Sensing Visual Attention by Sequential Patterns

MoBT4p.37

Yasuyuki Yamazaki, Hideitsu Hino, Kazuhiro Fukui
Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba, Japan

筑波大学 University of Tsukuba

Abstract

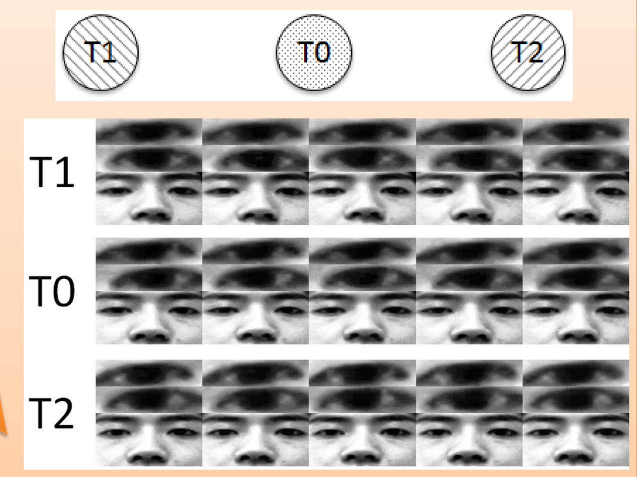
- We propose a concept of "**Vision Switch**" that reacts to human gaze and triggered by visual attention
- Using sequential **images from single camera**
- By the **analysis of sequential image patterns**
- Two mechanisms are introduced
  1. KOMSM* which is suitable for classifying sets of multiple images
  2. A kernel function for considering the head position

*Kernel Orthogonal Mutual Subspace Method

Each sequence is represented by a subspace
- Attention subspace
  - produced by the action of **gazing at a certain area**
- Non-attention subspaces
  - produced by the action of **gazing at other areas**

Introduction

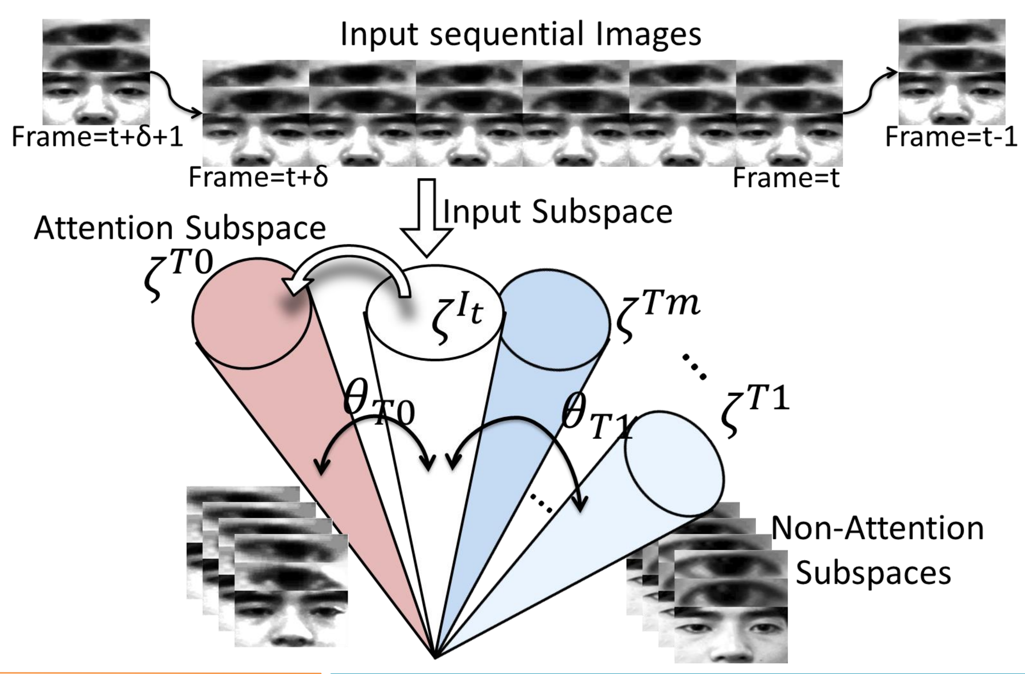
- **Vision Switch**
  - Triggered by visual attention
  - Under these three conditions
    1. No special equipment
    2. No constraint on head
    3. No special lighting



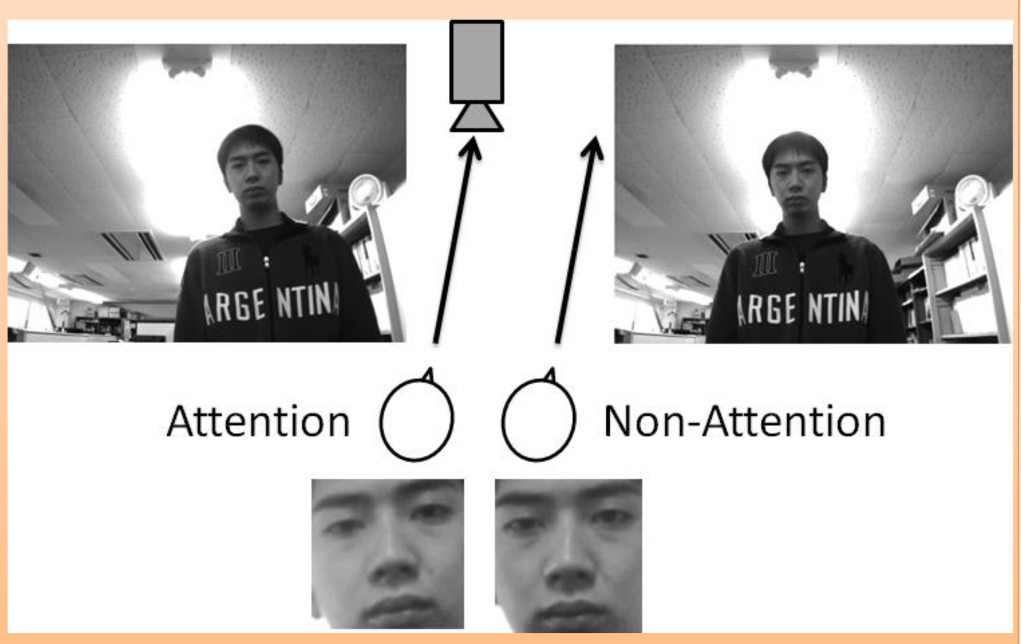
Example configuration for three areas and patterns corresponding to attention to each area

Implemented by **Gaze Detection**

- **Gaze Detection**
  - Is the problem to classify these images
  - A set of input sequential images are represented by a subspace
  - Calculate the angles between input subspace and (non-) attention subspaces
  - Attention is detected when the angle between ζ^{I_t} and ζ^{T0} (θ_{T0}) is the smallest



In this figure, the person moves parallel to the camera without moving head. Therefore, the cropped images are very similar



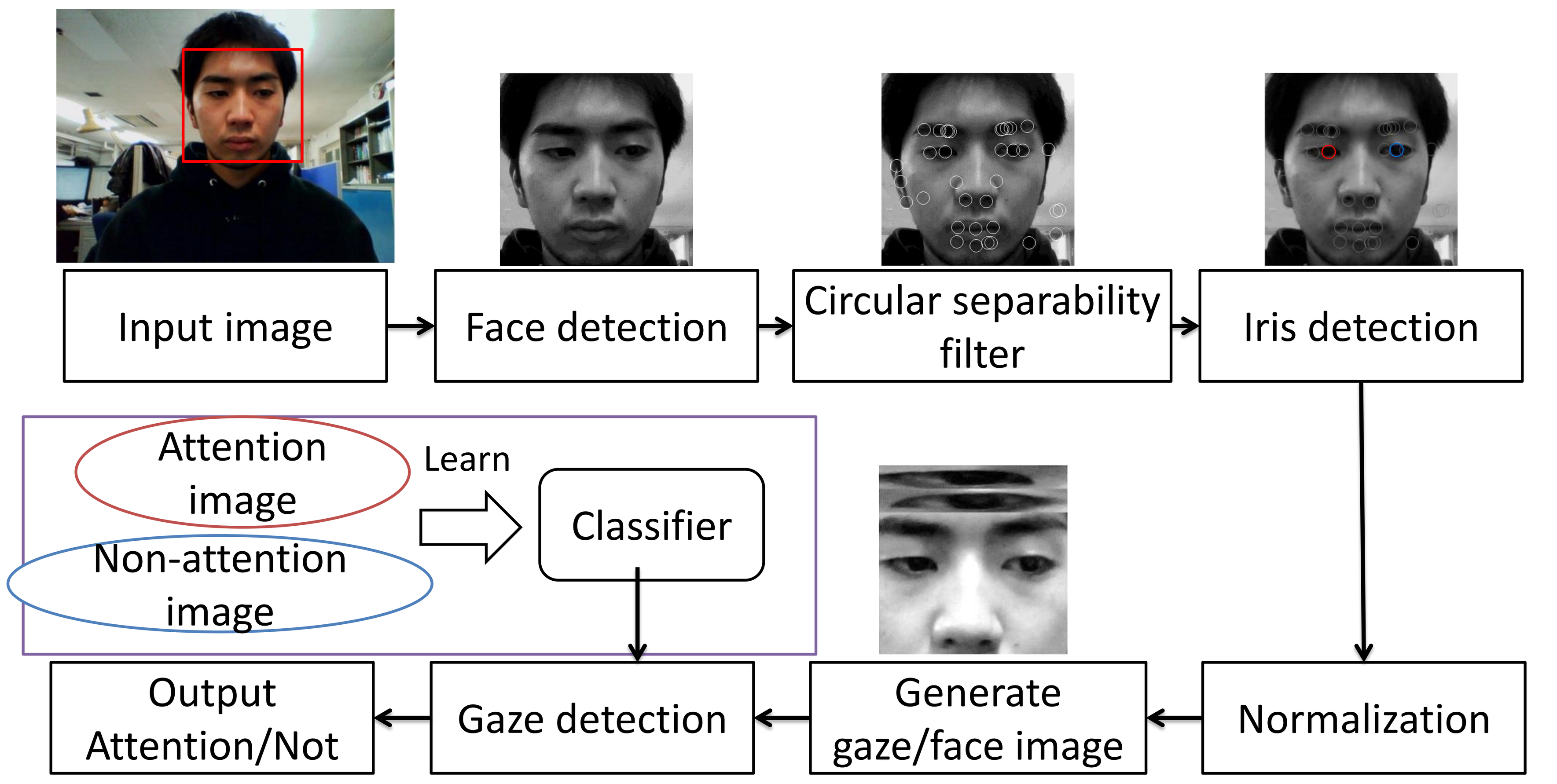
- **Problem**
  - We cannot detect attention only with the information of image
- **Solution**
  - Consider the head position as a feature and put it into the kernel

- **Additional kernel**
  - In KOMSM, we apply kernel function to image features (k_1)
  - To deal with changes in head position, we introduce an **additional kernel** k_2
  - k_2 represents **how far the head position is**
  - We use a kernel function of the form

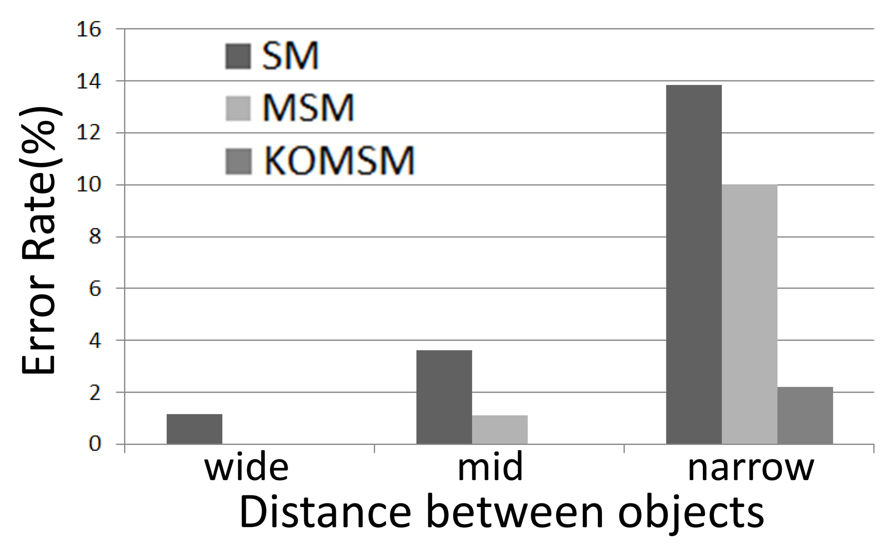
$$k(\{\boldsymbol{x}_i, \boldsymbol{p}_i\}; \{\boldsymbol{x}_j, \boldsymbol{p}_j\}) = \alpha_1 k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) + \alpha_2 k_2(\boldsymbol{p}_i, \boldsymbol{p}_j)$$

  - Here, $\boldsymbol{x}$ is the image feature, $\boldsymbol{p}$ is the head position (center of the detected face), α_1 and α_2 are the controlling parameters

Flow of gaze detection



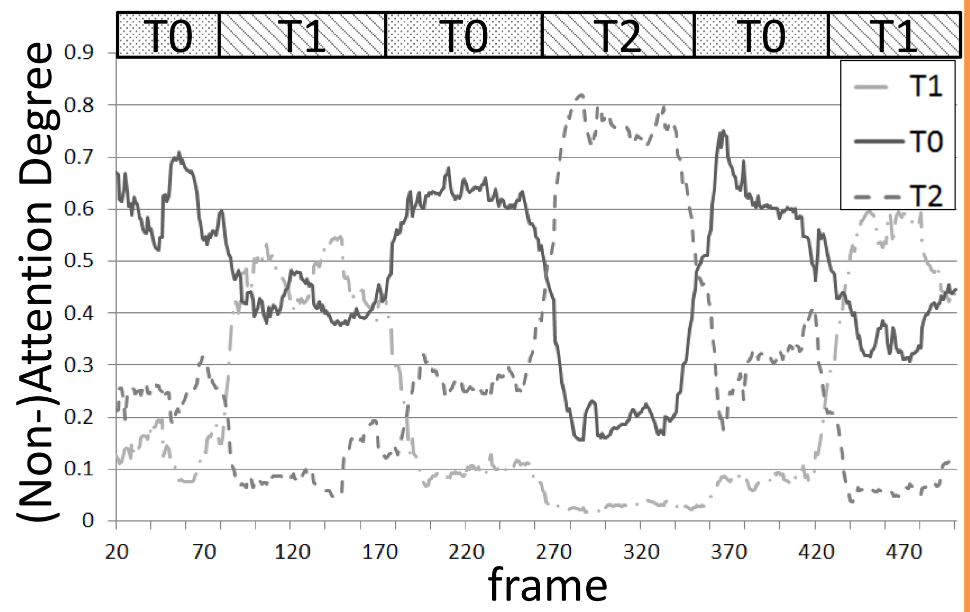
Experiment



Evaluate the performance of the proposed method when objects that are assumed as attention and non-attention are placed close each other

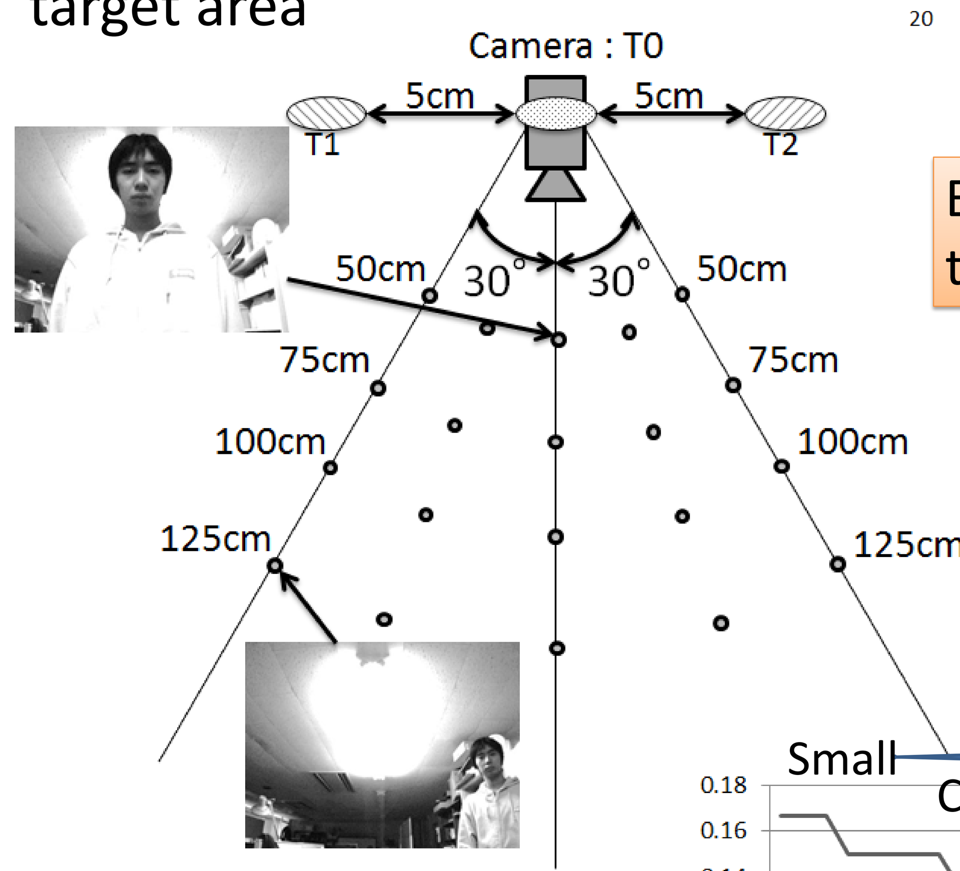
- KOMSM outperformed the other methods with 97.8% of accuracy in narrow-level

Evaluate the attention degree (transition of the attention) by assuming three objects

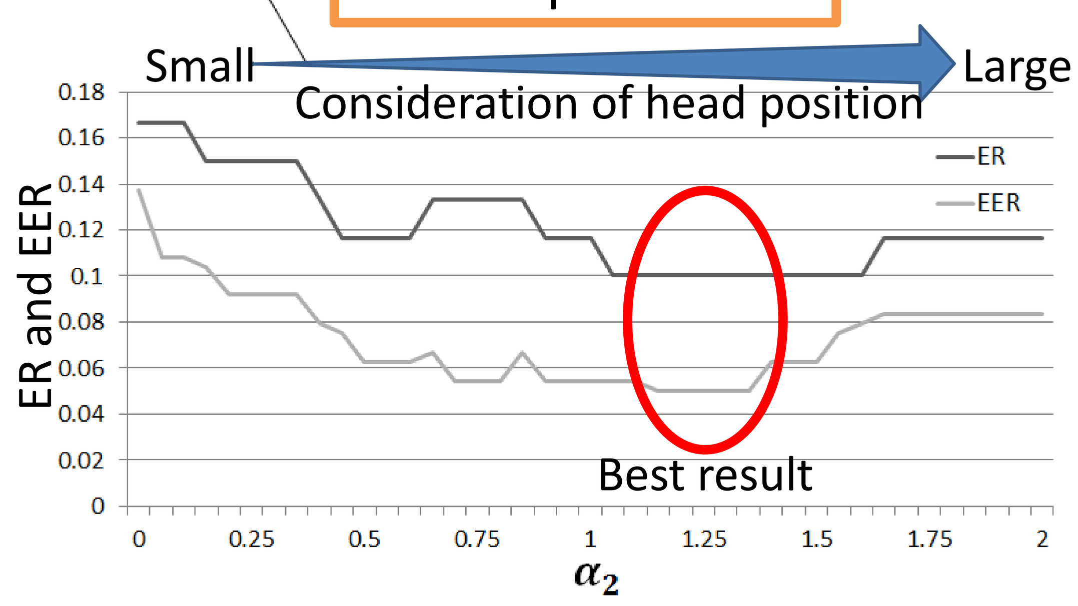


- We can see that the attention degree is high when the user is gazing at the corresponding target area

Evaluate the effect of introducing the location kernel k_2



Stand at each location and gaze at each point



- It is important that choosing appropriate value of α_2

Conclusion and Future Work

✓ We proposed a view-based method for sensing human visual attention
✓ We evaluated the effectiveness of proposed method including the robustness to the head position
☐ We will evaluate the proposed method in detail by using more participants