time because it has reasonably enough small size of Pose Space.
  - It has enough deformability of upper body motion.
- Pose Estimation with Sampling (Particle Filtering)
  - **Free from textures and colors**
    We use Chamfer matching which is robust in various environment.
  - **Independent of trained data**
    All process are not related with the learning result.
  - **Fast sampling**
    Hierarchical Key Pose is effective in sampling calculation

Pictorial Structured Filtering

- General Particle Filter [Yang, ICCV05]
  - A Bayesian sequential importance sampling technique which recursively approximates the posterior distribution using a finite set of weighted samples.
  - Consists of two steps : Prediction, Update.

Prediction

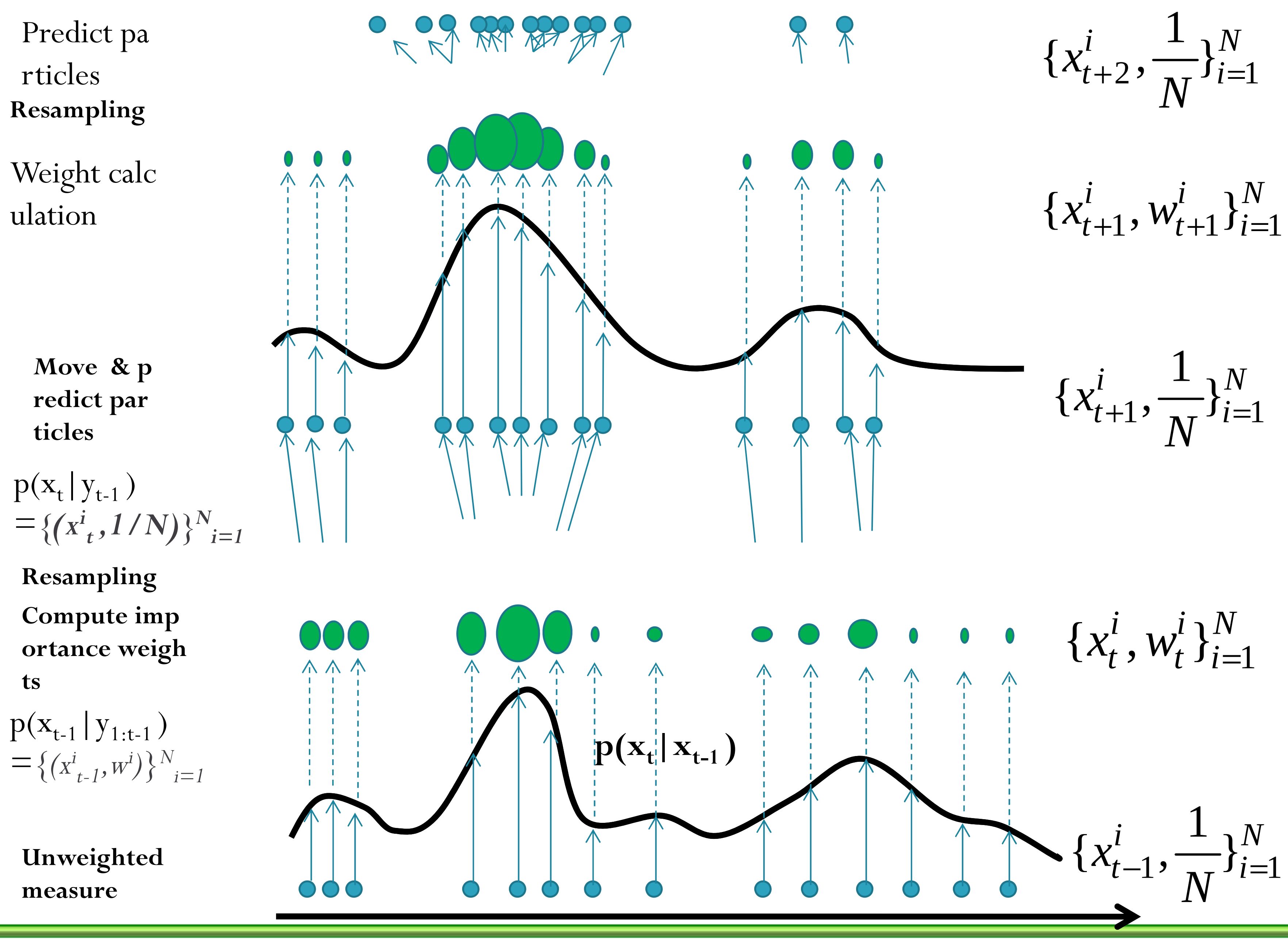
The probabilistic transition model $p(x_t|x_{t-1})$ is used to predict the posterior at time t.

$$p(x_t|z_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1})dx_{t-1}$$

Update

At time t , the state can be updated with the observation $p(z_t|x_t)$.

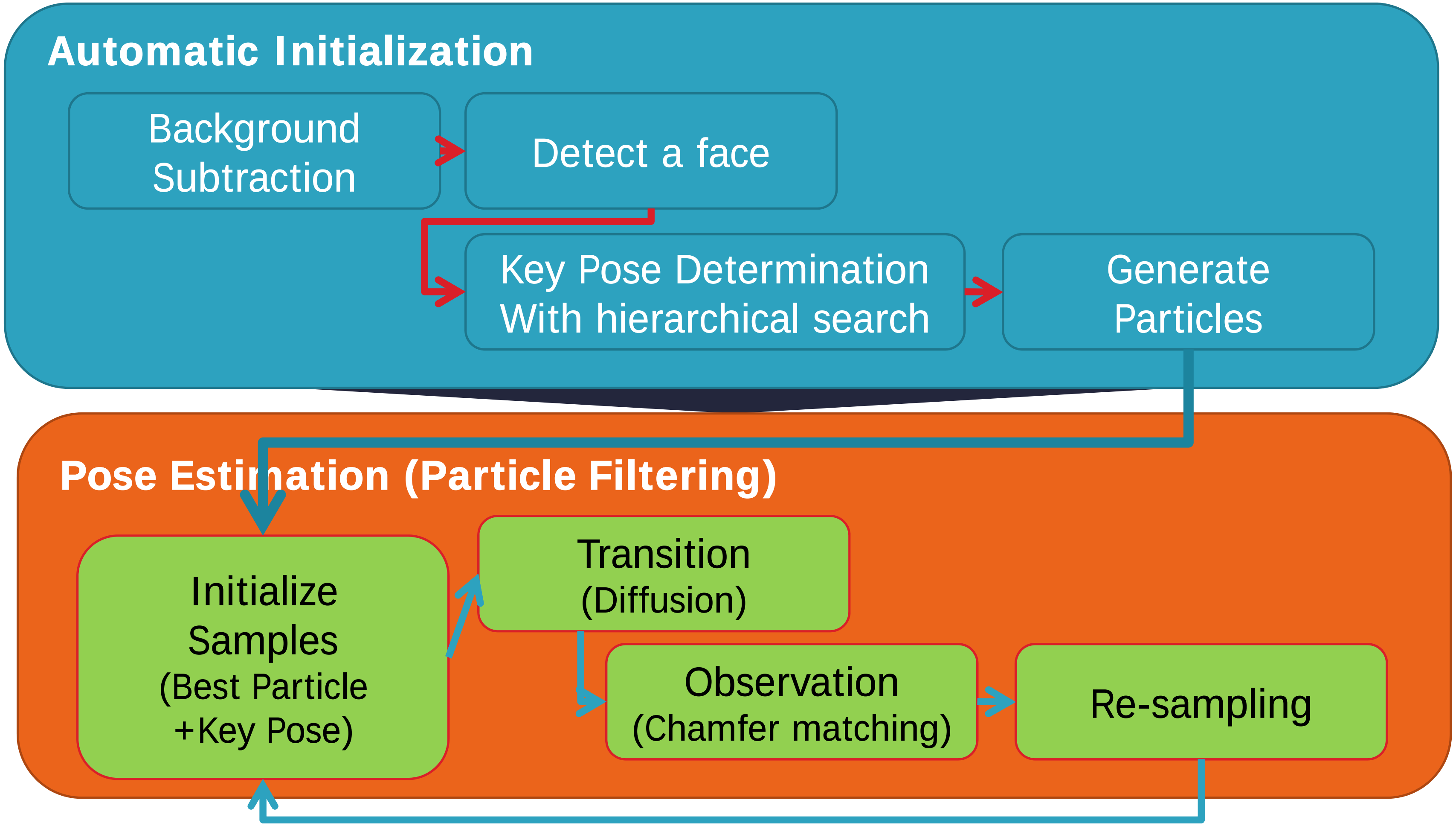
$$p(x_t|z_{1:t}) = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})}$$



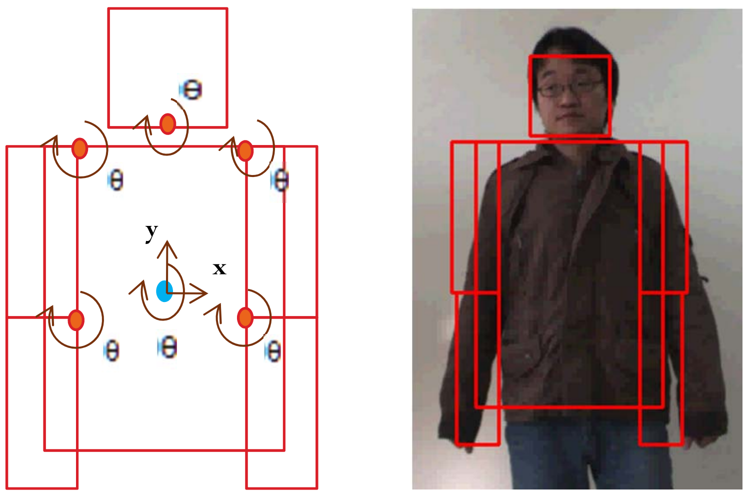
$i = 1, 2, \ldots, N = 14$ Particles

$w^i = p(y_t \mid x^i_t)\,{}^N_{i=1}$

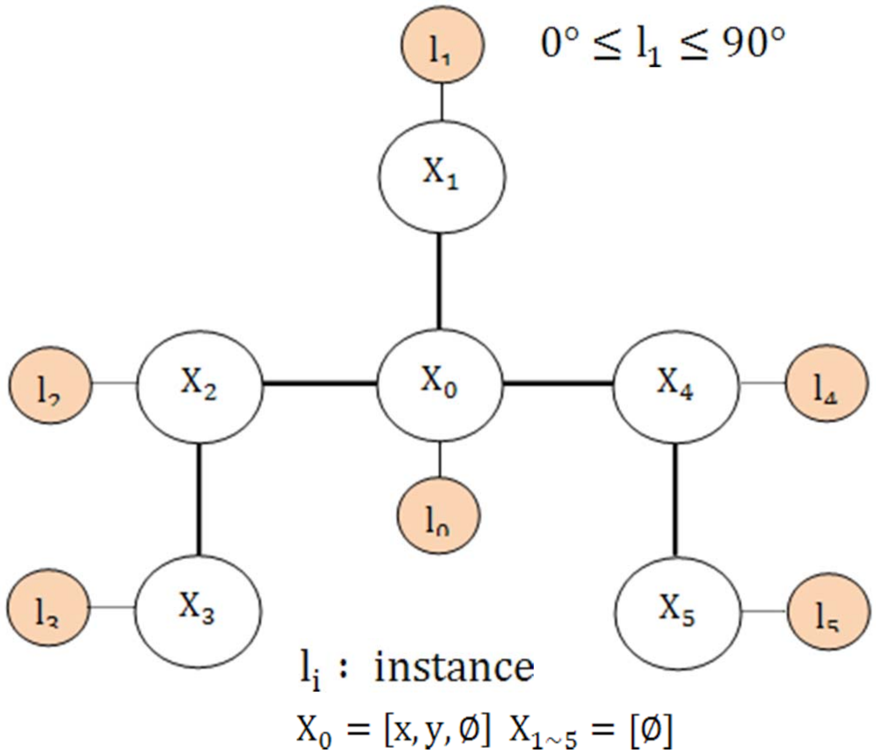
Procedure for Pose Estimation



Upper Body Modeling with Pictorial Structures

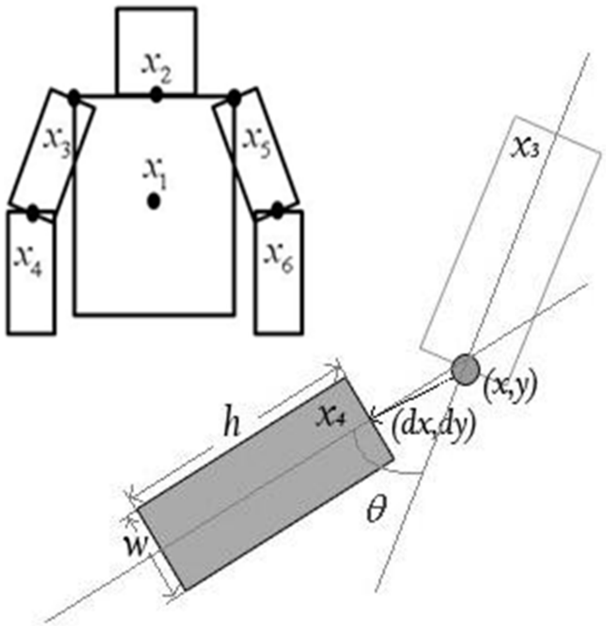


Pictorial Structures



Graphical Model

Pictorial Structured Model



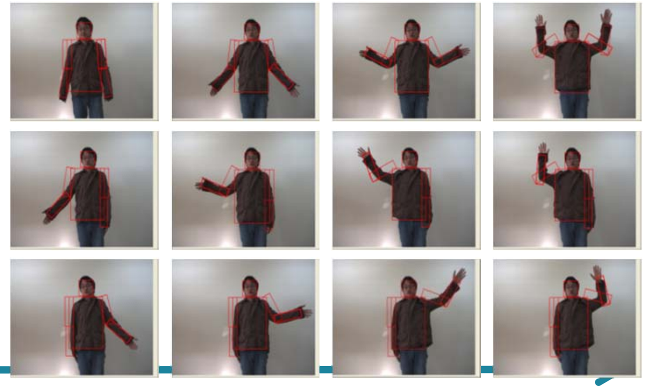
$$X_i = \{x, y, dx, dy, w, h\}$$

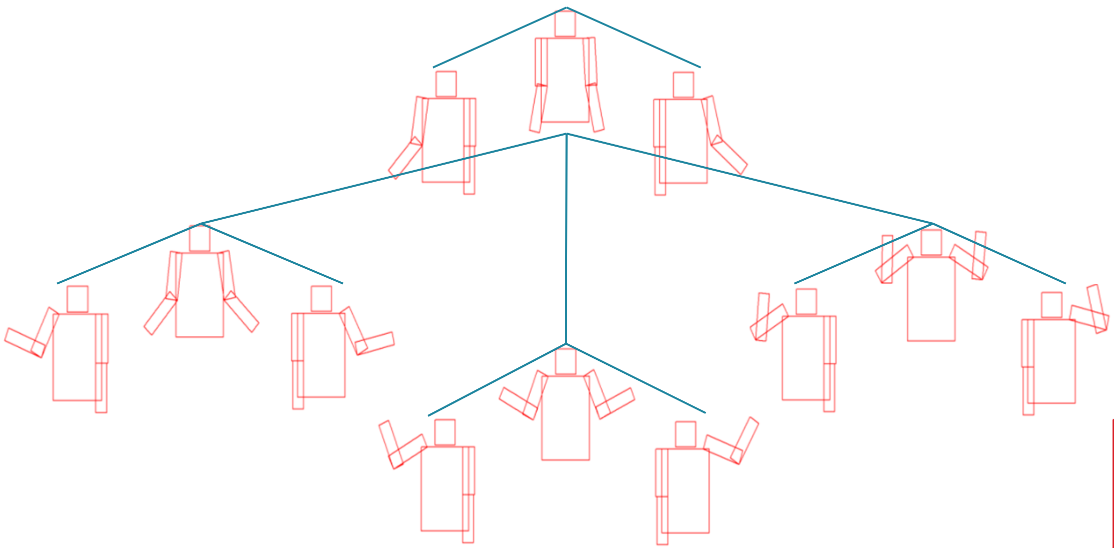
① ② ③

① Part location given by its parents
② Spring-like displacement
③ Rectangular size of part

Automatic Initialization

Key Pose Determination



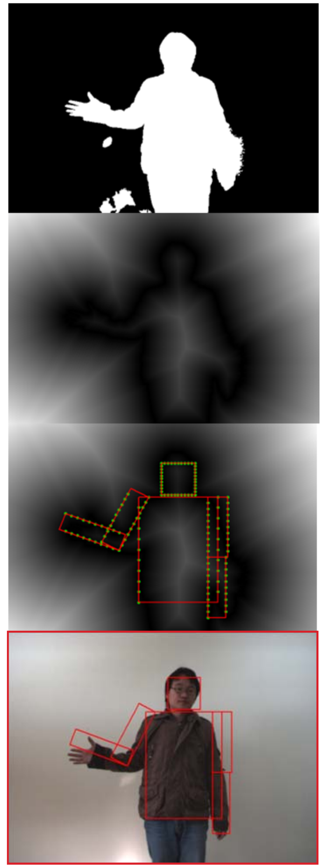


Input Image

Distance Transform

Chamfer Distance

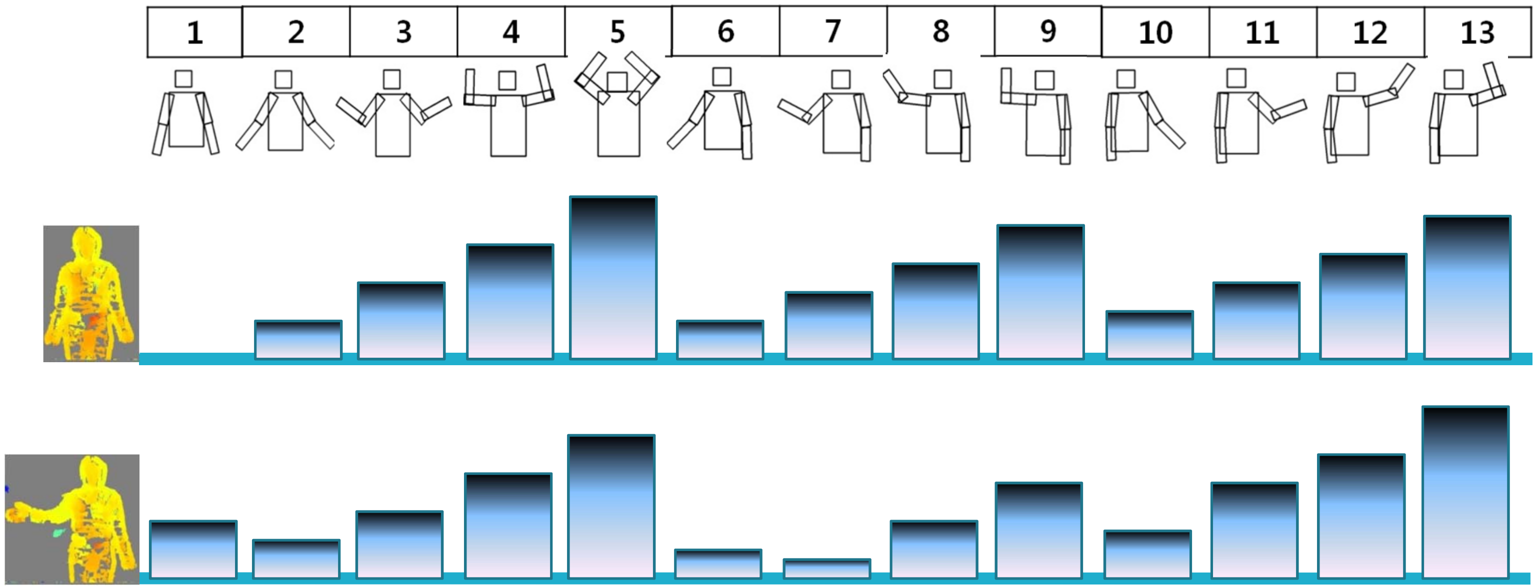
Key Pose



The Hierarchical Structure of Key Pose Space

Key Pose Library

- Building KPL.
  - Two exemplar visual features



Automatic Initialization

Key Pose Determination

- Training
  - Feature vector of a key pose

$$f = [d_i]^T \quad 1 \le i \le 12 \qquad d_i = m(kp_i, kp\_edge_i)$$

Where m() : chamfer distance

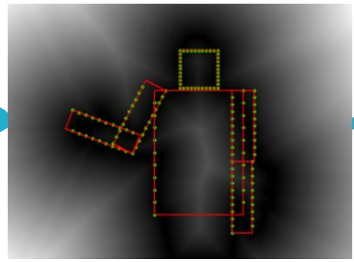
kp : key pose

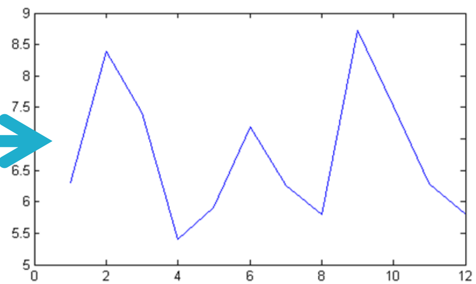
kp_edge : edge image of key pose image

The Hierarchical 12 Key Poses

d_i

$1 \le i \le 12$





Feature Vector f

Automatic Initialization

Key Pose Determination

- Key Pose Detection
  - Feature vector for Input Silhouette Image

$$q = [d_i]^T \quad 1 \le i \le 12 \qquad d_i = m(kp_i, \mathrm{Silouette\_edge})$$

  - Determine Key Pose with Diffusion Distance

$$kp = \arg\min dd(q, f_i)$$

Where dd() : Diffusion Distance

$$dd(h_1, h_2) = T(h_1) - T(h_2) \qquad T(h) = \int_1^{12} \prod_1^{t} g * h \, dx$$

Where g() : Gaussian Function

Automatic Initialization

Initialize
Particle Filter

- Particle (Sample) $s_j = [X_{i,t},\ X_{i,t-1},\ X_{i,t-2}, w]^T, \quad 0 \leq i \leq 5$
- Key Pose $kp_t = [l_{i,t}]^T, \quad 0 \leq i \leq 5$
- Initialization (for all particles) $X_{i,t} = l_{i,t},\ X_{i,t-1} = l_{i,t},\ X_{i,t-2} = l_{i,t} \quad , \quad 0 \leq i \leq 5$
- Transition (in pose estimation process)

$$X_{t+1} = A(X_{t-1} - X_t) + B(X_{t-2} - X_t) + C\ \mathrm{GaussRand}(\quad) + X_t$$

Similarity Measurement

- Best Configuration of Pictorial Structures

$$C^* = \operatorname{argmin}\left(\sum_{i=1}^{n} m_i(l_{i,t}) + \sum_{(v_i,v_j)} d_{i,j}(l_{i,t}, l_{j,t})\right)$$

Where m() : Part Matching Score
(Chamfer Distance)

d() : Link Score
(Euclidian Distance)

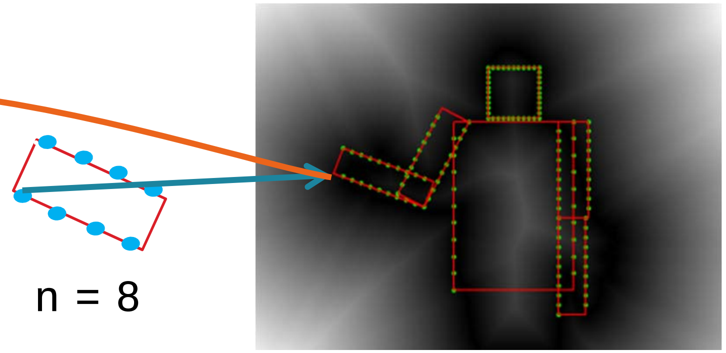
Similarity Measurement

Part Matching Score

$$m_i(l_{i,t}) = \frac{1}{n \times \max(\text{dist})} \sum_{(x,y) \in l_{i,t}} \text{dist}(x,y)$$

2	0	1	2	1	0	0	1

Chamfer Distances of Part edges



n = 8

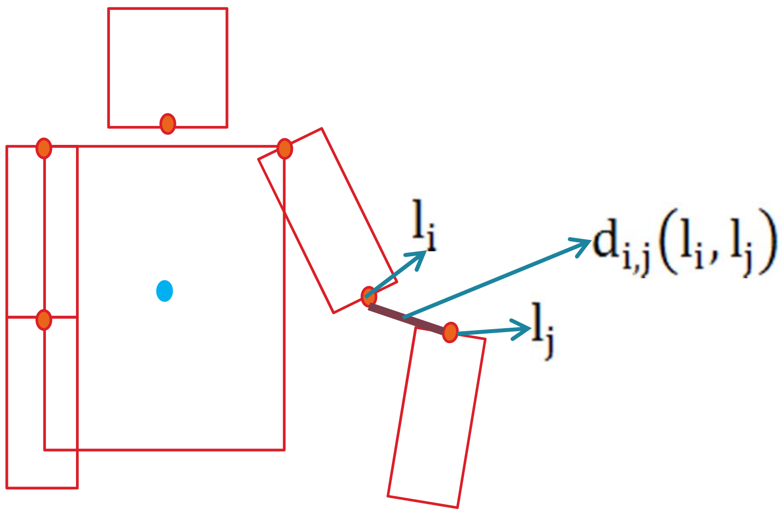
Chamfer Matching

Similarity Measurement

Link Score

$$d_{i,j}(l_i, l_j) = e^{\frac{\dot{d}(l_i,l_j)^2}{2\sigma^2}}$$

Where $\dot{d}(l_i, l_j)$: Euclidian distance of the link



Pictorial Structured Filtering

- Our Particle Filter
  - General Particle Filter is used in our tracking frame work with our sampling technique.
  - Consists of three steps : Sampling, Prediction and Update

    - **Sampling Number**
      - Particle Set (at the start point, t=0)

        [100%] = [Key Poses]

      - Particle Set (in the processing time, t>0)

        [n1 : n2 : n3] = [Best Weighted Particle
        : Best 5 Particles
        : Key Poses]

        # of Particle (N) = n1+n2+n2 = 2,000

Pictorial Structured Filtering

- **Sampling with Key Poses**
  - The diffusion distances of key poses are transformed as weights.
  - Weights are used to calculate how many particles of each key pose are need to be sampled.

- Normalized diffusion distances

$$d_i = dd(q, f_i) \quad , \quad 1 \le i \le 12 \qquad\qquad n\_d_i = \frac{d_i}{\sum d_i}$$

- Normalized weights

$$w_i = 1.0 - n\_d_i \qquad\qquad n\_w_i = \frac{w_i}{\sum w_i}$$

- Sample numbers of key poses

$$n_i = w_i \times N$$

where N : The number of particles

Algorithm Overview


Input Image


Silhouette


Edge

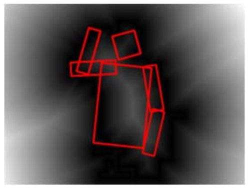

Face detection


Update
Key Pose Library
(KPL)

Tracking

(1) Prediction
1^{st} Order Markov Chain and KPL based-Proposal distribution

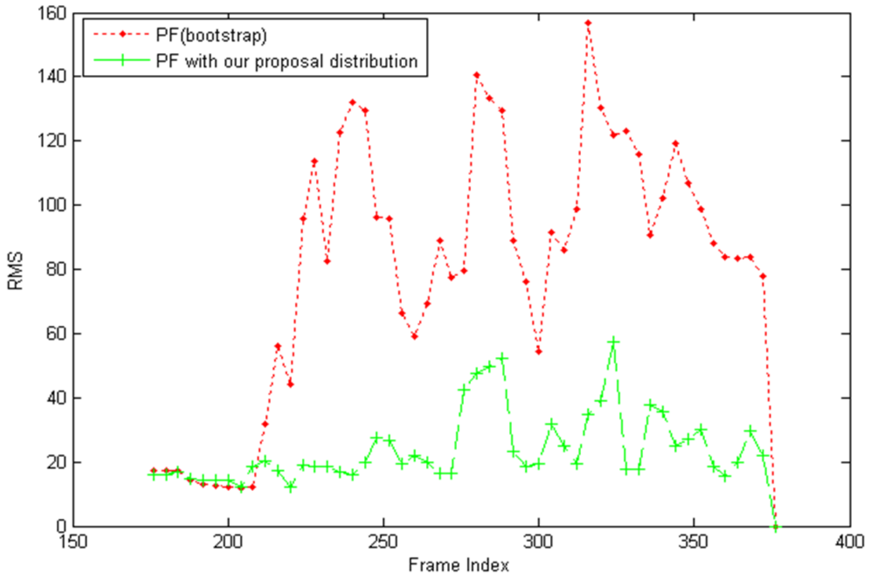
(2) Update
Edge & Silhouette likelihood

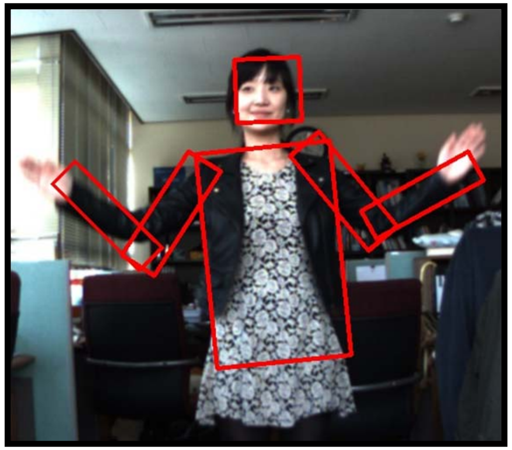


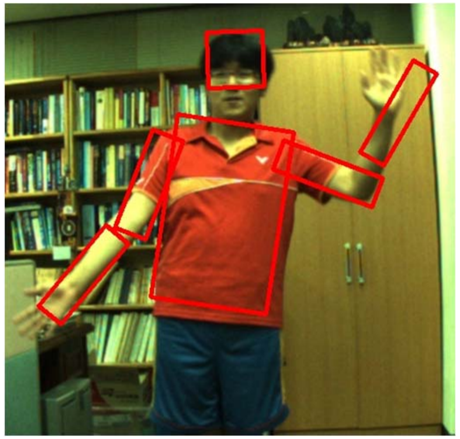

Estimated Upper Body Pose at time step t

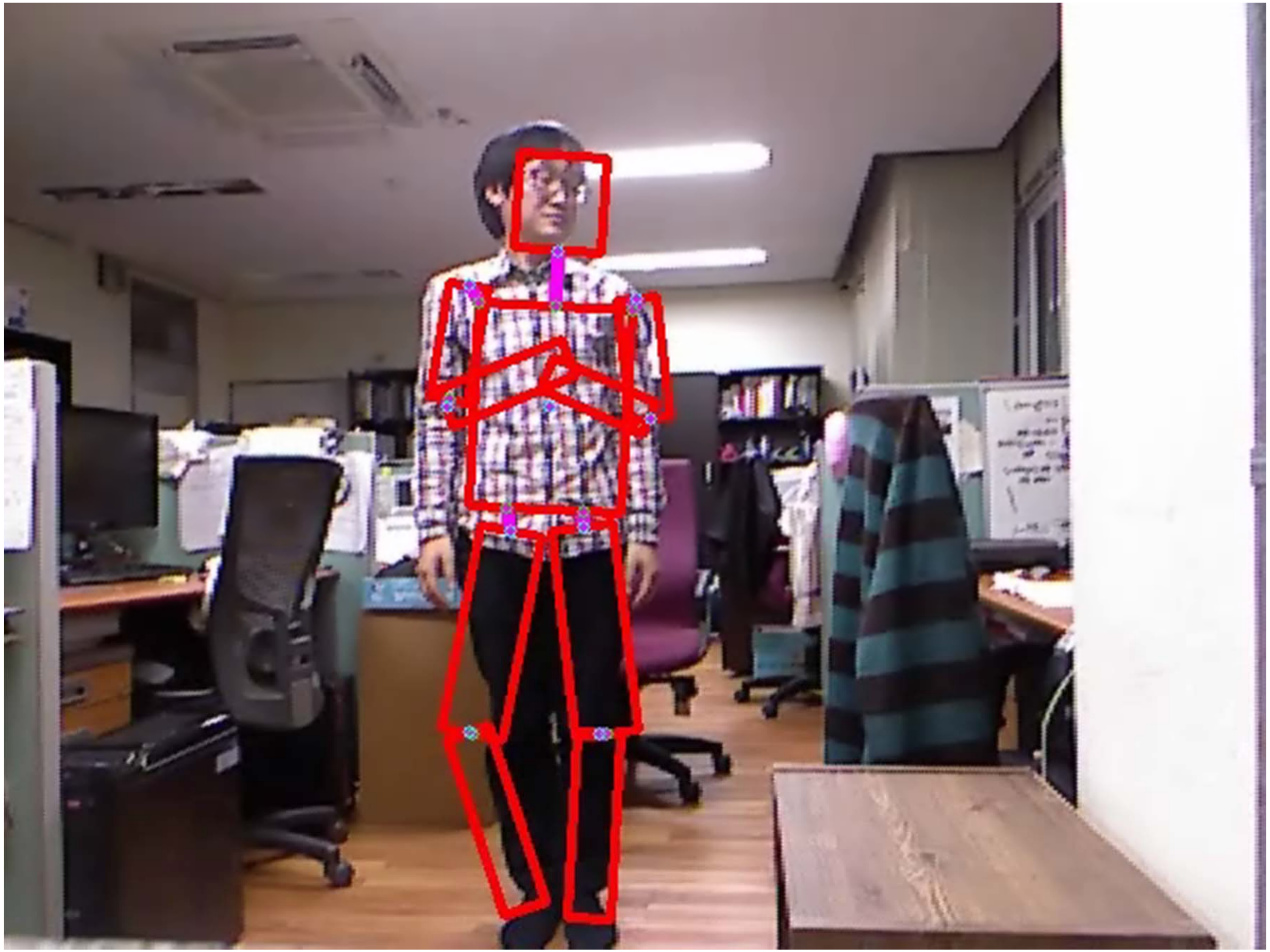
Experimental Results

Tracking Results



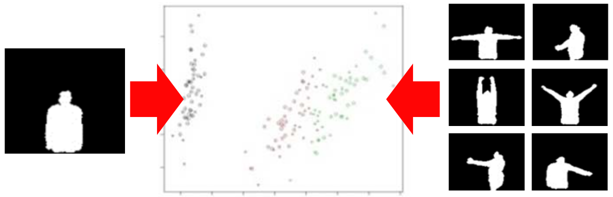




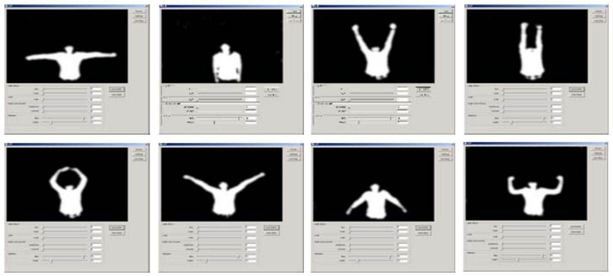


Game Interaction with Pose Recognition

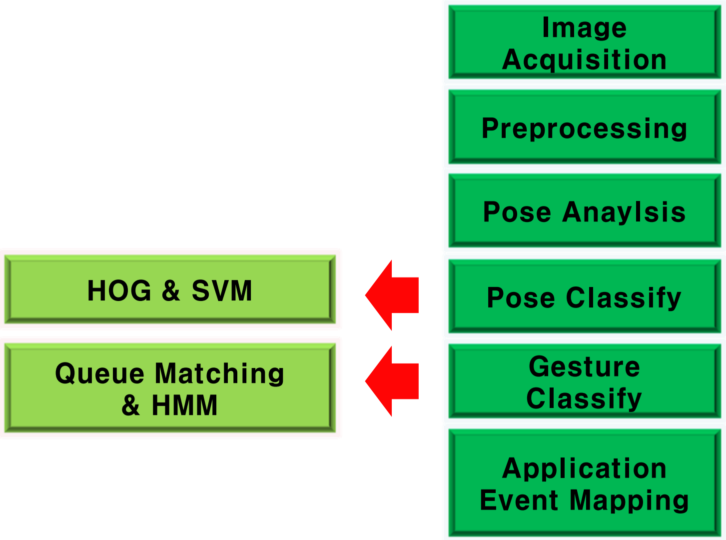
- Pose Recognition using HoG and SVM
- Periodic queue matching for robust analysis
- HMM with previously trained state transition probability



SVM의 포즈 분류 과정

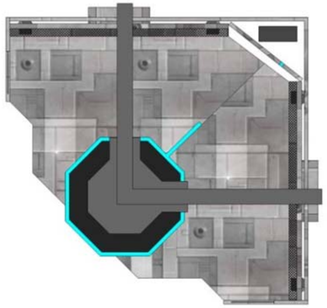


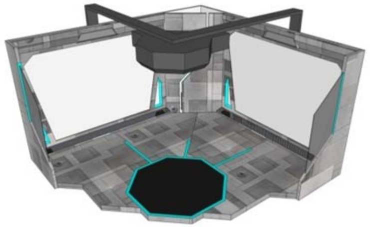
제스처 인식 프로그램 화면

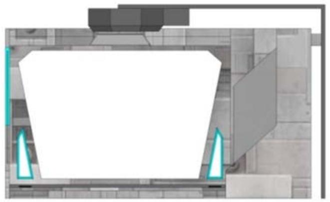


제스처 인식 과정

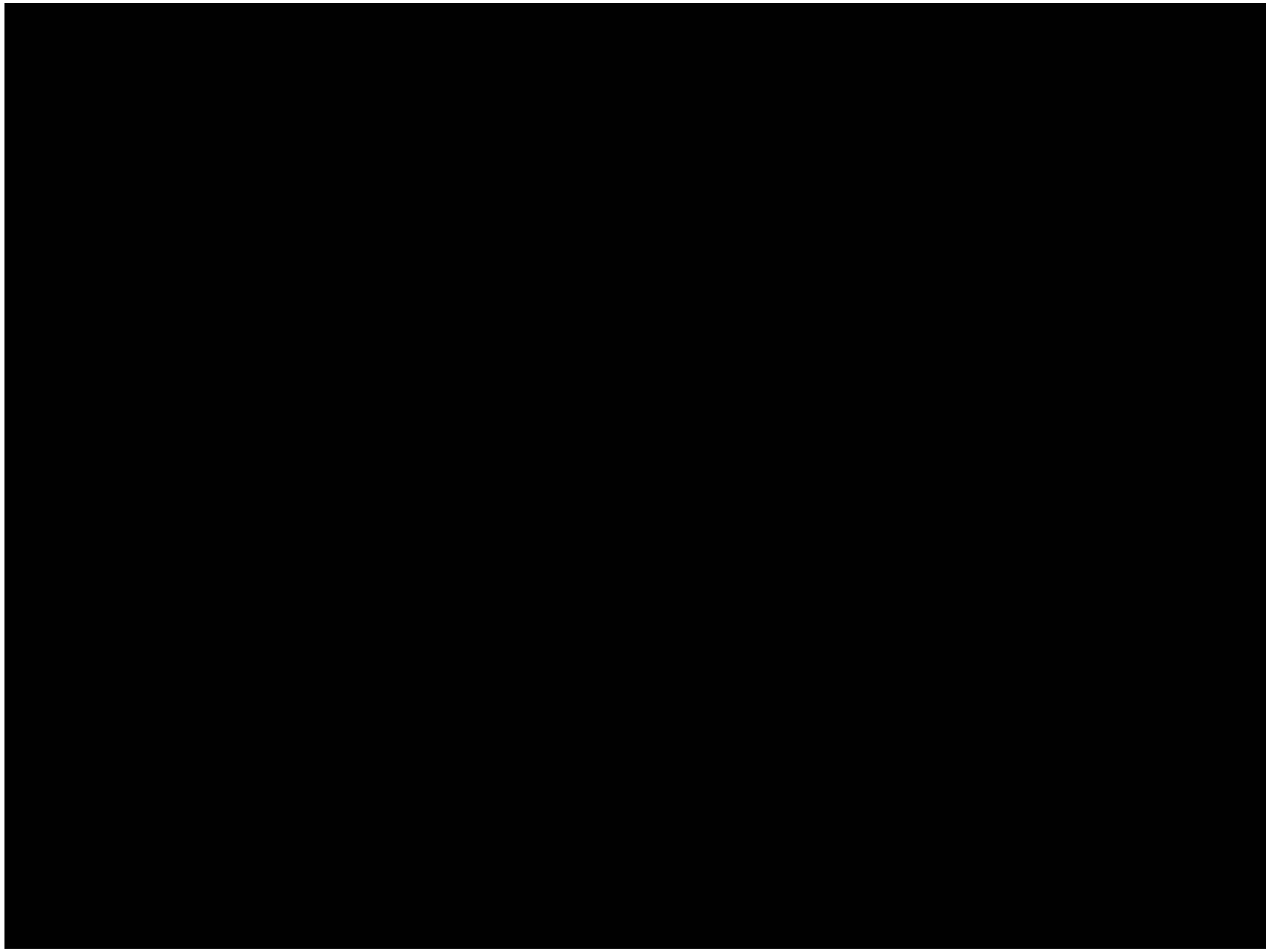
Game System



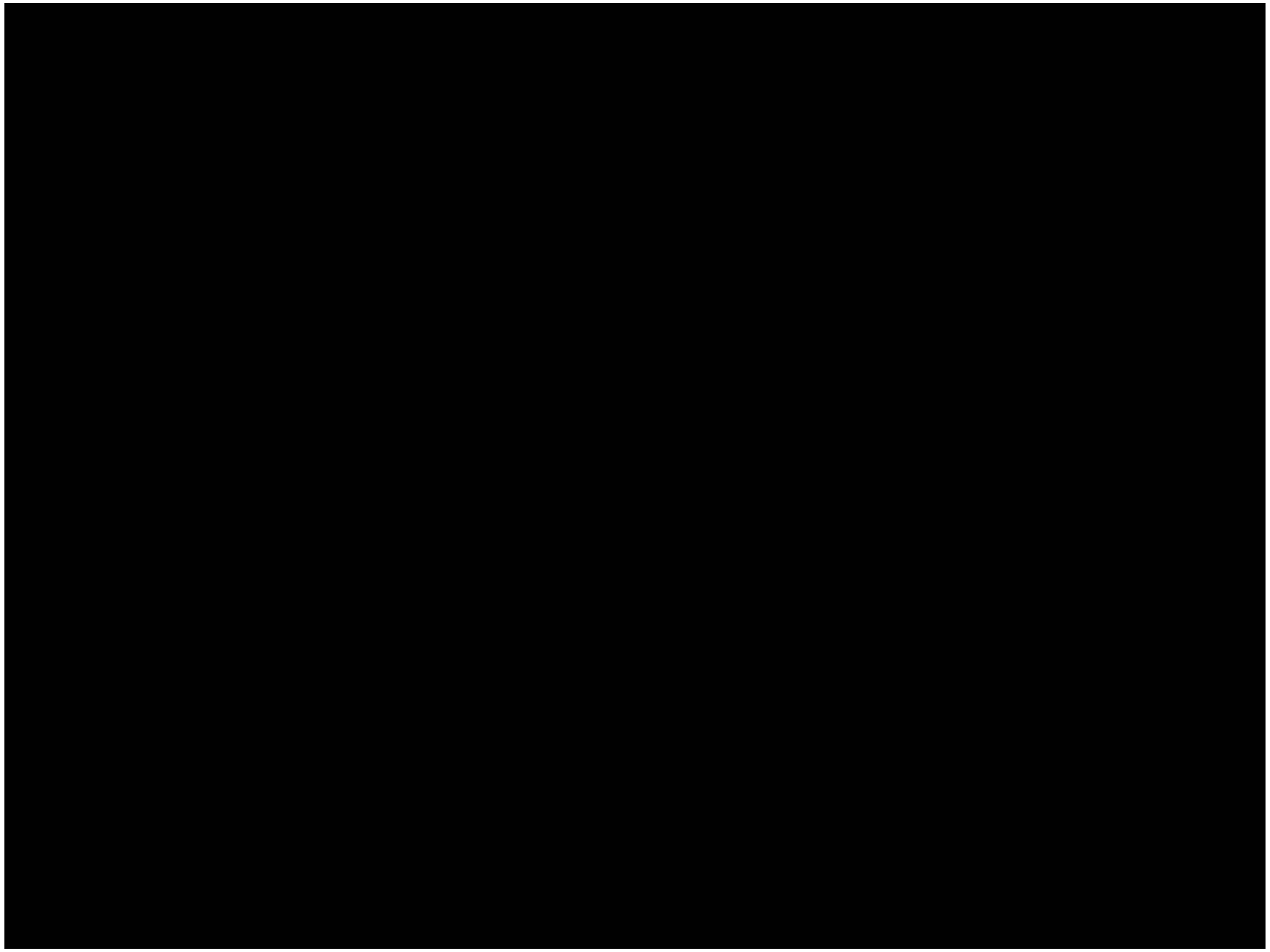




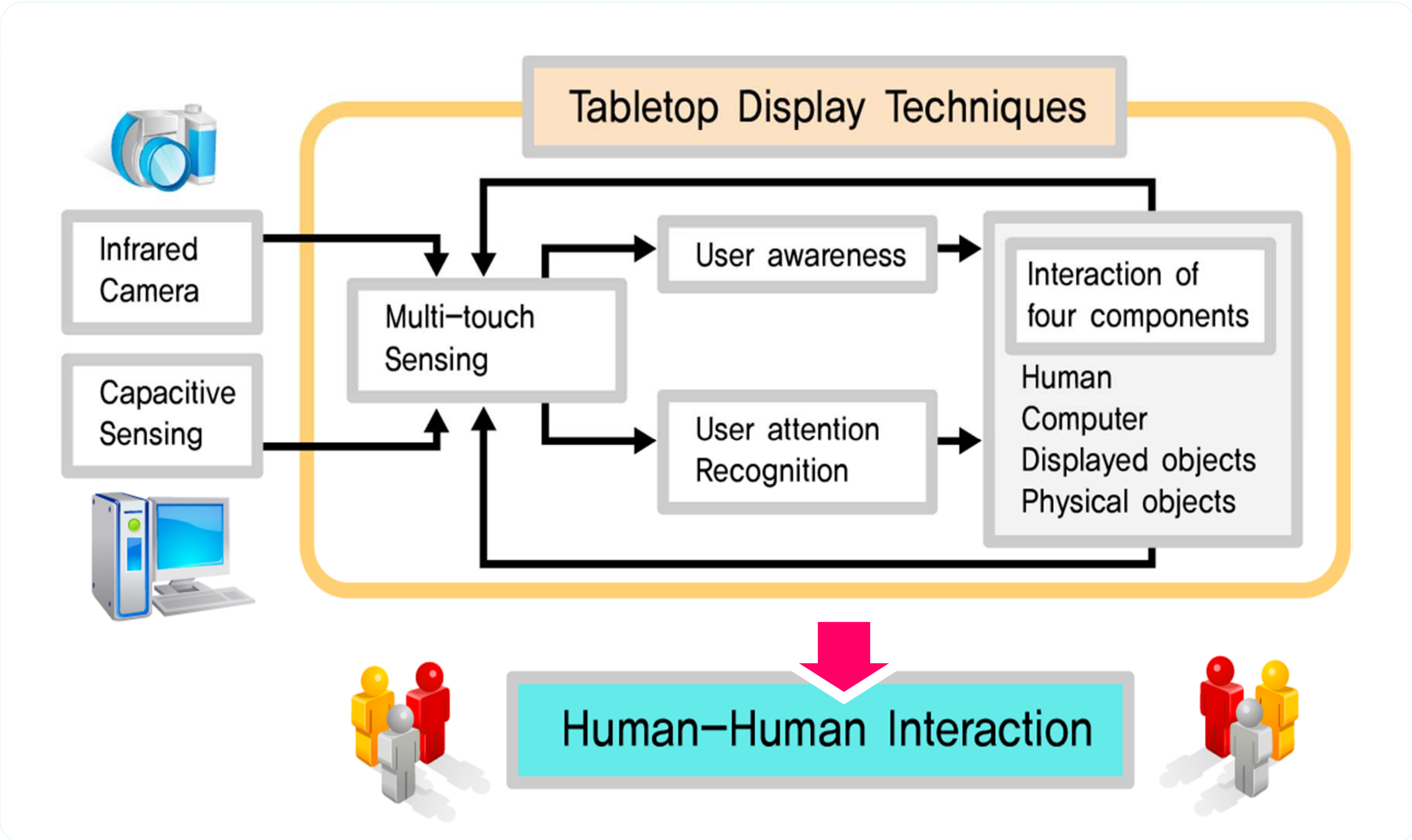
Gesture Game: demo(1)



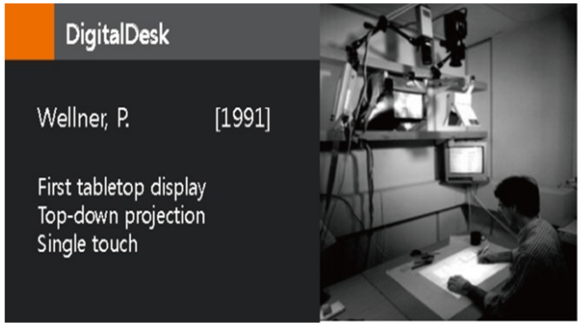
Gesture Game: demo(2)

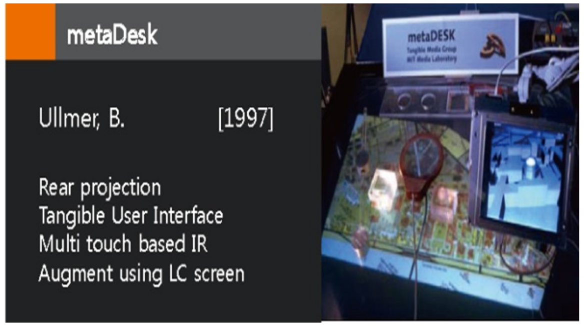


Tabletop Display: Architecture



Tabletop Display: examples

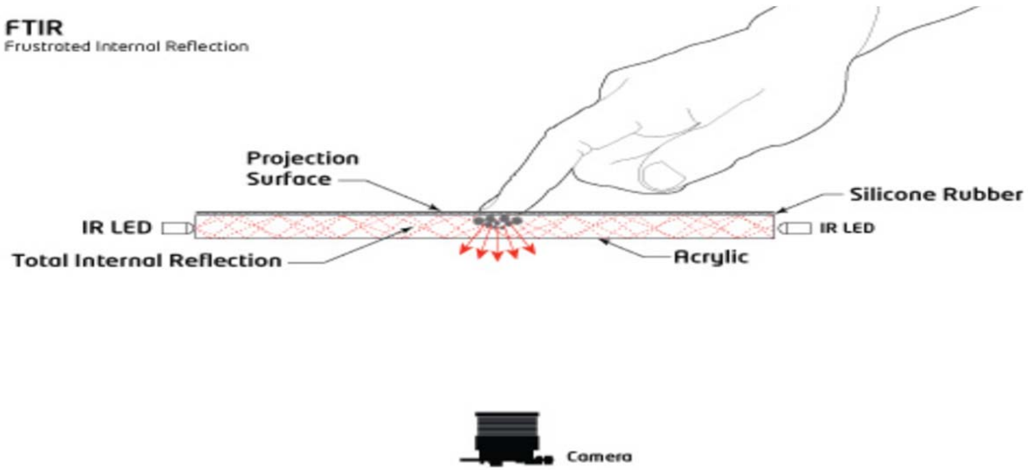




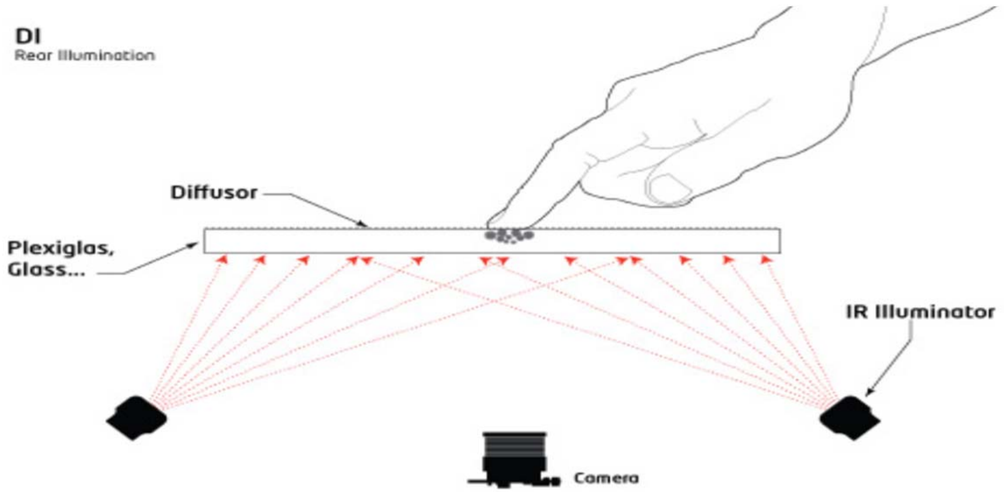
TouchFace-Ⅰ & Ⅱ

터치페이스 1호기

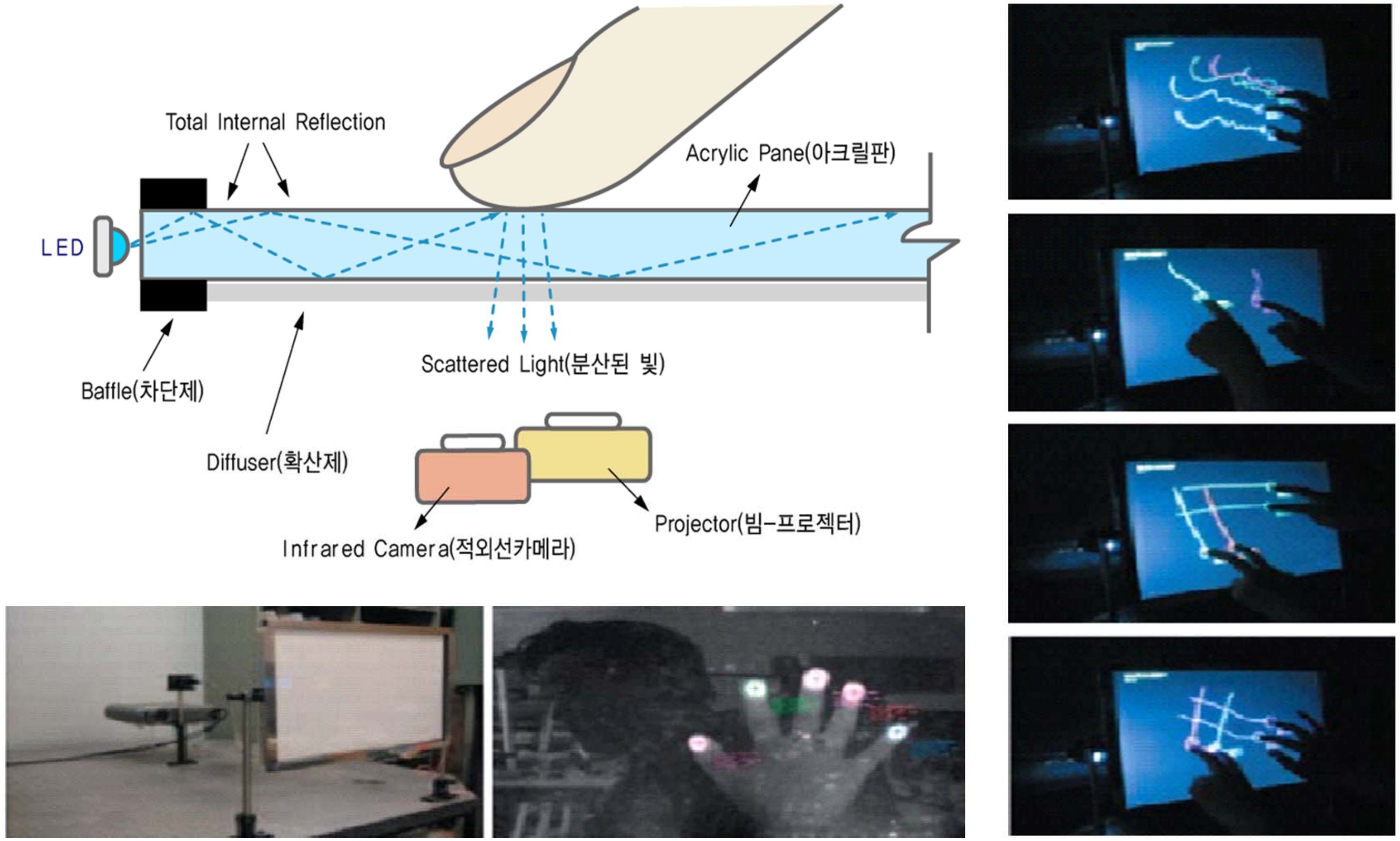
터치페이스 2호기



TouchFace-III & IV

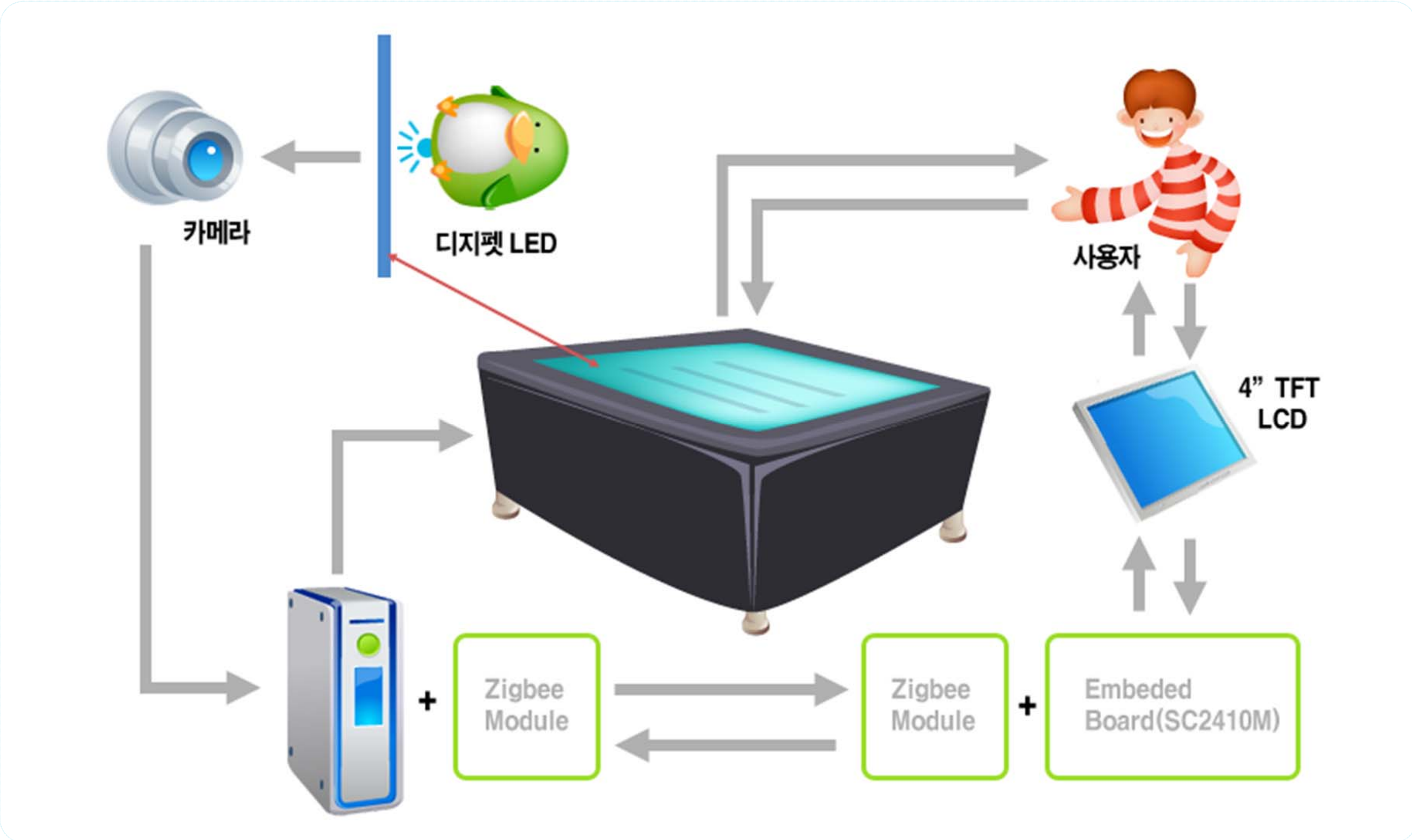


FTIR: Frustrated Total Internal Reflection

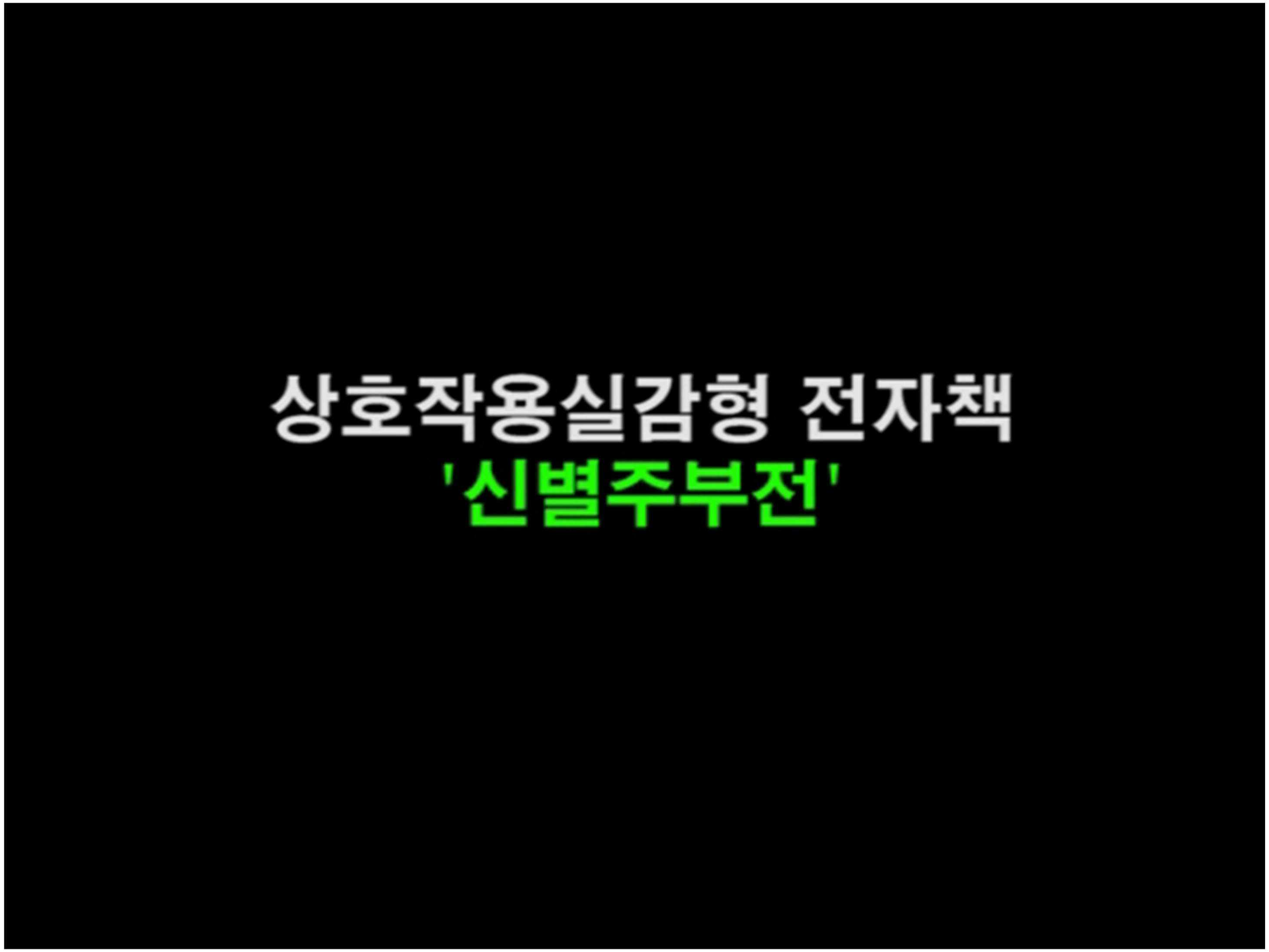


DigiPet: Abstract

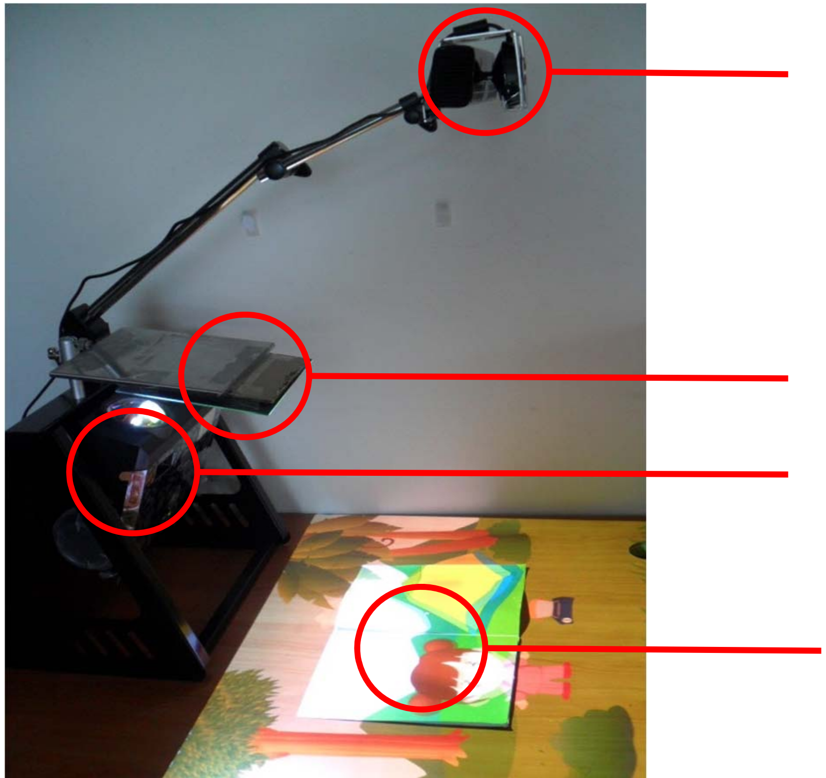
DEMO



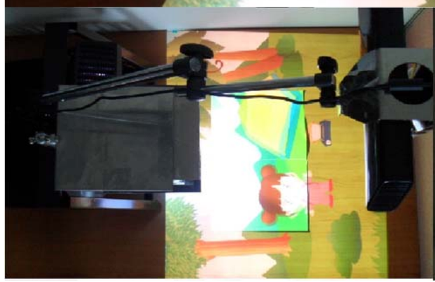
Content Interaction

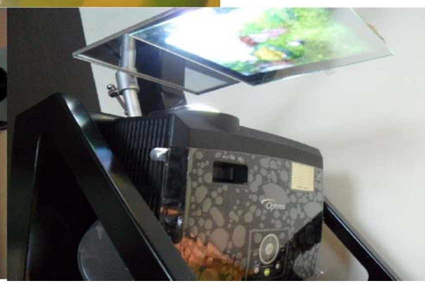


TouchFace-V



tangible-book :: **RealBook**





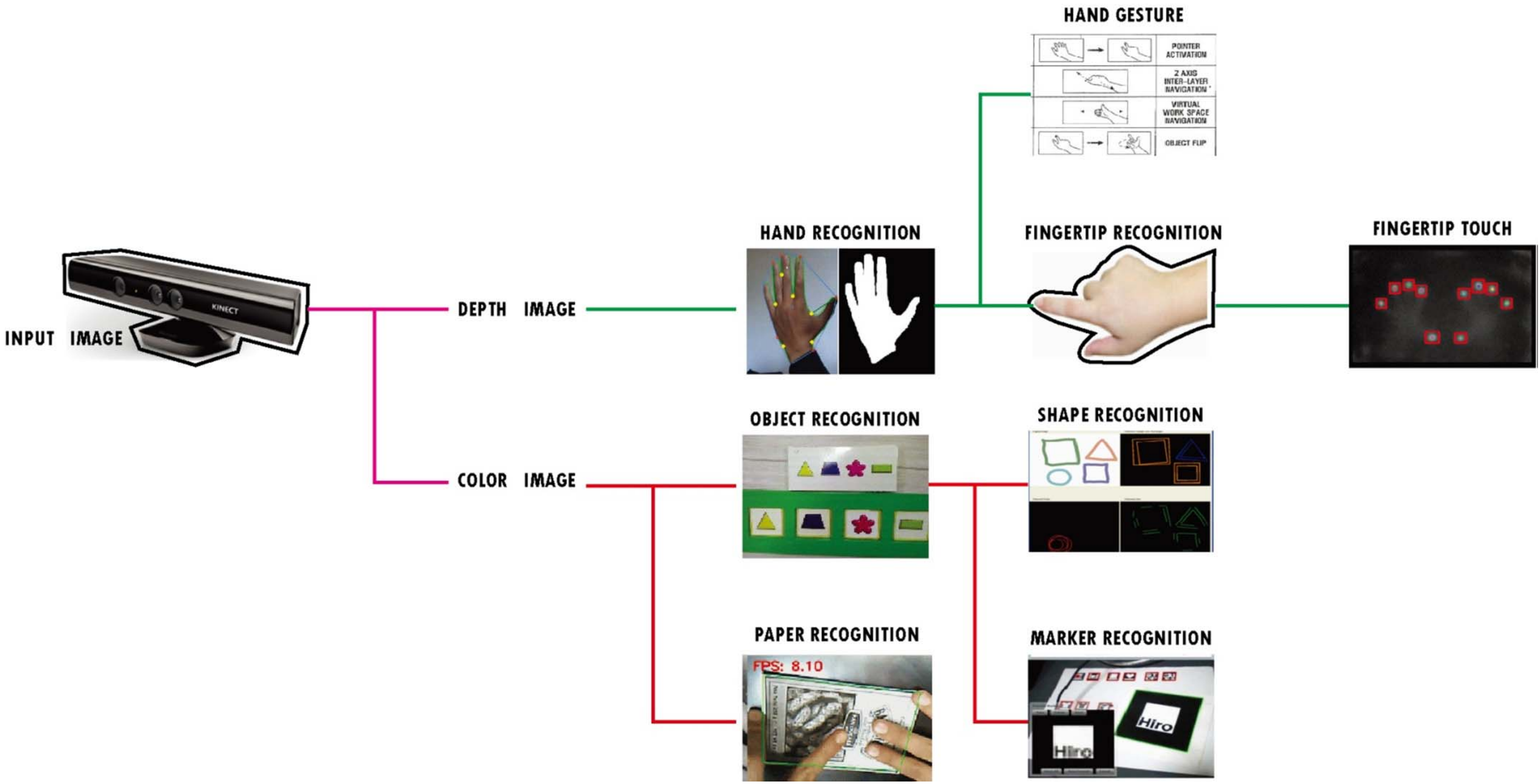
size : 40cm(w) x 50cm(l) x 70cm(h)

weight : 6.2kg

COMPACT

Algorithm: Abstract

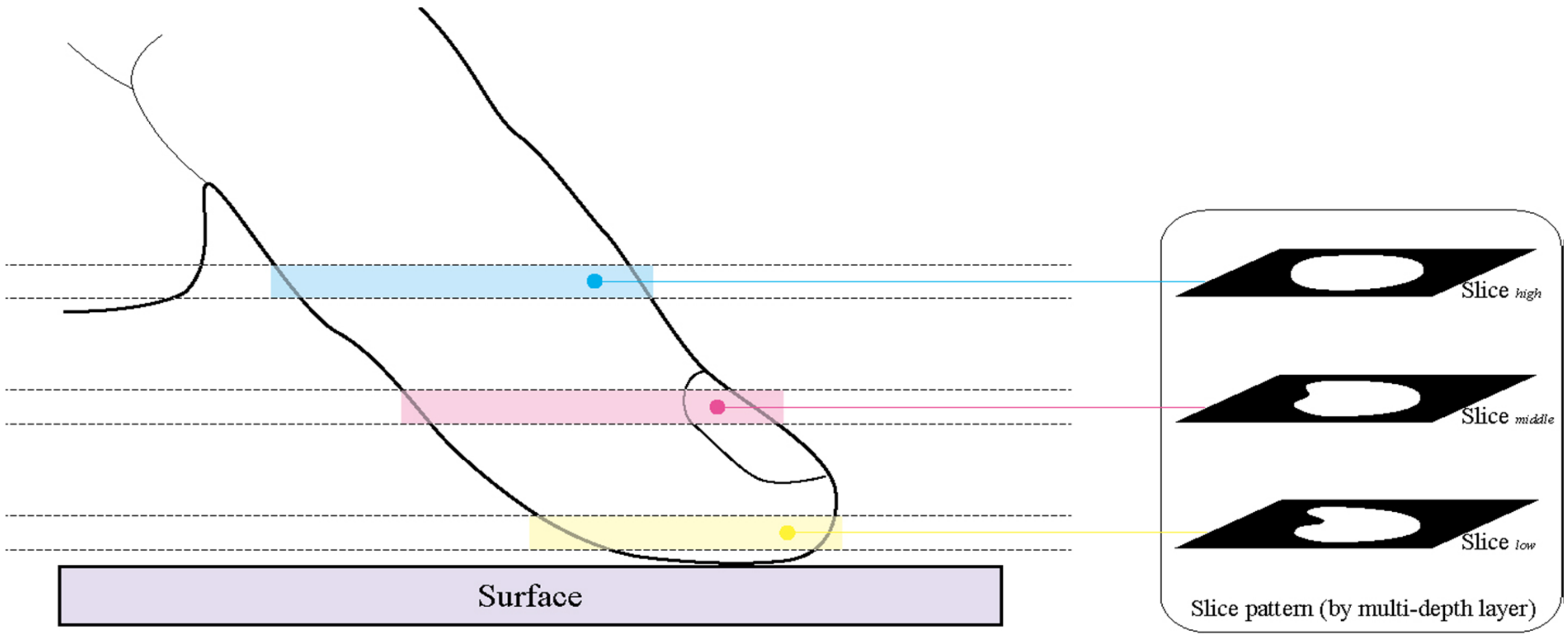
- A Compact Interactive Tabletop Projection-Vision Interface



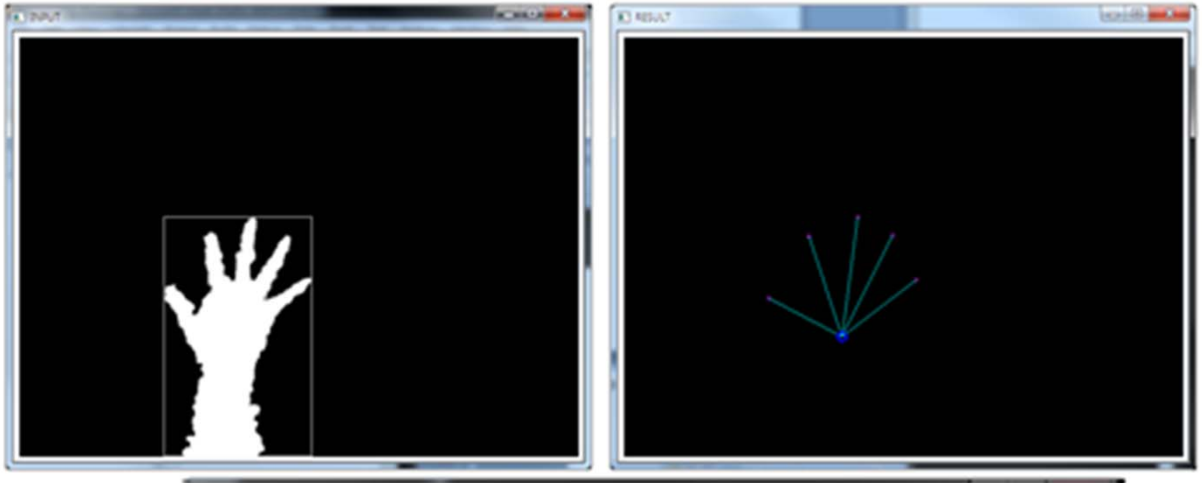
Precise Finger Touch Recognition

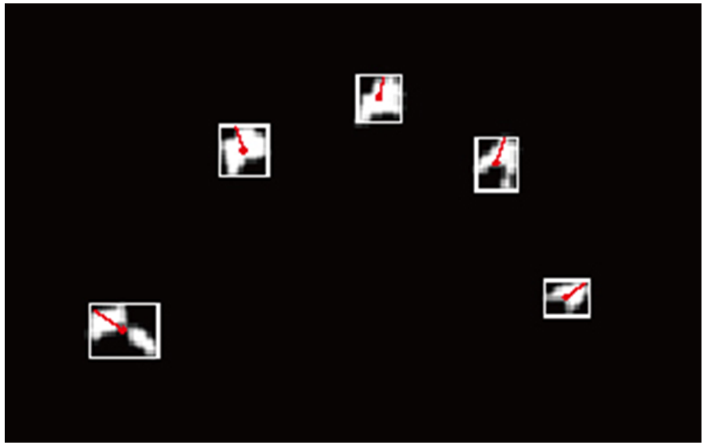
- **To extract unique patterns from multiple depth section**

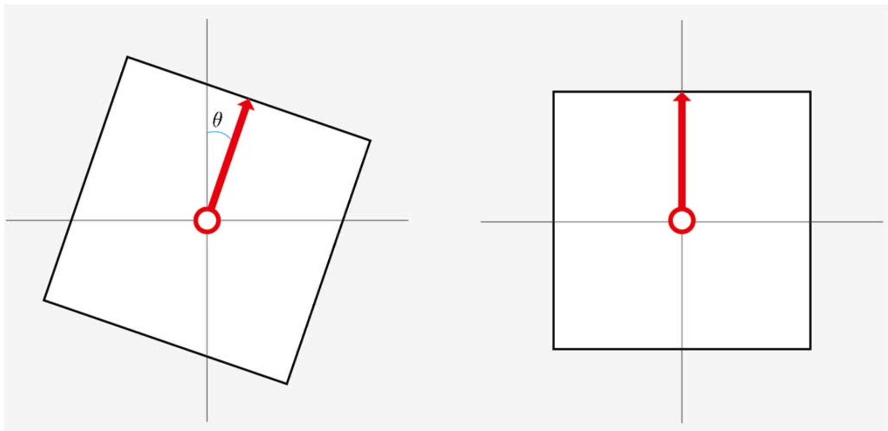
  → Multi-depth layer : **slice pattern**

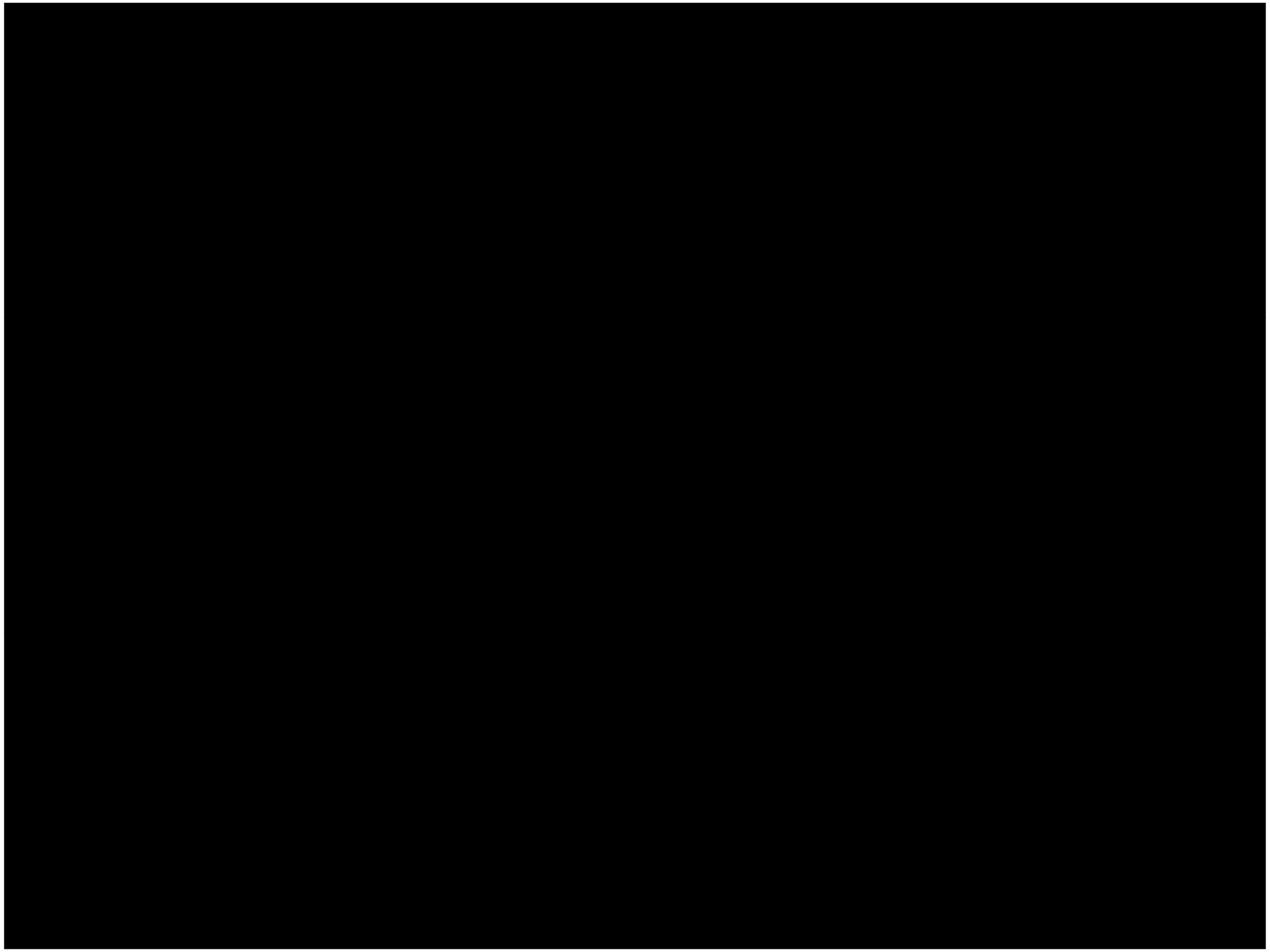


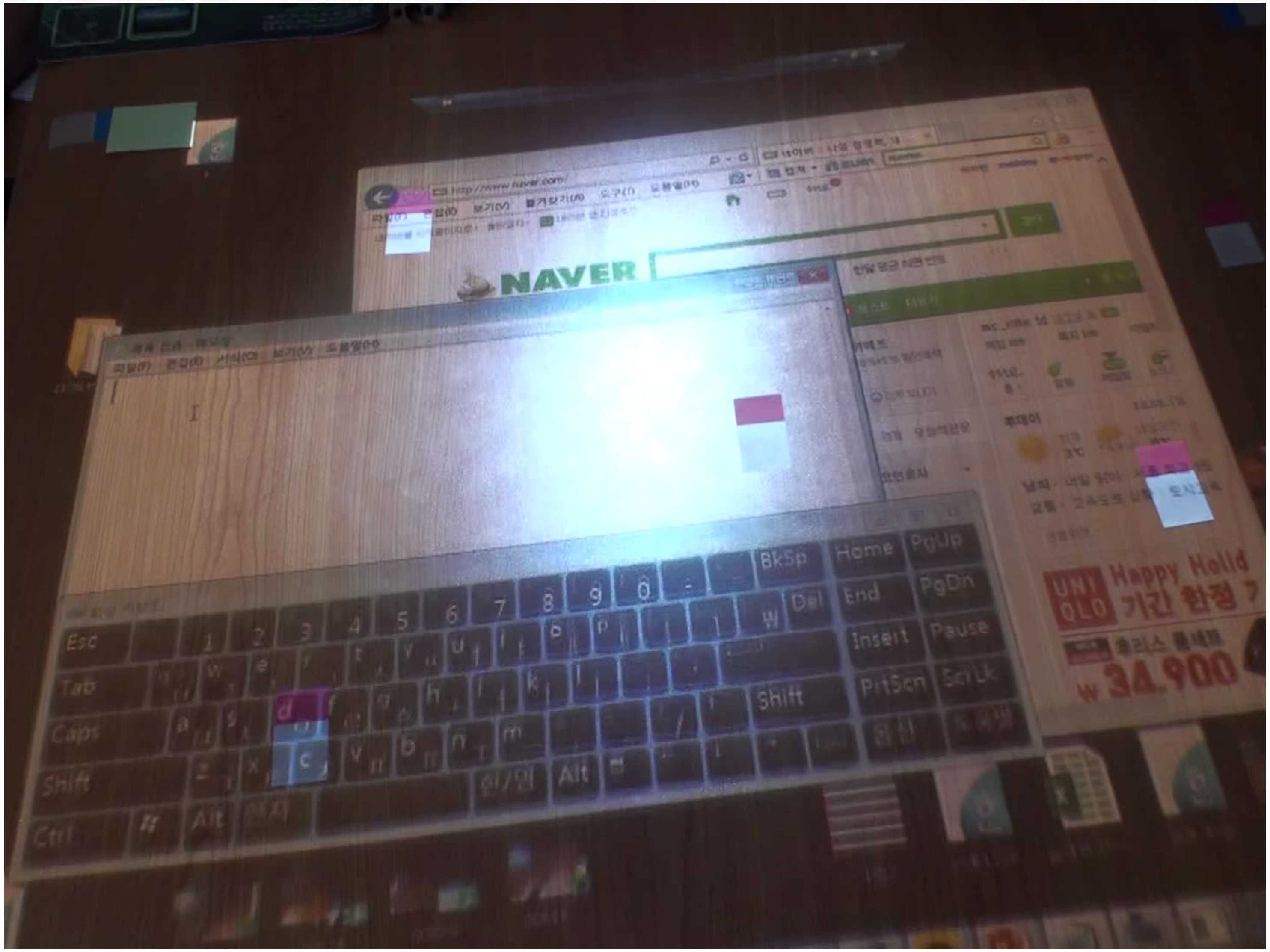
Pattern Analysis










NAVER
Esc
Tab
Caps
Shift
Ctrl
Alt
BkSp
Home
PgUp
Del
End
PgDn
Insert
Pause
PrtScn
ScrLk
UNI
QLO
Happy Holid
34,900

Conclusion

- Research paradigm of simple approach with real-time processing is presented
- Some visual recognition algorithms are presented for interactive systems
- Pose estimation algorithm for game interaction is presented
- A hierarchical pose estimation algorithm of upper body based on pictorial structures is proposed
- Real-time interaction on tabletop display is presented for content interaction

Thank You !!

leecw@jnu.ac.kr