Sensing Visual Attention by Sequential Patterns

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

***Abstract*—A method for sensing human visual attention is proposed. The method is based on the analysis of sequential image patterns of faces and irises observed at regular time intervals. The basic concept is to represent the set of image patterns produced by the action of gazing at a certain area as a nonlinear subspace in a high-dimensional pattern vector space. Such a space is called an *attention subspace*. In this framework, an input subspace from an unknown action is classified into an attention subspace of gazing at a certain area or into attention subspaces of gazing at other areas (named *non-attention subspaces*) by measuring the canonical angles between the input subspace and pre-computed dictionary subspaces. To maintain performance even in the presence of head movement, two mechanisms are introduced: 1) the kernel orthogonal mutual subspace method, which is suitable for classifying sets of multiple images; and 2) a kernel function for considering the head position in addition to a kernel function for pixel values. The stable performance of the proposed method including situations with head movements is demonstrated through experiments.**

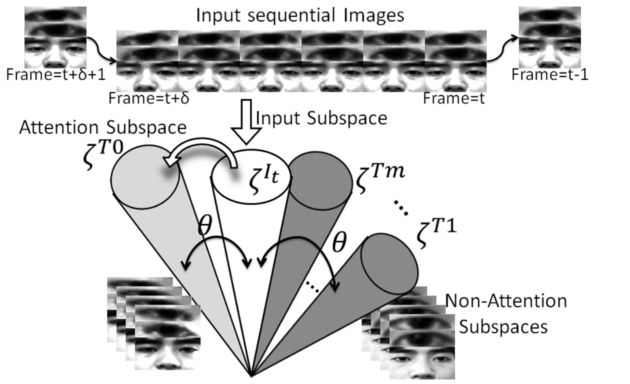


Fig. 1. Conceptual diagram of the process of comparing an input subspace to an attention subspace and non-attention subspaces. The input subspace ζ^{I_t} gradually changes as new input images arrive. The angle $\boldsymbol{\theta}$ indicates the set of canonical angles $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$.

I. Introduction

Recognizing visual attention is useful in various situations in human-computer interfaces [1]–[3] because attention is directly related to the intention of the person. By $visual\ attention$, we mean fixation of gaze on a certain predefined area. Visual attention typically entails the following actions: looking toward a direction of interest first by the movement of the eyeballs, and then gradually rotating the whole face (head) toward the direction [4]. Our ultimate goal is to sense visual attention as a signal for triggering operations in user interface systems [5]. Specifically, our aim is to construct a "vision switch" that operates under the following conditions.

(i) No special equipment is needed, and no constraints exist on head movement.

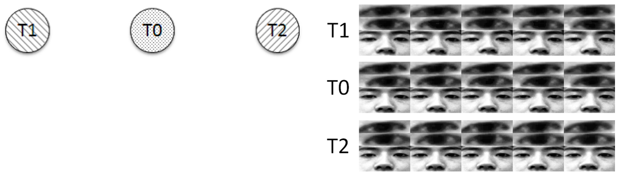
(ii) No special lighting such as infrared radiation is needed.

Although there are no existing techniques capable of sensing human visual attention under the above two conditions (to the best of the authors' knowledge), there have been some studies into gaze detection that are related to our work. For example, in view-based methods [6]–[8], gaze is estimated from the patterns of the eyes or face by using a neural network, morphable models, or other techniques, and the focus of attention is determined from temporal changes in the obtained direction of gaze [9]. The advantage of view-based methods is that they can be realized by simple algorithms that do not necessarily have a 3D model of the eyeball or head. The proposed method is also view-based in terms of images, which consist of patterns of the iris and of the whole face to take the directions of both the gaze and the face into consideration.

We consider the problem of sensing visual attention as a classification of an input sequence of patterns into different classes (attention and non-attention). The attention class is represented by the sequences of patterns that occur due to the action of gazing at a certain target area. Non-attention classes are represented by the sequences of patterns that occur due to the actions of gazing at other areas around the target area. We represent the sequences of image patterns of each action in visual attention by nonlinear subspaces (we call them $attention\ subspace$ and $non\text{-}attention\ subspace$) in a feature space.

A conceptual diagram of the proposed framework is shown in Fig. 1. The *attention subspace* ζ^{T0} for a target area $T0$ can be obtained by applying kernel principal component analysis (KPCA) [10] to the learning patterns collected from the actions of visual attention. The subspace-based representation [11]–[14] enables us to handle the problem of sensing visual attention efficiently by the following simple algorithm. We recognize an unknown action observed in a time interval as attention to a certain area when the subspace constructed by the input pattern is closer to a certain *attention subspace* ζ^{T0} than to any of the other *non-attention subspaces* $\zeta^{T1}, \ldots, \zeta^{Tm}$ for areas $T1, \ldots, Tm$. We define the distance between subspaces by the canonical angles [15] between them, which are obtained by simple linear algebra.

Although subspace methods are capable of recognizing an object without using complex 3D models, they tend to be sensitive to head movement when applied to recognizing attention. This sensitivity is partly because of variations in the view of the irises and whole face. Suppose a person gazes



(a) (b)

Fig. 2. (a) Example configuration for three areas. The action of gazing at $T2$ is regarded as visual attention. (b) Iris/face appearance patterns corresponding to attention to each area.

alternately at two different objects. The difference between the iris/face patterns obtained when gazing at one object and when gazing at the other object will be very small. However, these patterns can vary significantly depending on the head position of the person. This suggests that these patterns are dominated by head movement. Consequently, the performance of attention sensing is greatly reduced by even slight changes in head position. To mitigate this problem, we introduce the kernel orthogonal mutual subspace method (KOMSM) [14] to improve the separability of the subspaces and a kernel function for head location to improve the robustness against differences in head position. By introducing these mechanisms, we are able to produce stable and accurate classification between gazing and not gazing, which paves the way toward the realization of a "vision switch".

The remainder of this paper is organized as follows. In Section II, we outline the basic notions of the proposed method. Section III describes the theoretical framework of the proposed method. An experimental evaluation of the method is presented in Section IV. The last section is devoted to concluding remarks.

II. Basic Idea for Building the Vision Switch

In this section, we briefly introduce the basic notions and approaches for realizing the vision switch.

A. Introduction of Attention Degree

In the formulation of the problem of sensing visual attention to a certain area, we assume a simple situation in which three areas are arranged on a display as shown in Fig. 2 (a) as a running example. In this example, we assume gazing at $T0$ to be attention, and gazing at $T1$ or $T2$ to be non-attention. We note that the number and arrangement of the areas is arbitrary.

We represent the set of input sequential image patterns in the time period from t to $t+\delta$ by an input subspace, ζ^{I_t}, and the sequential patterns due to the action of gazing at objects $T0, T1$, and $T2$ by an attention subspace ζ^{T0} and non-attention subspaces ζ^{T1} and ζ^{T2}, respectively. Although the attention and non-attention subspaces ζ^{T0}, ζ^{T1}, and ζ^{T2} are static, the input subspace ζ^{I_t} is updated whenever a new pattern is input.

The relationship between two subspaces is completely described by the set of canonical angles between the subspaces. If the two subspaces coincide completely, all canonical angles are zero, whereas the canonical angles become larger as the two subspaces drift apart. Therefore, by averaging the canonical angles, we can obtain a value that indicates the distance between two subspaces. We then define the *attention degree* by

$$AD = \frac{1}{L}\sum_{i=1}^{L}\cos^2\theta_{T0i} \tag{1}$$

where θ_{T0i} is the ith canonical angle between the subspaces ζ^{I_t} and ζ^{T0}. Similarly, the *non-attention degrees* are defined by

$$NAD_{Tn} = \frac{1}{L}\sum_{i=1}^{L}\cos^2\theta_{Tni} \tag{2}$$

for object Tn $(n = 1, \ldots, m)$, where θ_{Tni} is the ith canonical angle between the subspaces ζ^{I_t} and ζ^{Tn} and we define $L = \min\{\dim(\zeta^{I_t}), \dim(\zeta^{T\hat{n}})\}$ $(\hat{n} = 0, \ldots, m)$. AD is regarded as the reliability of whether a person is gazing at an object $T0$. We recognize an input action as attention when AD has a significantly high value compared to NAD_{T1} and NAD_{T2}. The validity of this definition of attention degree is discussed in Section IV-B.

B. Problems to Be Solved

To implement our vision switch based on the subspace-based method using sets of sequential iris/face patterns, we have to simultaneously tackle the following problems.

(i) How to emphasize the difference between an attention subspace and other attention (non-attention) subspaces.

(ii) How to reduce the influence of changes in head position on classification performance.

To deal with these problems, we introduce the following two mechanisms.

For problem (i), we adopt KOMSM ([14]) as the classifier. In KOMSM, we represent a set of sequential iris/face pattern images from a gazing action as a nonlinear subspace, which is generated from the set of images by using KPCA [10]. In KOMSM, attention and non-attention subspaces are orthogonalized to each other in advance to increase the classification ability. Orthogonal subspaces have increased inter-class variations; namely, differences between the two subspaces become larger and within-class variation is limited. By doing this, we expect that the similarity between two orthogonal subspaces will not be greatly influenced by head movement. This implies that our method can detect visual attention stably. However, adopting only KOMSM cannot solve the second problem (ii).

Figure 3 shows an example of problem (ii) in which the head position of the user is shifted laterally. In this example, note that although the iris/face patterns are almost the same, the points at which the user is gazing are completely different. This kind of situation causes an overlap between the attention subspace and non-attention subspaces, making classification of both difficult. This observation suggests that we have to consider information about the position of the head in both the learning and test phases. To this end, we introduce a kernel function for considering the position of the head in addition to the kernel function used for representing the nonlinear subspace in KOMSM.

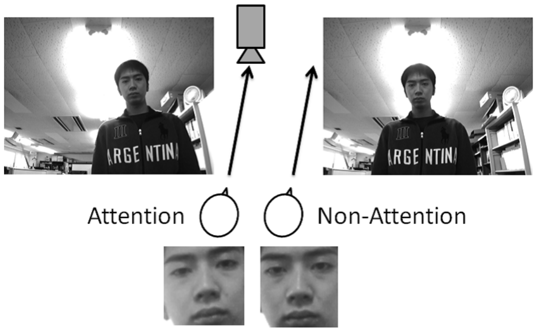


Fig. 3. Ambiguity of the appearance of a face. The two images are almost the same, in spite of gazing at different areas.

III. Theoretical Framework

This section describes a concrete algorithm for calculating the similarity between an input subspace and attention/non-attention subspaces under the framework of KOMSM [14].

A. Canonical Angles between Two Subspaces

We first describe how the distance between two given linear subspaces is calculated, and then explain how this is extended to nonlinear subspaces.

Suppose we already have a pre-computed m_D-dimensional dictionary subspace ζ^D. Given the sets of sequential iris/face patterns, we first represent the sets by an m_I-dimensional linear subspace ζ^I in a d-dimensional space and this subspace is generated from the image sets by using PCA. That is, the subspace ζ^I is the span of the eigenvectors of the empirical covariance matrix of the input samples. In this paper, for the sake of notational simplicity, we identify the subspace ζ by the matrix of eigenvectors that span the subspace. In the following discussion, for convenience, we assume $m_I \le m_D$, and that the m_I canonical angles $\{0 \le \theta_1 \le \cdots \le \theta_{m_I} \le \frac{\pi}{2}\}$ between ζ^I and ζ^D are uniquely defined.

A practical method for finding the canonical angles is by computing the matrix $\boldsymbol{U}^\top \boldsymbol{V}$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{m_I}]$ and $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m_D}]$. Vectors $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$ denote the ith d-dimensional orthonormal basis vectors of the linear subspaces ζ^I and ζ^D, respectively. Here, $\top$ means the matrix transpose. Let $\{\lambda_1, \ldots, \lambda_{m_I}\}$ be the singular values of the matrix $\boldsymbol{U}^\top \boldsymbol{V}$. The canonical angles can be obtained as $\boldsymbol{\theta} = \{\cos^{-1}(\lambda_1), \ldots, \cos^{-1}(\lambda_{m_I})\}$, and these angles are used to calculate the attention degree. The classification method based on the canonical angles (or attention degrees of input) and dictionary subspaces is called the mutual subspace method (MSM).

B. Orthogonalization of Subspaces

To boost the performance of the MSM, class subspaces are orthogonalized by using the framework from Fukunaga and Koontz's method [16] before measuring the canonical angles between them. Suppose there are m different classes. In the framework, the orthogonalization is performed by applying the orthogonalization matrix $\boldsymbol{O}$ to the reference subspaces. If we define the matrix $\boldsymbol{G} = \sum_{i=0}^{m} \boldsymbol{P}_i$ as the sum of the projection matrices $\boldsymbol{P}_i, i = 0, \ldots, m$ corresponding to the projection onto the class i subspace, then $\boldsymbol{O}$ is calculated as

$$\boldsymbol{O} = \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{B}^\top, \tag{3}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the ith largest eigenvalue of $\boldsymbol{G}$ as the ith diagonal component, and $\boldsymbol{B}$ is the matrix in which the ith column vector is the eigenvector of $\boldsymbol{G}$ corresponding to the ith largest eigenvalue. Subspaces $\zeta^{Ti}, i = 0, \ldots, m$ and the input subspace ζ^{I_t} are then transformed by the action of the orthogonalization matrix $\boldsymbol{O}$. That is, we obtain the orthogonalized subspace ζ' from a subspace ζ as

$$\boldsymbol{O} : \zeta \mapsto \boldsymbol{O}\zeta = \zeta'. \tag{4}$$

By this orthogonalization, we can expect better separability between the different subspaces corresponding to different classes. The method of classification according to orthogonalized subspaces is called the orthogonal mutual subspace method (OMSM).

C. KOMSM

A linear subspace is not suitable for representing a set of iris/face sequential patterns due to the highly nonlinear structures of the set. To overcome this problem, OMSM has been extended to a nonlinear method [14]. We introduce a nonlinear map $\boldsymbol{\phi}$ from the patterns $\boldsymbol{x}$ in a d-dimensional input space $\mathcal{I}$ to an f-dimensional feature space $\mathcal{F}$ as $\phi : \boldsymbol{x} \to \boldsymbol{f}$. To perform PCA and whitening on the feature space $\mathcal{F}$, we need to calculate the inner product $\boldsymbol{f}_i \cdot \boldsymbol{f}_j$ between the features $\boldsymbol{f}_i, \boldsymbol{f}_j \in \mathcal{F}$ obtained by mapping $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{I}$ to $\mathcal{F}$. However, this calculation is difficult because the dimension of the feature space $\mathcal{F}$ can be very high. We introduce a "kernel trick", which replaces the inner product $\boldsymbol{f}_i \cdot \boldsymbol{f}_j$ on the feature space $\mathcal{F}$ with the value of a kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. A common choice for the kernel function is the Gaussian kernel:

$$k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma_1^2}\right) \tag{5}$$

with a kernel bandwidth parameter $\sigma_1^2 > 0$.

PCA that uses the kernel trick is called KPCA [10], and the subspace in the feature space $\mathcal{F}$ is called the nonlinear subspace. Note that in KOMSM, each image set is represented by a nonlinear subspace, and the distance between two nonlinear subspaces can be calculated by using this kernel trick. For details, refer to [14].

D. Additional Kernel Function for Head Position

To deal with changes in head position, we introduce an additional kernel function defined as

$$k_2(\boldsymbol{p}_i, \boldsymbol{p}_j) = \exp\left(-\frac{||\boldsymbol{p}_i - \boldsymbol{p}_j||^2}{2\sigma_2^2}\right), \tag{6}$$

where $\sigma_2^2 > 0$ is the kernel bandwidth parameter, and $\boldsymbol{p}_i \in \mathbb{R}^2$ is the vector representing the coordinate (x, y) of the head position in the ith input image.

We propose using a kernel function of the form

$$k(\{\boldsymbol{x}_i, \boldsymbol{p}_i\}; \{\boldsymbol{x}_j, \boldsymbol{p}_j\}) = \alpha_1 k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) + \alpha_2 k_2(\boldsymbol{p}_i, \boldsymbol{p}_j) \tag{7}$$

for estimating subspaces with KPCA. In this kernel function, $\alpha_1, \alpha_2 \ge 0$ are controlling parameters for balancing the importance of the two kernel functions $k_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $k_2(\boldsymbol{p}_i, \boldsymbol{p}_j)$.

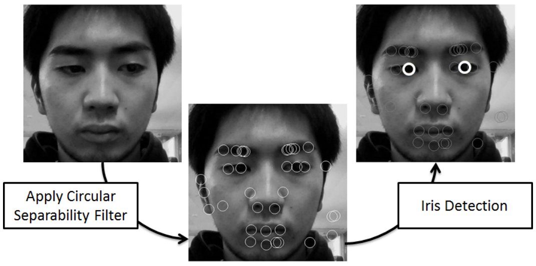


Fig. 4. Flow of iris detection procedure by using the circular separability filter.

Fig. 5. Extracted sequential patterns collected by giving visual attention to the object. To enhance the effects of the images around the irises, we extract the region around the irises and stack them above the whole face image.

E. Algorithm of Sensing Attention

Finally, we summarize the proposed algorithm for sensing attention in the following steps, as shown in Fig. 6. We consider the cases where the attention and non-attention areas are arranged as shown in Fig. 7 (a).

1. Detect two irises from an image for registration. Fig. 4 shows examples of irises detected by using a circular separability filter [17], [18], which can extract circular objects with high speed and accuracy from a given image.
2. Extract the patterns of the irises and normalize the whole face pattern. As shown in Fig. 5, we extracted the region around the irises and stacked them over the whole face image to enhance the importance of the iris area. The size of the extracted pattern is 32×32 pixels.
3. Using images for registration, perform KPCA on image sets for each class and generate subspaces $\zeta^{Ti}, i = 0, \ldots, 8$, then transform the subspaces $\zeta^{Ti}, i = 0, \ldots, 8$ by the orthogonalization matrix $\boldsymbol{O}$ to obtain the corresponding orthogonalized subspaces $\zeta^{T0\prime} \ldots \zeta^{T8\prime}$.
4. Whenever a new pattern is given, patterns are extracted in the same manner as in step 2 and the input subspace ζ^{I_t} is created by KPCA followed by orthogonalization.
5. Calculate the attention degree for all the objects from the orthogonalized input subspace $\zeta^{I_t\prime}$ and all the orthogonalized attention subspaces, $\zeta^{T0\prime} \ldots \zeta^{T8\prime}$.
6. The input action is recognized as visual attention on $T0$ when AD is the highest of all the $NAD_{Tn}(n = 1, \ldots, 8)$ values.

We used the input and all the attention subspaces with a reduced number of dimensions to calculate the canonical angles efficiently. In our preliminary experiments, we found that we can obtain good classification performance when the number of dimensions of the input subspace is set to $d = 5$.

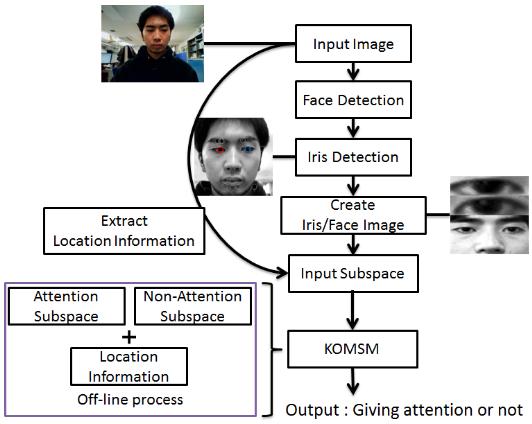


Fig. 6. Flowchart of the procedure for sensing visual attention.

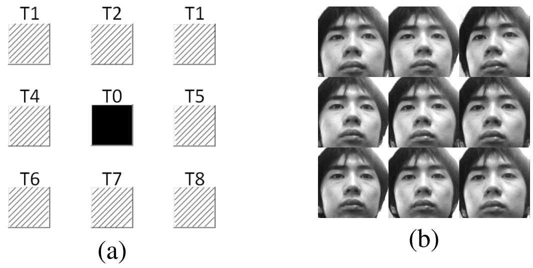


Fig. 7. Setting of attention and non-attention areas. (a) The attention area and eight adjacent non-attention areas. (b) Face appearance patterns corresponding to attention to each area.

Attention and non-attention subspaces are set to $d = 15$ for linear classifiers and $d = 45$ for KOMSM.

IV. Experiments

In this section, we perform several experiments to evaluate the performance of the proposed method. First, we evaluate the basic performance to confirm that the problem can be solved in our framework. Second, we estimate AD and NAD and then evaluate the validity of these estimates. Finally, we evaluate the robustness of the proposed method with respect to the position of the head and examine the effect of the location kernel.

A. Basic Performance

Casting the gaze detection problem as a classification problem as explained in the previous sections, we evaluate the performance of the proposed method when objects that are assumed as attention and non-attention are placed near each other. We first describe the location and the size of the objects in this experiment. There are 9 objects placed as shown in Fig. 7 (a). A user is instructed to gaze at the object displayed on a monitor. In this experiment, we considered 3 different types of distances between objects. Table I shows the details of the setting. Note that we measure the distance between subspaces as the minimum canonical angle in this experiment.

Fig. 7 (b) shows the face images when the distance level is set to the narrow level. The distance between the user and the monitor is set to 50 cm. The camera is located on top of the monitor. We do not fix the position of the head by any special

TABLE I. DISTANCE LEVELS AND CORRESPONDING HORIZONTAL AND VERTICAL DISTANCES BETWEEN EACH OF 9 ATTENTION AREAS.

Distance level	Horizontal distance	Vertical distance
Wide	15.4 cm	11 cm
Middle	8.8 cm	6.6 cm
Narrow	2.2 cm	2.2 cm

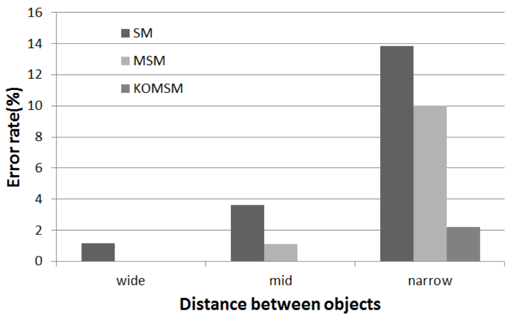


Fig. 8. Error rates for each distance level between objects.

devices. We collected 10 sequences of images, giving visual attention to each object. Throughout this paper, one sequence is assumed to be composed of 20 images, which means that we use 20 images when estimating a subspace. The total number of sequences of data is 90. We use the subspace method [11], MSM [12], and KOMSM as classifiers. In KOMSM, we did not use the location kernel. Instead, we used only a kernel function k_1 with a bandwidth parameter $\sigma_1 = 0.1$. The average error rate for each classifier is calculated by repeating the following procedure 10 times.

- Collect one sequence for all the areas $(T0, \ldots, T8)$ and generate nine input subspaces for them.
- Generate dictionary subspaces corresponding to each area by using the rest of the sequences.
- Calculate the canonical angles between dictionary subspaces and input subspaces.

The experimental results are shown in Fig. 8. From these results, we can see that KOMSM outperformed the other methods and gave perfect classification for the wide- and middle-level distances. Even when the distance between the areas was narrow, KOMSM obtained 97.8% classification accuracy.

B. Investigation of Attention Degree

We next estimate the attention degree, defined in Section II-A, and experimentally support the validity of the notion of the attention degree for measuring whether the user is gazing at a certain object.

Fig. 9 illustrates the setting used in the experiment. The areas $T0$, $T1$, and $T2$ are defined as the attention areas. We put a camera at the same location as $T0$. The distance between the attention area $T0$ and each adjacent non-attention area ($T1$ and $T2$) is 5 cm. The distance between the user and the camera is set to 75 cm, and the head position of the user is fixed in front of the camera. For estimating the dictionary

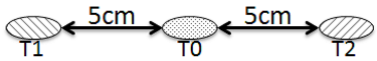


Fig. 9. Setting of the attention area and two adjacent non-attention areas.

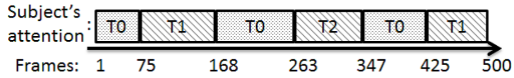


Fig. 10. Transition of subject's attention.

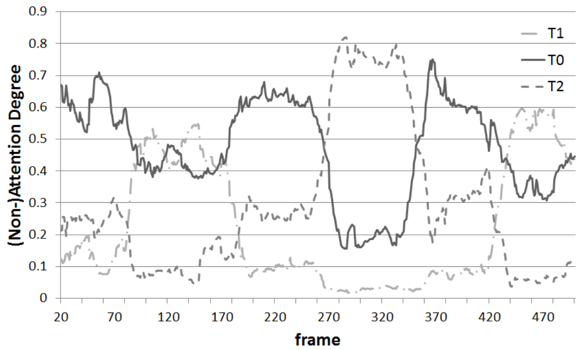


Fig. 11. Attention degree: from frame 20 to 146, giving attention to $T0$; from frame 147 to 262, giving attention to $T1$; from frame 263 to 384, giving attention to $T0$; from frame 385 to 400, giving attention to $T2$.

subspaces, 5 sequences of images of gazing at three areas $T0$, $T1$, and $T2$ were collected. In the testing phase, we recorded a video where a user is gazing at the three areas in turn, as shown in Fig. 10. From the recorded video, we sequentially estimated input subspaces in the following manner. For the initial input subspace, we used the first 20 frames of the video to generate an initial input subspace. We then updated the input subspace using the next frame and discarded the oldest frame. We performed this subspace updating for each new frame in the video and computed the attention degree with respect to the subspaces corresponding to $T0$, $T1$, and $T2$. The resulting attention/non-attention degrees AD, NAD_{T1}, and NAD_{T2} are regarded as a three-dimensional time series, as shown in Fig. 11. From this graph, we can see that the attention degree is high when the user is gazing at the corresponding target area and is low when the user turns their eyes away from the target area.

C. Stability against the Head Position

In this section, we evaluate the robustness of the proposed method against changes in head position and examine the effect of introducing the location kernel k_2. For this purpose, we collected images with various head positions by taking the images of a user standing at the points shown in Fig. 12 and gazing at the attention and non-attention areas shown in Fig. 9. Note that we measure the distance between subspaces as the minimum canonical angle in this experiment. We collected one sequence from each standing point and each gazing target, which amounted to 60 sequences in total. When we collected the images, we also recorded the position of the user as

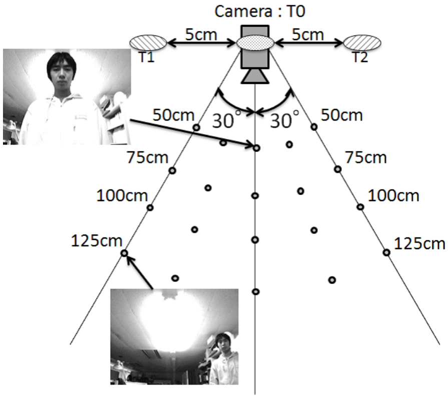


Fig. 12. Head positions for evaluation.

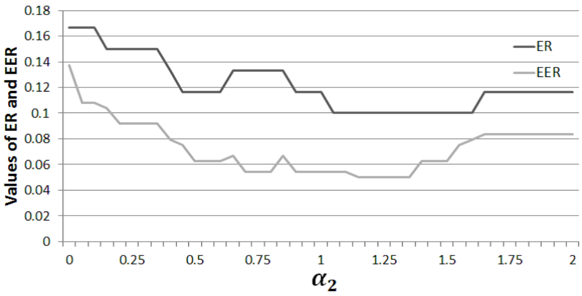


Fig. 13. Effectiveness of the location kernel.

reference data for use in the location kernel. We fixed the parameter $\alpha_1 = 1$ in Eq. 7, and varied α_2 from 0 to 2 in steps of 0.05 to evaluate the effect of the location kernel. For the bandwidth parameters, we set $\sigma_1 = 10^{0.2}$ and $\sigma_2 = 10^{-0.2}$, which we estimated on the basis of preliminary experiments. We evaluated error rates (ER) and equal error rates (EER) of the classification by our method with kernel functions corresponding to different α_2 values as follows.

- Since we have three sequences gazing at $T0$, $T1$, and $T2$ for each position, we choose three sequences corresponding to one position and use them for the input subspaces.
- Construct dictionary subspaces for the attention area and the two non-attention areas using the remaining 57 sequences, and calculate the canonical angles between the input subspaces.

We performed this procedure 20 times using different positions as input data. Fig. 13 shows the estimated ERs and EERs in terms of α_2. When $\alpha_2 = 0$, we could not obtain good performance because we did not use the head position information in the classification, and similar sequences with different head positions were misclassified. However, when the location kernel was introduced and α_2 was set to an appropriate value, the classification performance improved since the head position information was also used in the classification. However, large values of α_2 caused poor performance. This suggests that it is important to control the kernel combination parameters.

V. Conclusions

We proposed a view-based method for sensing human visual attention by analyzing sequences of image patterns of faces and irises without any restraint on head movement. The proposed method is capable of sensing attention efficiently by using the framework from KOMSM. In addition, to remove the effects of head movement, we introduced the notion of a location kernel to KOMSM. Experiments using real data showed that the proposed method is robust to head movement. In future work, we will evaluate the proposed method in detail by using more participants in more complicated situations.

Acknowledgments

This work was supported by JSPS KAKENHI for Scientific Research (B) Grant Number 24300290 and by JSPS KAKENHI for Young Scientists (B) Grant Number 25870811.

References

[1] D. W. Hansen, and Q. Ji, "In the Eye of the Beholder: A Survey of Models for Eyes and Gaze", *IEEE Transactions on PAMI*, vol. 32, no. 3, pp. 478-500, 2010.

[2] R. J. K. Jacob, "What You Look at Is What You Get: Eye Movement-based Interaction Techniques", *CHI'90*, pp. 11-18, 1990.

[3] A. Sharma and P. Abrol, "Eye Gaze Techniques for Human Computer Interaction: A Research Survey", *IJCA*, vol. 71, no. 9, pp. 18-29, 2013.

[4] P. Morasso, E. Bizzi, and J. Dichgans, "Adjustment of Saccade Characteristics During Head Movements", *Experimental Brain Research*, vol. 16, no. 5, pp. 492-500, 1973.

[5] T. Chino, K. Fukui, and K. Suzuki, "GazeToTalk: A Nonverbal Interface with Meta-Communication Facility", *ETRA*, pp. 111, 2000.

[6] S. Baluja and D. Pomerleau, "Non-Intrusive Gaze Tracking Using Artificial Neural Networks", *CMU CS Technical Report*, CMU-CS-94-102, 1994.

[7] T. D. Rikert and M. J. Jones, "Gaze Estimating using Morphable Models", *FG'98*, pp. 436-441, 1998.

[8] F. Lu, T. Okabe, Y. Sugano, and Y. Sato, "A Head Pose-free Approach for Appearance-based Gaze Estimation", *BMVC*, pp. 126.1-126.11, 2011.

[9] R. Stiefelhagen, M. Finke, J. Yang, and A. Waibel, "From Gaze to Focus of Attention", *VISUAL'99*, pp. 761-768, 1999.

[10] B. Schölkopf, A. Smola, and K-R. Müller, "Nonlinear Principal Component Analysis as a Kernel Eigenvalue Problem", *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.

[11] E. Oja, "Subspace Method for Pattern Recognition", *Research studies Press Ltd.*, 1983.

[12] K. Maeda and S. Watanabe, "A Pattern Matching Method with Local Structure", *Transactions on IEICE*, vol. J68-D, no. 3, pp. 345-352, 1985 (in Japanese).

[13] O. Yamaguchi, K. Fukui, and K. Maeda, "Face Recognition Using Temporal Image Sequence", *FG'98*, pp. 318-323, 1998.

[14] K. Fukui and O. Yamaguchi, "The Kernel Orthogonal Mutual Subspace Method and Its Application to 3D Object Recognition", *ACCV*, vol. 4844, pp. 467-476, 2007.

[15] H. Hotelling, "Relation Between Two Sets of Variates", *Biometrica*, vol. 28, no. 3, pp. 322-377, 1936.

[16] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen-Loéve Expansion to Feature Selection and Ordering", *IEEE Transactions on Computer*, vol. C-19, no. 4, pp. 311-318, 1970

[17] K. Fukui and O. Yamaguchi, "Facial Feature Point Extraction Method Based on Combination of Shape Extraction and Pattern Matching", *Systems and Computers in Japan*, vol. 29, no. 6, pp. 49-58, 1998.

[18] Y. Ohkawa, C. H. Suryanto, and K. Fukui, "Fast Combined Separability Filter for Detecting Circular Objects", *MVA*, pp. 99-103, 2011.