

Introduction

○○○
○

Gaze Tracker

○○○
○

Robot

○
○○○

Camera

○○
○

Robot and Camera

○○○○○
○

Robot's perception in natural interaction skills learning

Bozorgmehr Aminian

Idiap

August, 2018

1 Introduction

- Motivation
- Approach

2 Gaze Tracker

- Data
- Calibration

3 Robot

- Setup
- Kinematics

4 Camera

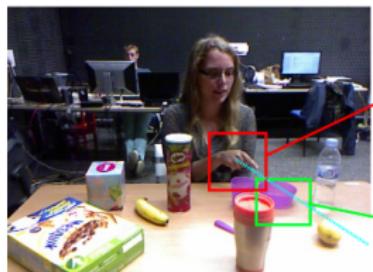
- Intrinsic
- Extrinsic

5 Robot and Camera

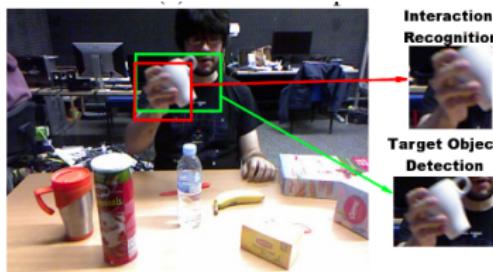
- Calibration
- Synchronization

Provides the required measurements about :

- People
- Scene object
- Gestures
- Verbal and non-verbal communication



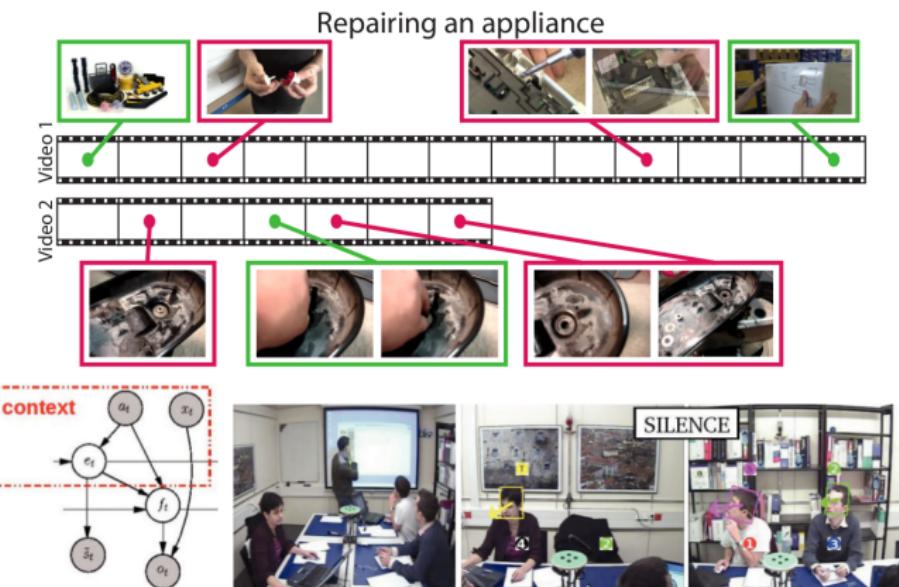
(a) Point Example



(b) Show Example

Understanding of the teacher's natural behavior :

- Segmentation (start, end, reference-action sequence)
- Attention (person/object of interest)
- Feedback (positive/negative, interruption)



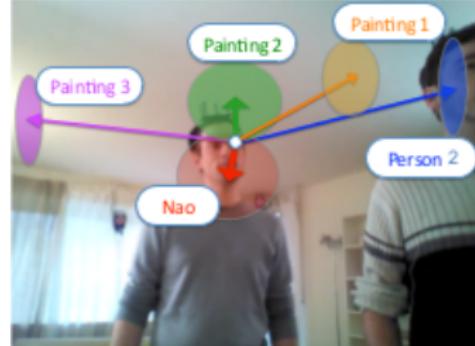
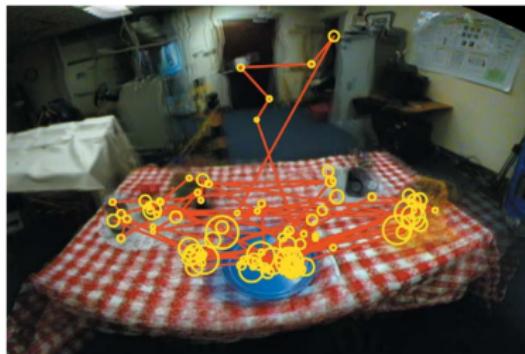
Use sensors in a every day's life framework (natural behavior) :

- RGB camera
- Depth sensor (kinect v2)
- Microphone

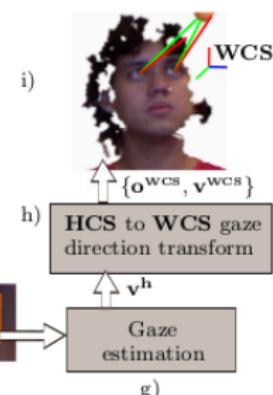
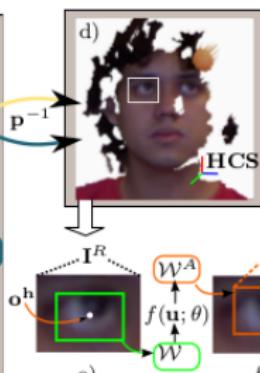
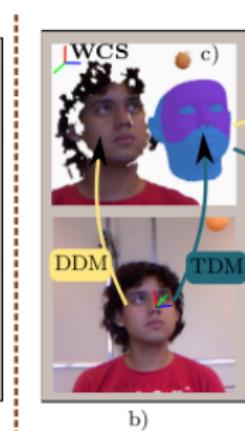
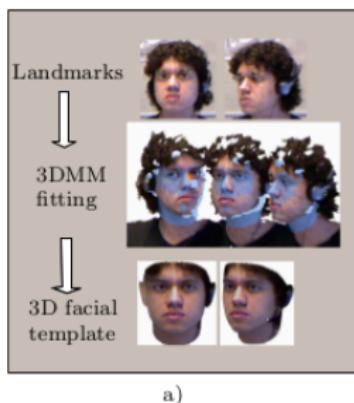


Gaze tracking and attention modeling :

- Unsupervised gaze calibration
 - Head tracker (3DMM, ICP).
 - Gaze tracker (G3E, kNN, H-SVR, R-SVR).
 - Gaze calibration (interaction prior based, LSMed).
- Contextual interaction models for gaze activity interpretation (VFOA)



Gaze Tracker



Introduction

○○○
○

Gaze Tracker

○○○
○

Robot

○
○○○

Camera

○○
○

Robot and Camera

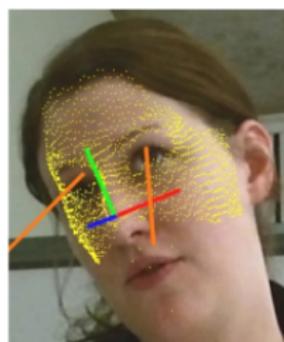
○○○○○
○

Gaze Tracker

Wrong Head pose
⇒ Wrong cropping



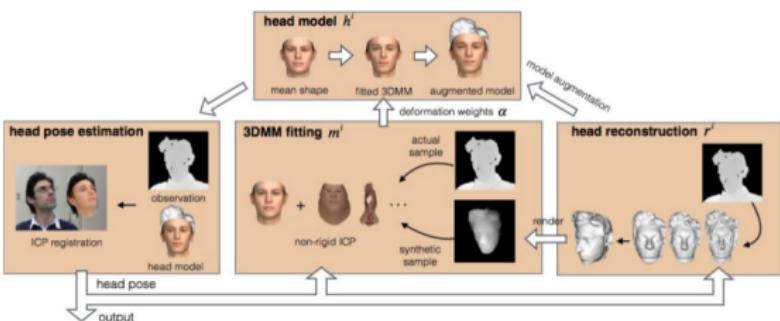
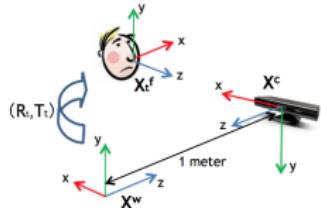
Correct Head pose
⇒ Correct cropping



Head Pose :

Rotation matrix $R_t \in SO(3)$, $T_t \in \mathbb{R}^3$ vector at time $t \in \mathbb{R}$. The head pose is defined by

$$p_t := (R_t, T_t) \quad (1)$$



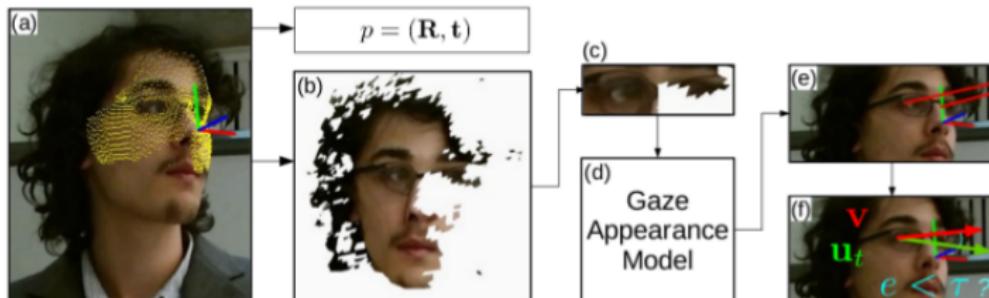
Eye Center :

Coordinate of the eye center $o_t \in \mathbb{R}^3$.

Gaze Direction :

Tilt angle $\phi_t \in [-\pi/2, \pi/2]$ and pan angle $\theta_t \in [-\pi/2, \pi/2]$ at time t in a *head coordinate system*. The gaze direction is defined by

$$g_t = (\phi_t, \theta_t) \quad (2)$$



Unitary Gaze Direction Vector :

There exists a mapping $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ such that the unitary gaze direction vector is given by

$$\Phi(g_t) = v_t, \quad \forall t \in \mathbb{R}. \quad (3)$$

Ground Truth :

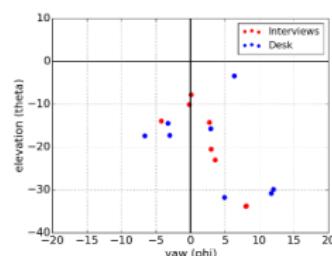
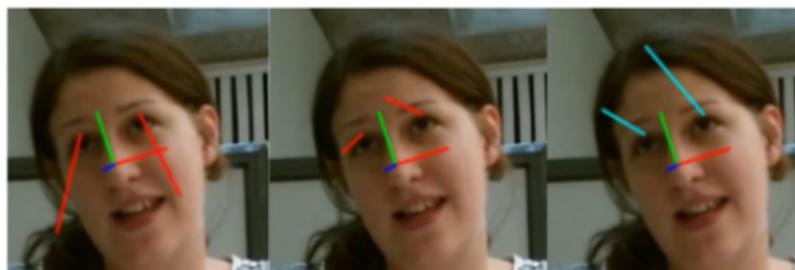
For an object coordinate $x \in \mathbb{R}^3$. The ground truth is defined by

$$\bar{v}_t = \frac{x - o_t}{\|x - o_t\|} \text{ and } \bar{g}_t = \Phi^{-1}(\bar{v}_t) \quad (4)$$

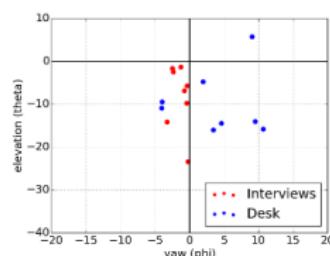
Gaze Calibration

For a set of parameters $\alpha = \{\alpha_1, \dots, \alpha_k\}$. The gaze calibration consists in the optimization problem

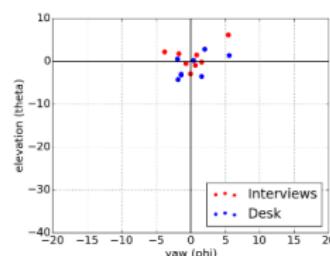
$$\text{minimize } \|G(p_t, g_t; \alpha) - \bar{g}_t\| \text{ for } t \in \mathbb{R}. \quad (5)$$



(a) Baseline



(b) Lm



(c) Lm + Spk-Med

Introduction

○○○
○
○

Gaze Tracker

○○○
○

Robot

●
○○○

Camera

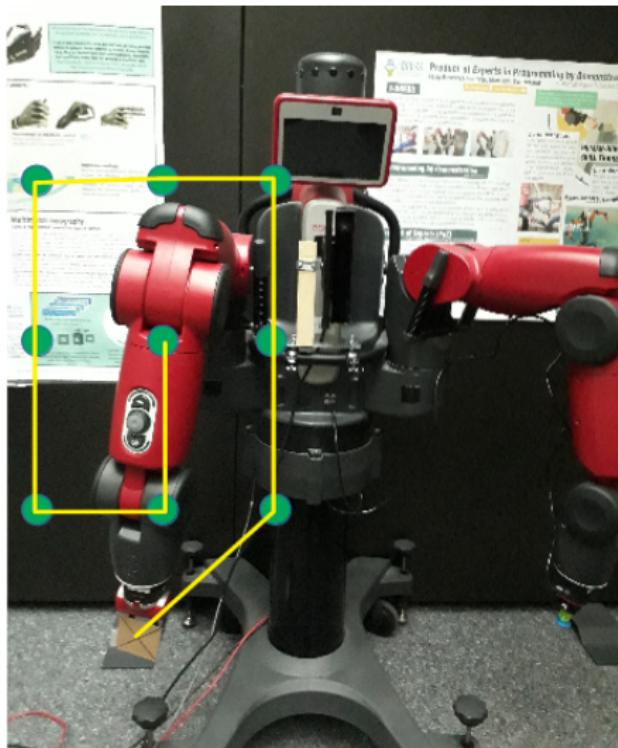
○○
○

Robot and Camera

○○○○○
○

Setup

Gaze Calibration Experiment :



Forward kinematics :

For $\theta \in \mathbb{R}^k$, define $X : \mathbb{R}^k \rightarrow \mathbb{R}^n$ such that

$$X = X(\theta). \quad (6)$$

Inverse kinematics :

For $Y \in \mathbb{R}^n$, find $\theta \in \mathbb{R}^k$ such that

$$X(\theta) - Y = 0 \quad (7)$$

- Cost function :

$$E(\theta) = \frac{1}{2} \|X(\theta) - Y\|^2 \quad (8)$$

- Gradient :

$$\begin{aligned}\frac{\partial E}{\partial \theta}(\theta) &= (X(\theta) - Y)^T \frac{\partial X}{\partial \theta}(\theta) \\ &= (X(\theta) - Y)^T J \\ &=: d\theta\end{aligned}$$

- Update rule

$$\theta^{m+1} = \theta^m - \eta d\theta^m \quad (9)$$

- Constraints :

$$\theta^- \preceq \theta \preceq \theta^+ \quad (10)$$

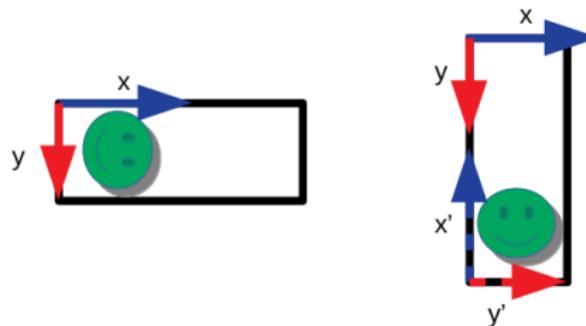
- Indicator : for $i = 1, \dots, k$,

$$[\text{ldc}(\theta)]_i = \mathbb{1}_{\{\theta_i^- \leq \theta_i \leq \theta_i^+\}^c} = \begin{cases} 0 & \text{if } \theta_i^- \leq \theta_i \leq \theta_i^+, \\ 1 & \text{otherwise.} \end{cases} \quad (11)$$

- Update rule (identity matrix $I_k \in \mathbb{R}^{k \times k}$)

$$\theta^{m+1} = \theta^m - (\eta I_k - \lambda \text{diag}(\text{ldc}(\theta^m))) d\theta^m, \quad \lambda \geq \eta. \quad (12)$$

Camera rotation :



$$x' = -y \text{ and } y' = x \quad (13)$$

Hence

$$P = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (14)$$

Camera intrinsic :

- Camera matrix :

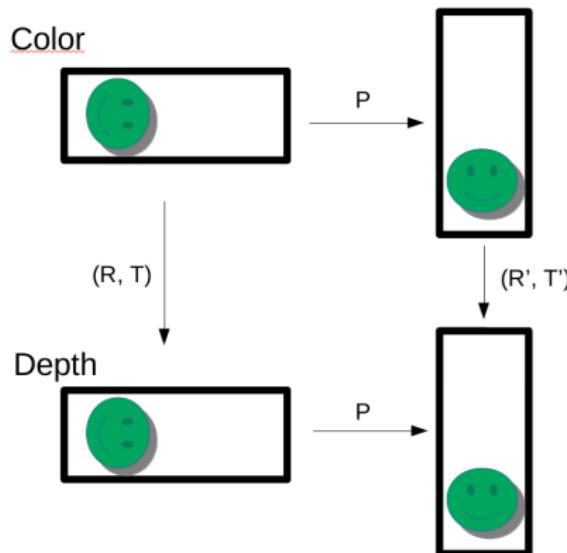
$$K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}. \quad (15)$$

- After rotation :

$$K' = \begin{pmatrix} f_y & 0 & c_y \\ 0 & f_x & s_x - c_x \\ 0 & 0 & 1 \end{pmatrix} \quad (16)$$

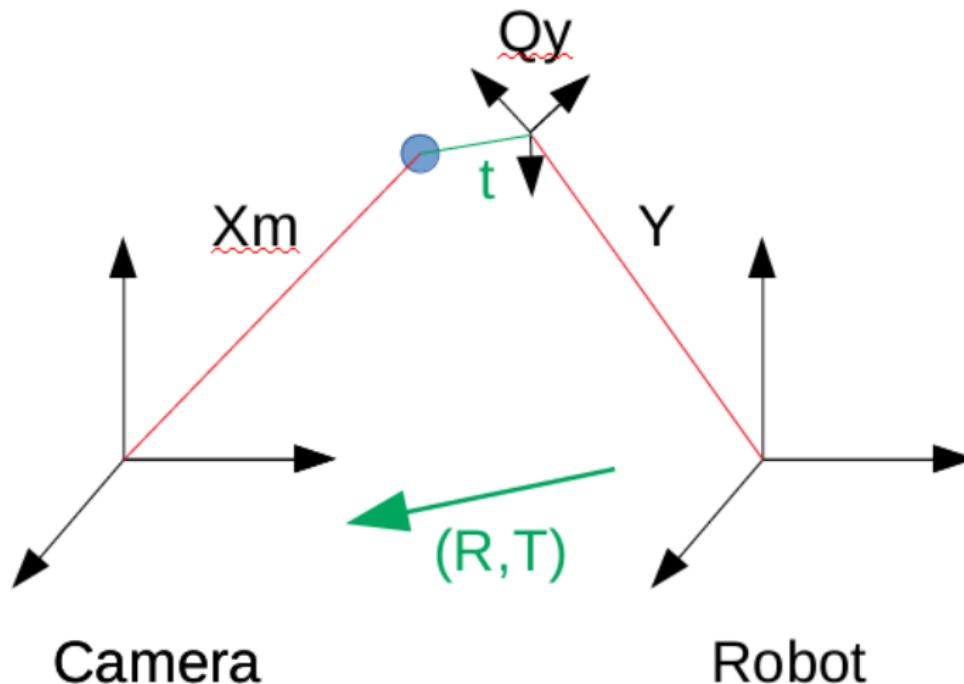
where (s_x, s_y) is the size of the original image.

Camera extrinsic :



$$R' = PRP^{-1} \text{ and } T' = PT \quad (17)$$

Robot-camera calibration



- \mathcal{F}_X \mathcal{F}_Y : camera frame, resp. robot frame.
- X, Y : coordinate of the end-effector in \mathcal{F}_X , resp. in \mathcal{F}_Y .
- X_m, Y_m : coordinate of the marker in \mathcal{F}_X , resp. in \mathcal{F}_Y .
- \mathcal{F}_{Q_Y} : end-effector frame.
- t : coordinate of the marker in \mathcal{F}_{Q_Y} .
- $Q_Y \in \mathbb{R}^{3 \times 3}$: basis vector of the frame \mathcal{F}_{Q_Y} expressed in frame \mathcal{F}_Y . It corresponds to the rotation matrix from \mathcal{F}_Y to \mathcal{F}_{Q_Y} .

- Known data (n images) : $X_m = \{X_m^i\}_{i=1}^n$, $Y = \{Y^i\}_{i=1}^n$ and $Q_Y = \{Q_Y^i\}_{i=1}^n$.
- Unknowns : rigid transform (R, T) from \mathcal{F}_Y to \mathcal{F}_X and vector $t \in \mathbb{R}^3$:

$$RY + T = X \text{ and } RY_m + T = X_m \quad (18)$$

- Remark :
 - $R = R_\gamma R_\beta R_\alpha$ where (α, β, γ) are Euler angles.
 - $Q_Y = Q_Y(q_Y)$ where q_Y is the quaternion coordinates of the orientation of the end-effector in \mathcal{F}_Y .

- Solve for images $i = 1, \dots, n$:

$$R(Y^i + Q_Y^i t) + T = X_m^i \quad (19)$$

- Substract mean :

$$R(\bar{Y}^i + \bar{Q}_Y^i t) = \bar{X}_m^i \quad (20)$$

where

$$\bar{Y}^i = \left(Y^i - \frac{1}{n} \sum_{k=1}^n Y^i \right), \quad \bar{Q}_Y^i = \left(Q_Y^i - \frac{1}{n} \sum_{k=1}^n Q_Y^i \right)$$

$$\bar{X}_m^i = \left(X_m^i - \frac{1}{n} \sum_{k=1}^n X_m^i \right)$$

- Cost function :

$$E(\alpha, \beta, \gamma, t) = \frac{1}{2} \sum_{i=1}^n \| R(\bar{Y}^i + \bar{Q}_Y^i t) - \bar{X}_m^i \|^2 \quad (21)$$

Procrustes problem :

For $X, Y \in \mathbb{R}^{3 \times N}$

$$R = \underset{\Omega}{\operatorname{argmin}} \|\Omega X - Y\| \text{ subject to } \Omega^T \Omega = I \quad (22)$$

Procrustes solution :

For $M := YX^T = U\Sigma V^T$ a singular value decomposition (SVD),
then

$$R = UV^T. \quad (23)$$

Synchronize robot and camera time :

- Robot has its own time.
- Camera's time is computer time (- latency)
- Set robot's time to computer time (latency is small for robot topics).

Play Video