Expectation maximization of prefrontal-superior temporal network by indicator component-based approach

T. Koshizen, Bernd Heisele, Hiroshi Tsujino

Honda Research Institute Co. Ltd.,
8-1 Honcho, Wako-Shi, Saitama-ken, 351-0114 (Japan Office),
145 Tremont Ave., Boston, MA., 02111-1208 (Americas Office)
koshiz@jp.honda-ri.com

Abstract

In CNS*02, we hypothetically provided the cross-supramodal integration system, which is hypothetically dedicated by a bidirectional neuronal network where prefrontal cortex (PFC) is interacted with hippocampus (HC) in order to calculate the coherent relationship between the supramodalities. The parameter to learn the coherence may also be dedicated by Anterior Cingulate (AC) cortex. In this paper, we attempt to propose top-down attention control system based on the outcome that we presented at CNS*02. That is, PFC is presumably interacted with superior temporal (ST) neuron in order to extract indicator component representing the whole view of face or object, by maximizing the expectant value where attention modulation can be taken into account of distinguishing different faces. As a result, we postulate an objective of the top-down attention control is essentially to compute the abstraction of face/object information based on the Expectation Maximization (EM) algorithm since the voluntary movements of facial viewpoints must play an important role of integrating spatial and temporal property.

Key words: Prefrontal and Superior-temporal Cortex, Distributed Cortical Neuronal Network, Spatiotemporal Attention, Cross Supra-modality, EM Algorithm

1 Introduction

PF cortex is involved in a broad array of cognitive functions, including learning, memory, attention, executive function, planning, and judgment [1]. It is known that PFC has also the executive committee by consolidating hippocampus and other cortical areas. In CNS*01, we suggested the computational

model of anterior-cingulate (AC)-prefrontal-posterior parietal (PP) cortex in relation to attention demanding and modulation process [2]. Conclusively, this work has indicated the aim of PFC-based executive neuronal network was to compute the spatiotemporal attention mechanism. Furthermore, we in CNS*02, presented the computational model of spatiotemporal attention to learn the coherent relationship across the cross-supramodal integration [3]. From recent neurobiological results, it is indicated that the rewiring mechanism can be induced by the inhibitory neuronal network onto excitatory neurons of cerebral cortex where PF neuron may be crucial role of representing the supramodal density and the cross-supramodal correlation by interacting AC cortex. In this paper, we propose the top-down attention control scheme that is hypothesized the computation with respect to the connection of PF and superior temporal (ST) cortex. ST cortex is presumably known a conjunctive point between the 'dorsal' stream specialized by motion (temporal attention) property and the 'ventral' stream (what) specialized by form (spatial attention) property. Additionally, the ST is known for the multimodal response as well as the cross-modal response when interacting with posterior parietal (PP) cortex. Importantly, recent physiological results also demonstrated the neurons in ST involve in the computation of face perception [4]. This is because the two core areas of face processing engage in the categorization of a stimulus as a face, and the identification of a specific individual, by the ST neuron incorporated with inferior temporal (IT) neuron, in order to implement the facial perception consisting of the 'rough' and the 'finer' computations through their neurons belonging in PF and Orbito-Frontal (OF) cortex. The PF-based executive neuronal network involving ST cortex is hypothetically represented as shown in Fig.1, and AC, Basal Ganglia (BG) and Hippocampus (HP) are taken part in the network. The key idea of our speculation is, PF cortex, which may expertise the top-down attention control where attention modulation allows the maximum expectant to yield the semantic abstraction representing specific face or object information in accordance with the supramodal computation. In this sense, the computational role of top-down attention may be implicated to calculate the expectation value where the visual motion may be crucial for maximizing it by adding the biasing signals to the visual form. It evokes generalization properties of biological motion perception using a new class of stimuli that were generated by the spatial and temporal characteristic of morphing among different viewpoint patterns. It has been demonstrated using several monkey's experiments shows that the PF neuron involves in the reward-based learning where spatial information becomes more accurate when reward outcome is expected; more accurate representations of spatial information would as a consequence lead to more accurate behavior, e.g., [5]. In this paper, we suggest the top-down attention control scheme based on the PFC-STS network related to a biological Expectation Maximization (EM) learning algorithm to extract the indicator component based on the top-down attention control.

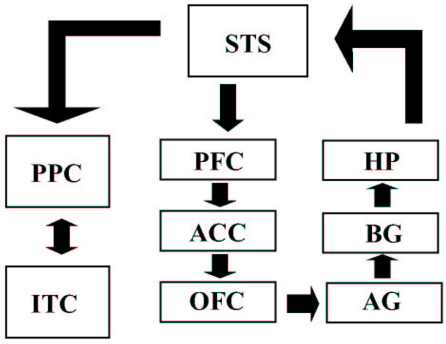

Fig. 1. Hypothetical connection of PFC and STS

2 Computational Model

In this section, we describe the computation model of our proposal, called indicator facial component-based learning approach. The overview of proposed system is shown in Fig.2 (left) where the single component classifier initially developed by [5] is hierarchically reorganized by the multiple component classifiers using Support Vector Machine (SVM) training algorithm which performs pattern recognition for a two-class problem by determining the separating hyperplane that has maximum distance to the closest points of the training set. The closest points, the maximum distance are called Support Vector (SV) and Margin respectively [6]. In our framework, the proposal system aims to extract the indicator facial component that is used to abstract the informative source to distinguish different faces. Face component-based learning algorithm aims to detect faces of different sizes and arbitrary positions in a gray value input image [4]. More precisely, Fig.2. (Right) shows the schematic drawing where component classifiers independently detect components of the face on the first level. This classifier allows the components to extract the features around eyes, nose and mouth. On the second level, the geometrical configuration classifier performs the final face detection by linearly combining the results of the component classifiers. We eventually obtain the output of SVM component classifier indicates if there is a face inside the window or not. Generally, one of the main problems in the component-based object recognition is the selection of the components; how to find discriminated components that allow to distinguish a particular object from rest. To address the question, we employ the proposed system shown in Fig.2 where top-down attention control must be taken into account of maximizing the expectation value where the viewpoint of faces are fluctuated by attention modulation to differentiate the semantic abstraction of indicator component.

Let the expectation value $\mathcal{E}$ be

$$\mathcal{E}(\Omega_\tau) \approx \int logP(\mathcal{O}, \Omega; \tau) d\mathcal{O} \tag{1}$$

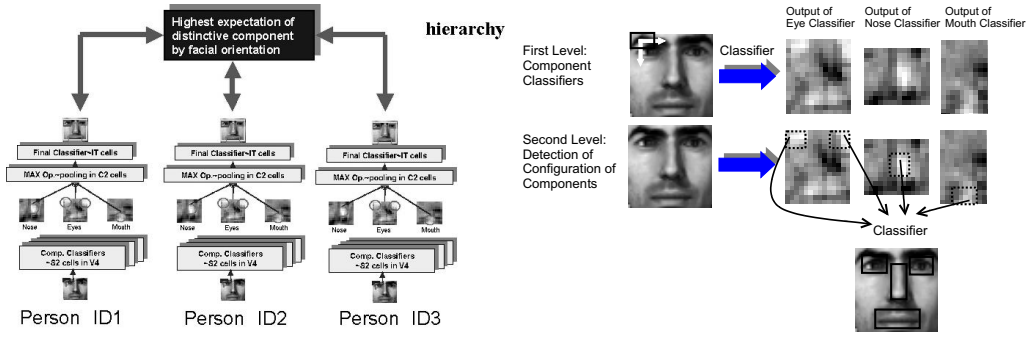

Fig. 2. Proposed multiple classifier system to attain indicator component(Left) and conventional two-level single classifier system (Right)

where Ω represents the probabilistic attention variables *e.g.* see [2] in accordance with certain viewpoints τ. $\mathcal{O}$ is the outputs that are calculated from each SVM component classifier shown in Fig.2 (Left). Additionally, to calculate the probability density function $P(\mathcal{O}, \Omega; \tau)$ we assume the hypothetical mapping $h : \mathcal{O} \mapsto \Omega$. The mapping h will be mathematically defined as,

$$P(\Omega_\tau) \equiv P(y = 1|\mathcal{O}_\tau) = \frac{1}{1 + exp(\mathcal{O}_\tau/\gamma_\tau)} \tag{2}$$

where, y is the classification label of positive examples and $\gamma \in \Omega$ is the predictive parameter of attention modulation to extract the indicator component that maximizes the expectation value $\mathcal{E}$. γ_τ represents the correspondance between the supramodalities such as visual form and motion. Importantly, in our framework the attention Ω can be regarded as the kind of 'hidden variables' calculating to determine the optimal subset of ranked features that is learned from each SVM component classifier. The computation of Eq. (2) is initially suggested by [8]. Mathematically, γ_τ is somehow related to the differentiation of the expectation $\mathcal{E}$ over Ω_τ,

$$\gamma_\tau = \frac{\partial \mathcal{E}(\Omega_\tau)}{\partial \Omega_\tau} \tag{3}$$

Note that the right side of Eq.(3) computes the functional splope of the expectation value $\mathcal{E}$. In principle, we can maximize the logarithm of the joint distribution (which is proportional to the posterior):

$$\Omega_{\tau+1} = argmax_{\Omega_\tau} \quad \mathcal{E}(\Omega, \Omega_\tau) \tag{4}$$

where τ denotes the parameter representing the viewpoint of a facial component. It is important to remind that the expectation value $\mathcal{E}$ is calculated in the **E**-step by evaluating the current guess Ω_τ where in the **M**-step we are optimizing $\mathcal{E}(\Omega, \Omega_\tau)$ with respect to the *free variable* τ (facial viewpoint) to accordingly obtain the new estimate $\Omega_{\tau+1}$.

To implement our indicator component-based approach, training images are captured over the circumstances of various illuminations and the unique (black) background. After the images are collected, pixel values are used as inputs to each layer of a SVM component classifier as shown in Fig.2. The cropped image is then converted into gray values and is re-scaled to 40×40 pixels. Histogram equalization is also applied to remove variations existing in image brightness and contrast. The 1,600 gray values of each face image are then normalized to the range between 0 and 1. Each image is represented by a single feature vector of length 1,600 - the total number of pixels in the image. These feature vectors serve as the inputs to the SVM face classifier during the training process. With respect to the training dataset it includes 974 images of all six subjects in our database. The rotation in depth is again up to about $\pm 42°$. Fig.3 (left) implicates the facial viewpoints has the rotation by right to left, or left to right within $-10°$ to $+10°$. By contrast, the rotation of facial viewpoints is only left by $+12°$ to $+42°$ in Fig.3 (right). Furthermore, Fig.4 shows the expectation values in both cases, and which are calculated based on Eq.(1). Conclusively, their results demonstrate that the expectation value is used for selecting component features by qualifying the input data structure in accordance with different facial viewpoints. We suggest the computation of top-down attention control by PF neuron may be originated to maximize the expectation value where attention class Ω is modulated over different component features, in order to restrict the quantity and quality of training data mapping into the feature space in order to find the optimal subset of selected features, as the indicator component.

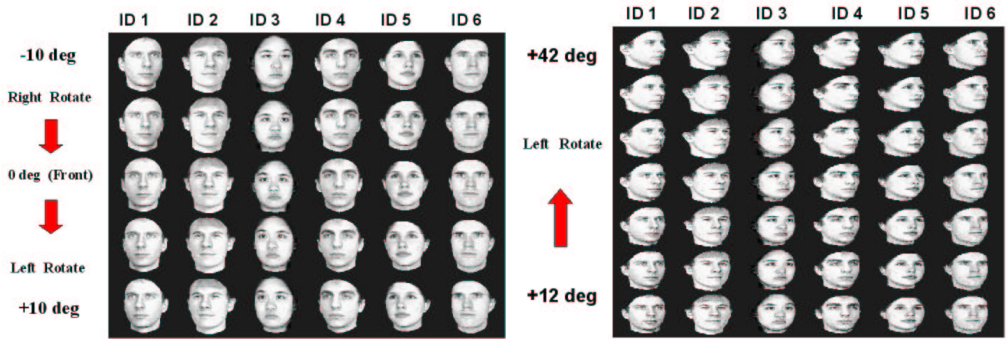

Fig. 3. Facial movement pattern by $-10°$ to $+10°$ (Left) and $+12°$ to $+42°$ (Right)

3 Discussion and Future work

We present the proposed learning system as it relates to component-based face recognition where top-down attention control can be taken into account of facial movements, which are dedicated the possibility that is 'experience-dependent' learning with the intrinsic nature of facial perception. Additionally,

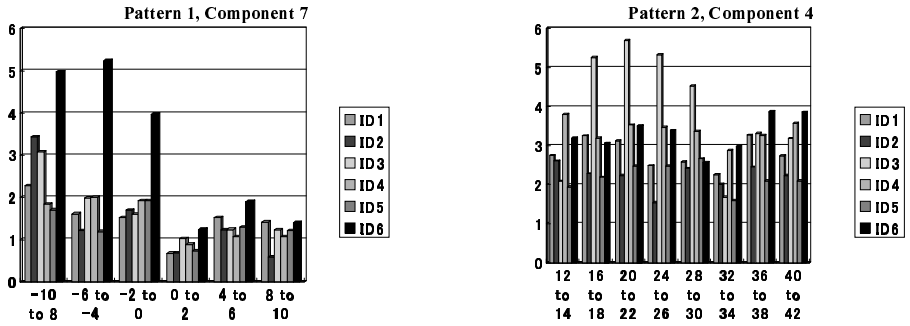

Fig. 4. Histogram shows the learning result by SVM. Vertical axis denotes the margin, while horizontal axis represents the viewpoint discretized by every 2°. ID6 shows the most distinction for the left-side nose component (Left) with $-10°$ to $+10°$. By contrast, ID3 shows the most distinction for the right-side eye component (Right) with $+12°$ to $+42°$

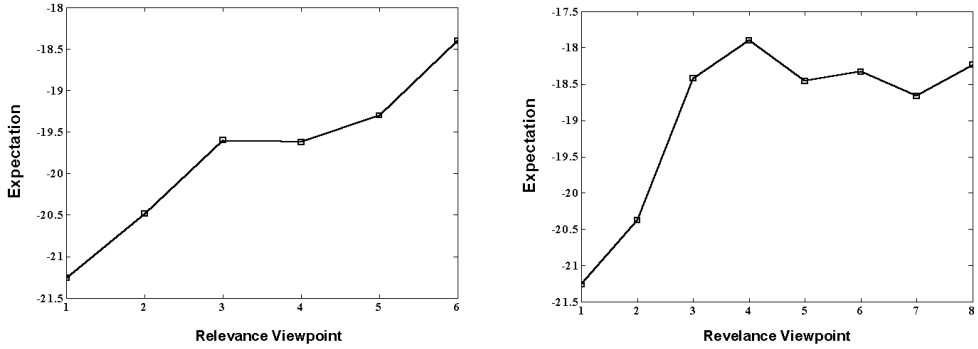

Fig. 5. Expectation value calculated by Eq.1 under the rotation of $-10°$ to $+10°$ (Right) and $+12°$ to $+42°$ (Left)

we assume faces may engender both face-specific and general object processing, and that face-specific processing might be revealed only if the general object system was occupied by concurrent object processing. That is, it is possible that the general object recognition system is itself not *monolithic*, and may become differentiated by experience. It hypothetically allows us to provide the neuronal computation across PFC and STS/IT. Prior to the mechanism of human face recognition, many studies so far indicated that it is the intrinsic

nature itself rather than the experience-dependency. However, several lines of evidence make this conclusion unlikely based on infant experience of face processing [9][10][11][12]. Furthermore, the development of face processing has indicated its significance for social cognition. For example, infants orientate more rapidly to peripheral visual targets when cued by the direction of eye gaze presented by face [13]. Semantic priming refers to the fact that recognition of a word or object of a particular category *e.g.*, animals, is better and faster when preceded by a stimulus of the same category *e.g.*, cat proceeded by dog than when preceded by a stimulus of a different category *e.g.*, cat preceded by pencil. Behavioral studies of face priming indeed have been carried out to infer the processes involved in face recognition. We showed the PFC-STS model suggests the EM algorithm that is basically originated from top-down attention control process biased by visual motion. In our framework, the top-down attention regulates each SVM component classifier by calculating the expectant value to extract indicator component. Future work may allow us the top-down attention to be applying in emotional perception and even in more general-domain for object perception by extending the class of indicator component which is obtained from animated faces.

References

[1] Miller, E.K. and Cohen, J.D., (2001), *An integrative theory of prefrontal cortex function*, Annual Review Neuroscience, Vol.24, pp.167-202.

[2] Koshizen, T., Akatsuka, K. and Tsujino, H., (2002), *A computational model of attentive visual system induced by cortical neural network*, Neurocomputing, Vol.44-46C, pp.879-885.

[3] Koshizen, T., Yamada, S. and Tsujino, H., (2002), *Semantic rewiring mechanism of neural cross-supramodal integration based on spatial and temporal properties of attention*, Neurocomputing, to appear.

[4] Haxby, J.V. et al., (2000), *The distributed human neural system for face perception*, Trends Cognitive Science, Vol.4, pp.223-233.

[5] Kobayashi, S., Lauwereyns, J. Koizumi, M., Sakagami, M. and Hikosaka, O, (2002), Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex, J. of Neurophysiology, Vol.87, pp. 1488-1498.

[6] Heisele, B., Poggio, T. and Pontil, M., (2000), "Face detection in still gray images", A.I. memo 1687, Center for Biological and Computational Learning, MIT, Cambridge, MA.

[7] Vapnik, V., (1998), "Statistical learning theory", John Wiley and Sons, New York.

[8] Wahba, G., (1999), *Support vector machines, reproducing kernel hilbert spaces and the randomized GACV*, In B. Scholkopf, C.J.C. Burges, and A.J.Smola, editors, Advances in Kernel Methods - Support Vector Learning, pages 69-88., Cambridge, MA, 1999. MIT Press.

[9] Valenza, E., Simion, F., Machi Cassia, V., Umilta, C. (1996), *Face preference at birth*, JEP:HPP, Vol. 22, pp.892-903.

[10] Simion, F, Valenza, E., Umilta C. and DallaBarba, B. (1998), *Preferential orienting to faces in newborns: a temporal-nasal asymmetry*, Journal of Exp Psyhol Hum Percept Perform, Vol.24, pp.1399-1405.

[11] Farah M.J., Rabinowitz, C., Quinn, G.E. and Lui, G.T., (2000), *Early commitment of neural substrates for face processing*, Journal of cognitive neuropsychology.

[12] de Schonen S, Mancini, J., Leigeois, F. (1998), *About functional cortical specialization:the development of face recognition*, In the development of sensory, motor and cognitive capacities in early infancy: from perception to cognition, edited by Simon F, Butterworth G., Hove:Psychology Press, pp.103-120.

[13] Farroni, T., Johnson, M.H., Brockbank, M. and Simion, F. (2000), *Infants' use of gaze direction to cue attention:the importance of perceived motion*, Visual cognition, Vol.7., pp.705-718.