Appears in Proceedings of the Annual Meeting of the Cognitive Science Society, 2016 (CogSci 2016)

Hidden Markov Modeling of eye movements with image information leads to better discovery of regions of interest

Stephan Brueggemann (u3003097@hku.hk)
Department of Psychology, The University of Hong Kong, Pokfulam Road, Hong Kong

Antoni B. Chan (abchan@cityu.edu.hk)
Department of Computer Science, Tat Chee Avenue, Kowloon, Hong Kong

Janet H. Hsiao (jhsiao@hku.hk)
Department of Psychology, The University of Hong Kong, Pokfulam Road, Hong Kong

Abstract

Hidden Markov models (HMM) can describe the spatial and temporal characteristics of eye-tracking recordings in cognitive tasks. Here, we introduce a new HMM approach. We developed HMMs based on fixation locations and we also used image information as an input feature. We demonstrate the benefits of the newly proposed model in a face recognition study wherein an HMM was developed for every subject. Discovery of regions of interest on facial stimuli is improved as compared with earlier approaches. Moreover, clustering of the newly developed HMMs lead to very distinct groups. The newly developed approach also allows reconstructing image information at each fixation.

Keywords: Eye-tracking; Face Recognition; Hidden Markov Model; Machine Learning;

Introduction

Eye movements provide a direct insight into ongoing cognitive processes. Although mental processes cannot be observed per se, recording eye movements is an unobtrusive measurement of what a person is processing at a particular moment. Thus, eye tracking can be used to study attention, memory, language, problem solving and decision making. There exist different approaches to analyze eye-tracking recordings. Below, we summarize common eye movement analysis techniques in face recognition.

Face recognition is a cornerstone process of meaningful social interactions since it helps us to identify familiar individuals irrespective of the viewpoint, lighting conditions and emotional expression of a face. In previous studies, attempts have been made to better understand the spatial and temporal characteristics of face recognition.

Spatial eye-tracking analyses

One goal of studies on eye-tracking in face recognition is to identify which facial regions people look at when successfully recognizing another person. There exist different approaches for doing so. In a region of interest analysis, the percentage of fixations on a predefined region of interest (ROI) is computed (Henderson et al., 2005). However, there exists a lack of an objective way to identify ROIs. Statistical fixation maps aim to close this gap by constructing ROIs in a data-driven way.

A statistical fixation map can be constructed by plotting all fixations of a subject on an average face, with fixations being subsequently smoothed by convolving Gaussian kernels. Using fixation maps, it was shown that more fixations are placed on the nose area and the eye region compared with other areas (Caldara & Miellet, 2011; Hsiao & Cottrell, 2008).

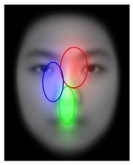
Temporal eye-tracking analyses

Face recognition is a process that can span over several seconds. Therefore it is important to understand the temporal characteristics as well. An early attempt to do so was the scan path theory (Noton & Stark, 1971). According to the theory, fixations are made on facial features in a sequence (scan path). For the same stimulus, the same sequence emerges. Studies have shown that this assumption does not always hold. For example, Walker-Smith and colleagues (1977) showed that the same scan path only emerges about 65% of the time when the same stimulus is presented.

Fixation locations during face recognition do not follow a strict, a priori planned path. More precisely, it has been indicated that saccades are constantly influenced by top-down (Yarbus, 1965) and bottom-up inputs (Mannan, Ruddock, & Wooding, 1997). It can be argued that eye movements can be treated as a Markov stochastic process. In this process, the future state depends only on the current state. Probabilistic time series models are a good fit for understanding eye-movement strategies (Chuk, Chan, & Hsiao, 2014).

Hidden Markov Models

Chuk and colleagues (2014) proposed to use a statistical time series model, the hidden Markov model (HMM) to analyze eye movement recordings in cognitive tasks. The model has several advantages over existing analysis techniques. It combines spatial and temporal analyses, and it produces data-driven ROIs, which can be of different size. Furthermore, the model accounts for individual differences by identifying subject-specific ROIs and transition probabilities.



Prior values	Red	Green	Blue
	0.47	0.35	0.17
Transition probabilities (from\to)	To Red	To Green	To Blue
From Red	0.60	0.21	0.19
From Green	0.42	0.56	0.02
From Blue	0.45	0.06	0.49

Figure 1: The Figure shows a hidden Markov Model with its prior values and transition probabilities. The red, blue and green ellipses on the face represent the hidden states. The table shows the prior values (i.e. the probability that the first fixation is at a given hidden state). The transition probabilities indicate the probability that a subject's fixation moved from one hidden states to another. The Figure is obtained from Chuk and colleagues (2014).

An HMM summarizes a subject's eye-fixation strategies (Figure 1). First, the regions of the face where the most fixations were present are identified (hidden states). In Figure 1, one hidden state is between the eyes, one hidden state is below the right eye and one hidden state covers the lower nose and mouth region. The locations of hidden states are estimated from the fixation locations. The HMM describes the participant's eye fixation strategy by the transition from one hidden state to another hidden state. The HMM specifies the prior probabilities of the initial fixation's hidden states (prior probability vector). The transition matrix describes the probability of moving from one hidden state to another hidden state. More precisely, when a subject's fixation is at a certain hidden state, transition probabilities indicate at which hidden state the next fixation will be.

It was shown that HMMs can successfully model spatial and temporal information and capture individual differences in face recognition strategies (Chuk, Chan, & Hsiao, 2014; Simola, Salojarvi, & Kojo, 2008; Wedel, Pieters, & Liechty, 2003). In Chuk and colleagues (2014), an HMM was developed for each subject to describe the individual's eye-movement patterns. Individual differences were found in both fixation locations and transition probabilities. Moreover, the eye-movement strategies of participants could be classified into one of two groups, namely holistic and analytic, demonstrating individual differences even within the same culture. In addition, it was also shown that correct and incorrect recognitions were associated with distinct HMMs. The main difference between the two groups of HMMs was found in the transition probabilities.

Hidden Markov models with fixation locations and image information

Fixation locations were the input for the HMM model by Chuk and colleagues (2014). 2-D Gaussians were fitted to the fixation locations to identify regions of interest on the face. In the present study, we propose a new way to identify hidden states. We suggest identifying hidden states based on both fixation locations and image information. The newly proposed input features advance the original model in several ways.

First, fixation locations alone are not always the optimal input features because they can be compromised. Recordings with an eye-tracker are not always accurate and over- and undershooting by the eye itself can further introduce noise to the fixation locations. Including image information as an additional input feature ensures that hidden states are identified based on fixation locations and corresponding observed visual stimuli.

Secondly, including image information makes the model more expressive. Fixation location and image information do not always closely correspond. In other words, we are able to better discriminate between fixation locations that are similar regarding their coordinates but correspond to different facial features. Thirdly, face stimuli vary slightly. In Figure 2, it can be seen that a position with a given coordinate can correspond to different facial features. Fixations of a hidden state in the newly proposed HMM will be similar in both fixation location and observed image patch.

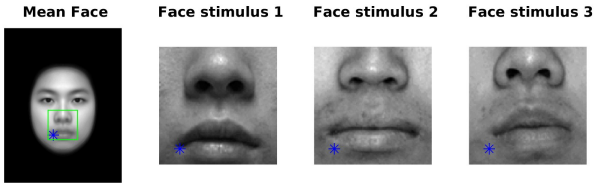


Figure 2: The Figure illustrates the variation in the structure of faces. On the left, the mean average face of all stimuli is shown. The three other images are parts of face stimuli that were used in the experiment. The blue star is located at X=168, Y=363. Although the location is exactly the same, the point belongs to slightly different areas of the face (corner of the mouth vs. area below the mouth).

In short, the newly developed HMM will advance the original model in three ways. First, it allows the identification of regions of interest on the face more accurately. Second, the current approach allows us to read out the image information for every hidden state. We are able to better understand what kind of image information a participant used during face recognition. As a final step, we cluster the newly developed, individual HMMs to investigate group differences.

Methods

Experimental Setup

In the present study, the data by Chuk and colleagues (2014) was re-analyzed. More details can be found in the original paper. In short, participants had to perform a face recognition task. 32 Chinese subjects (16 males) participated in the experiment. The experiment consisted of two blocks. In the training block, participants were instructed to study faces (targets) and in the testing phase they had to recognize the targets among new faces (distractors). In the testing phase, participants indicated whether they recognized a face or not by pressing one of two buttons on a response pad. In order to familiarize participants with the experimental task, they needed to complete a very short practice version of the experiment. The EyeLink 1000 Tower Mount eye tracker was used to record the eye movements from participants. Before the start of the experiment, the standard nine-point calibration procedure was performed. At the beginning of each trial, a drift correction was performed.

In the testing phase, face stimuli were displayed for 5 seconds within which participants were required to respond. The stimulus set consisted of 40 (20 males) gray scale frontal-view Asian faces. The faces had a neutral expression. Stimuli had an inter pupil size of 60 cm and they were all cropped to a size of 320 x 420 pixels. The screen was viewed at a distance of 60 cm. The horizontal visual angle was 6 degrees and the vertical angle was 8 degrees. Faces were aligned by vertical and horizontal eye positions. Participants were not familiar with the face stimuli.

Data Analysis

Input features The fixation location and the corresponding image information were used as input for the model. For every fixation, we extracted a 50x50 pixel image patch around the fixation location on the face stimuli. The image patch was foveated using the formula from Geisler and Perry (1998). More precisely, the spatial resolution of the visual system strongly decreases away from the fixation location. Geisler and Perry (1998) developed a foveated multiresolution pyramid which transforms every image into 5-6 regions of different spatial resolutions (see Figure 3). We performed a Principal Component Analysis (PCA) on all image patches for the purpose of dimensionality reduction. All image patches were converted to vectors. PCA was used to identify its principal components of the set of images. Each image patch was represented with its principal component coefficients, which we used as input features for the HMM.

HMMs with a different number of PCA coefficients can be learned. Firstly, we developed an HMM with a partial representation of image information. 5 coefficients account for 78.43% of variance in the stimuli. For this model, we used the X- and Y-fixation locations and the first 5 PC coefficients as input features. The 5 PC coefficients help to obtain hidden states which tend to be similar in image information.

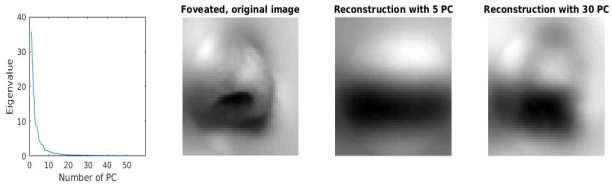
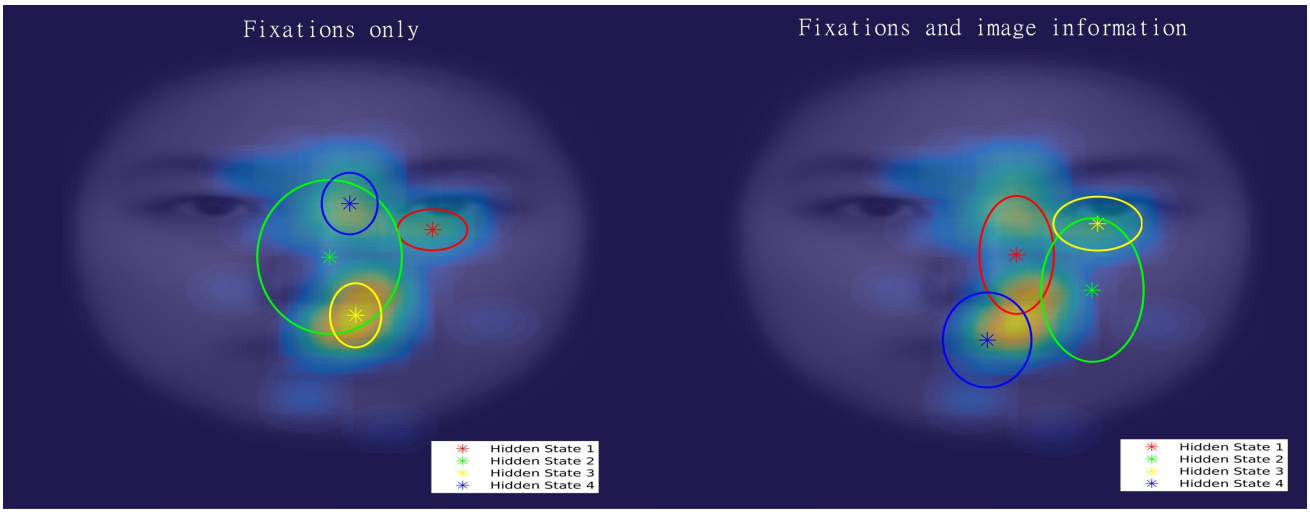


Figure 3: On the left, the eigenvalues for the first 50 eigenvectors are shown. The second image to the left shows an extracted 50 x 50 pixels image patch which was foveated using the formula by Geisler and Perry (1998). The two images on the right show the reconstruction of the original image using 5 and 30 coefficients respectively.

Secondly, we built an HMM with the purpose to obtain improved individual HMMs and to reconstruct the perceived visual stimuli per hidden state. The first 30 components are used as features. They account for 96.62% variance. We matched the number of fixation dimensions with the number of image information dimensions to ensure equal weighting. More precisely, to match the 30-dimensions of the PCA representation, we replicated the X- and Y-coordinates of the fixation 15 times. We did this to balance the influence of the fixations locations and image information on the positioning of the centers of the hidden states. The final feature vector has 60 dimensions. The model with 30 coefficients helps to obtain image patches that are highly similar in image information in a hidden state and furthermore allows reconstructing image information at the fixation.

Hidden Markov Model An HMM was estimated for every participant. Parameters of the HMM were estimated in a two-stage process. Firstly, regions of interest (ROIs) were estimated by learning a Gaussian mixture model (GMM) on the feature vectors. Each Gaussian component in the GMM corresponds to one ROI. The variational Bayesian framework for Gaussian mixture models (VBGMM) was used to estimate the number of GMM components and Gaussian parameters (Bishop, 2006). The VBGMM puts priors on the GMM components and on the GMM parameters, and it tries to find the maximum a posteriori (MAP) estimate. The first step was repeated 2000 times. The model with the highest log likelihood was chosen. Models where a hidden state had a component weight below 0.1 were rejected. In a second step, the transition and prior probabilities of the hidden states are estimated using the forward-backward algorithm (Bishop, 2006).

Clustering HMMs First, we developed HMMs for every participant to model their individual eye-movement strategy during face recognition. In a second step, we investigated if there exist different groups of eye movement patterns among participants. To cluster HMMs, we used the variational hierarchical EM algorithm (VHEM) for HMMs (Coviello, Chan, & Lanckriet, 2012, 2014). The VHEM clusters



Priors

HS 1	HS 2	HS 3	HS 4
0.35	0.52	0.03	0.1

Transition Probabilites

	To HS 1	To HS 2	To HS 3	To HS 4
From HS 1	0.34	0.39	0.24	0.04
From HS 2	0.47	0.08	0.15	0.3
From HS 3	0.87	0	0.13	0
From HS 4	0.9	0	0	0.1

Priors

HS 1	HS 2	HS 3	HS 4
0.43	0.18	0.03	0.35

Transition Probabilites

	To HS 1	To HS 2	To HS 3	To HS 4
From HS 1	0.15	0	0.2	0.65
From HS 2	0.34	0.22	0	0.44
From HS 3	0	0	0.22	0.79
From HS 4	0.22	0.26	0.17	0.35

Figure 4: The Figure shows the fixation heat maps and overlaid HMMs of the first subject. Moreover, prior values and transition probabilities are shown. On the left, an HMM was developed only using fixation locations as input. On the right, we developed an HMM with input features consisting of fixation locations and full image information (30 coefficients). The first HMM has hidden states that closely correspond to the heat map. In contrast, the second HMM has hidden states that center more on facial features.

HMMs into groups and presents common ROIs and transitions for each group. Moreover, heat maps were developed using the fixations of all subjects clustered together. We applied the VHEM algorithm to HMMs with partial image information.

Results

HMM with full image representation

Regions of Interest Figure 4 shows a comparison of the different HMMs that were obtained using the fixation-only input features and the input features consisting of fixation locations and image information for the first subject. The HMM which was developed only using fixation sequences has hidden states which closely correspond to the heat map. Hidden state 2 covers a great part of the central face and therefore makes it difficult to understand where exactly the fixation was present. The hidden states of the HMM with the newly proposed input features correspond to facial features rather than to the heat map. The first hidden state is clearly located on the nose and the second hidden states is on the cheek next to the nose. Moreover, the third hidden state covers the right eye and the fourth hidden state is on the mouth. The HMM with the newly proposed input features seems to better capture the regions of interest on the face stimuli of the first subject.

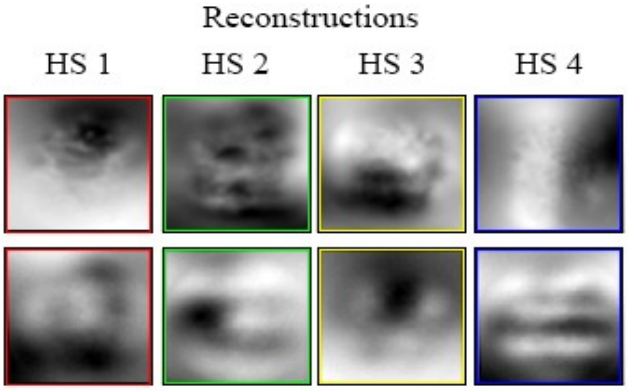


Figure 5: Images show the reconstructions of the image patches per hidden state. The top panel shows the reconstructions of the fixation-only HMM. The bottom panel shows the reconstructions of the image patches of the newly proposed HMM.

Reconstructions of image information Figure 5 shows the reconstructions of the image information for each hidden state of the HMM with the newly proposed input features (bottom panel). It is possible to identify important face regions. The first reconstruction shows the center of the nose and the second reconstruction shows the right side of the

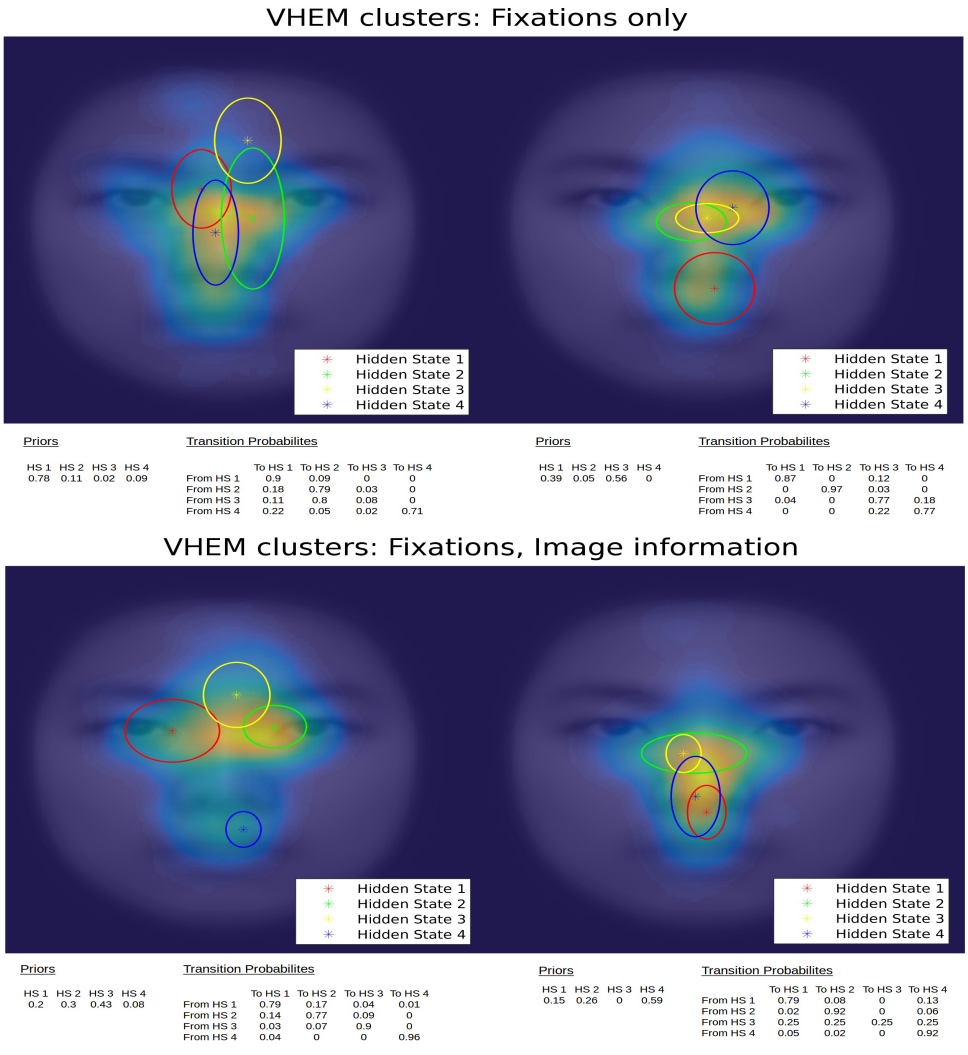


Figure 6: The Figure shows the VHEM clusters for HMMs which were developed based on fixation locations (top panel) and HMMs developed using fixation locations and image information (bottom panel). Heat maps summarize fixation locations of all participants who are part of a given cluster. VHEM clusters and the difference in heat maps of the latter are much more distinct. Prior and transition probabilities are rounded to two decimal places.

nose and parts of the right cheek. Reconstructions from the third and fourth hidden state show the right eye and the mouth respectively. The newly proposed model allows reconstructing the perceived visual stimuli at fixation, although reconstructions are not always very clear. We compared the reconstructions from the newly proposed HMM with reconstructions by the fixation-only HMM. In the top panel, averaged image patches from the hidden states of the fixation-only HMM are shown. They look more washed out and noisy than the reconstructions from the new model.

HMM with partial image representation

Group Differences In Figure 6, results of the HMM clustering are illustrated. After having developed HMMs for every participant, the VHEM algorithm was used to cluster the HMMs into groups to investigate if different fixation strategy groups exist. The top panel shows clustering results of the HMMs which were solely developed using fixation locations. Two different clusters exist. The HMM which is representative of the first cluster (top panel, left HMM) has hidden states which cover a large part of the central, upper area of the face. The hidden states of the second cluster (top panel, right HMM) are located more in the central area of the face.

The VHEM clusters obtained from the HMMs with partial image information are much more distinct. The first group (bottom panel, left) has three hidden states centered on the eye area and one hidden state on the mouth area. The HMM representative of the other VHEM cluster has all its hidden states on the nose area.

Visually, there exists a clearer difference in heat maps for VHEM clusters of HMMs with partial image information (bottom panel). The heat map of the first cluster indicates that many fixations fell on the eye area. The heat map of the second clusters shows that most fixations were on the nose area. The difference in heat maps of the VHEM clusters of HMMs without image information is less pronounced (top panel). In short, it is clear the clustering of HMMs with partial image information resulted in much more distinct clusters than the clustering of HMMs without image information

Discussion

In the present study, a new HMM approach is introduced. We showed that using image information in addition to fixation locations as input features has several benefits. The newly developed HMMs have better ROIs which are based on fixations that are similar in fixation location and image information. Moreover, the newly developed HMMs with full image representations allow to reconstruct the image information that was associated with each ROI. Lastly, clustering the newly developed HMMs resulted in very distinct groups confirming the findings by Chuk and colleagues (2014). The newly introduced HMM approach can be used for different cognitive tasks to investigate spatial and temporal characteristics of eye-movement strategies.

Acknowledgments

We are grateful to the Research Grant Council of Hong Kong (project 17402814 to J. H. Hsiao and CityU 110513 and G-CityU109/14 to A. B. Chan). We thank Tina Liu for collecting the data for the study and we thank T. Chuk for helpful discussion.

References

Bishop, C. M. (Ed.). (2006). Pattern recognition and machine. New York: Springer.

Caldara, R., & Miellet, S. (2011). imap: A novel method for statistical fixation mapping of eye movement data. Behavior Research Methods, 43, 864 - 878.

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden markov models. Journal of Vision, 14, 1 - 14.

Coviello, E., Chan, A. B., & Lanckriet, G. R. G. (2014). Clustering hidden markov models with variational hem. Journal of Machine Learning Research, 15, 697 - 747.

Coviello, E., Lanckriet, G. R., & Chan, A. B. (2012). The variational hierarchical em algorithm for clustering hidden markov models. In P. Barlett (Ed.), Advances in neural information processing systems. New York: Curran Associates, Inc.

Geisler, W. S., & Perry, J. S. (1998). A real-time foveated multi-resolution system for low-bandwidth video communication. Human Vision and Electronic Imaging, 3299, 294 - 305.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. Memory and Cognition, 33, 98 - 106.

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. Psychological Science, 19, 998 - 1006.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S.. (1997). Fixation patterns made during brief examination of two dimensional images. Perception, 26, 1059 - 1072.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vision Research, 11, 929 – 942.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9, 97 - 113

Simola, J., Salojarvi, J., & Kojo, I. (2008). Using hidden markov model to uncover processing states from eye movements in information search tasks. Cognitive Systems Research, 9, 237 - 251.

Walker-Smith, G. J., & Findlay, A. G. G. J. M. (1977). Eye movement strategies involved in face perception. Percep- tion, 6, 313 326.

Wedel, M., Pieters, R., & Liechty, J. (2003). Evidence for covert attention switching from eye-movements. reply to commentaries on liechty et al., 2003. Psychometrika, http://doi.org/10.1007/BF02295611.

Yarbus, A. L. (Ed.). (1965). Eye movements and vision, translated from Russian by basil haigh. New York: Plenum Press.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. (1985). Matching familiar and unfamiliar faces on internal and external features. Perception, 14, 737 - 746