Journal of Eye Movement Research
12(2):4

# Effects of Individuality, Education, and Image on Visual Attention: Analyzing Eye-tracking Data using Machine Learning

Sangwon Lee
Yonsei University
Seoul, South Korea

Yongha Hwang
University of Michigan
Ann Arbor, United States

Yan Jin
Qingdao University of Technology
Qingdao, China

Sihyeong Ahn
Yonsei University
Seoul, South Korea

Jaewan Park
Yonsei University
Seoul, South Korea

Machine learning, particularly classification algorithms, constructs mathematical models from labeled data that can predict labels for new data. Using its capability to identify distinguishing patterns among multi-dimensional data, we investigated the impact of three factors on the observation of architectural scenes: individuality, education, and image stimuli. An analysis of the eye-tracking data revealed that (1) a velocity histogram was unique to individuals, (2) students of architecture and other disciplines could be distinguished via endogenous parameters, but (3) they were more distinct in terms of seeking structural versus symbolic elements. Because of the reverse nature of the classification algorithms that automatically learn from data, we could identify relevant parameters and distinguishing eye-tracking patterns that have not been reported in previous studies.

Keywords: Eye tracking, visual attention, individual differences, art perception, architectural design, machine learning, classification, region of interest

## Introduction

In the design research community, the benefit of using eye-tracking is distinct: it provides quantifiable information about a viewer's visual attention in a non-intrusive manner. Its significance is based on the fact that the visual appearance of a product plays a critical role in consumer response (Crilly, Moultrie, & Clarkson, 2004) and on the hypothesis that eye-tracking data instantly externalize what people think (Just & Carpenter, 1980) or aim to accomplish (Just & Carpenter, 1976).

Eye-tracking is subject to certain restrictions though: processing eye-tracking data computationally has no universal standard (Kiefer, Giannopoulos, Raubal, & Duchowski, 2017), the link between eye-tracking measures and those of domain problems is hidden (Mayer, 2010), experimental setups may not fully represent actual practice (Venjakob & Mello-Thoms, 2016), and the quality of analysis often depends on the ability to

Received January 3, 2019; Published July 16, 2019.
Citation: Lee, S., Hwang, Y., Jin, Y., Ahn, S, & Park, J. (2019). Effects of Individuality, Education, and Image on Visual Attention: Analyzing Eye-tracking Data using Machine Learning. *Journal of Eye Movement Research*, *12*(2):4.
Digital Object Identifier: 10.16910/jemr.12.2.4
ISSN: 1995-8692

build customized software (Oakes, 2012). However, eye-trackers have been shown to provide objective measures that can be associated with high-level design problems, such as usability (Nielson & Pernice, 2010), training effects (Nodine, Locher, & Krupinski, 1993; Park, DeLong, & Woods, 2012), preference (Reid, MacDonald, & Du, 2012), and cultural influences (Dong & Lee, 2008).

Eye-tracking data have been linked to design problems through layers of mediating parameters. Human eye movement consists primarily of two phenomena: 1) fixation—a relatively stationary period of eye movement and 2) saccades—rapid movements between fixations. The low-level parameters, e.g., fixation duration, fixated positions, and saccade amplitude, have been combined as quantitative indicators of a specified design problem. For example, changes in fixated positions were related to the specified task (Yarbus, 1967), and the distribution of gaze durations (cumulative fixation duration within a cluster) has been found to quantify the difference between individuals with and without artistic training (Nodine et al., 1993). The mean fixation duration and saccadic amplitude have been established to encode individualities (Castelhano & Henderson, 2008), and the shapes of reading patterns have demonstrated the effects of cultural background (Dong & Lee, 2008).

The limitation of this practice is that the determination of the criterion parameters for a specified problem is not always straightforward or successful. That is, the selected eye-tracking parameters were often not effective indicators of the target effect. The consequence is the lack of agreement on the mapping between parameters and design problems (Ehmke & Wilson, 2007), and studies whose findings partially support or lie outside the scope of the initial goal (Weber, Choi, & Stark, 2002; Koivunen, Kukkonen, Lahtinen, Rantala, & Sharmin, 2004; Kukkonen, 2005; Reid et al., 2012; Lee, Cinn, Yan, & Jung, 2015). One major cause of such phenomenon is the difficulty of dealing with multi-dimensional data; it is beyond human intuition to compare multiple high-dimensional data simultaneously. For example, we can visually inspect the scanpaths of two images, but comparing multiple scanpaths in two groups is significantly more challenging (Lorigo et al., 2008). We can compare mean fixation durations at once, but quantifying the differences of fixations over a period of time allows for multiple parameterizations, which makes it more complicated to gain insights and test hypotheses quickly. A more profound task is to relate the parameters embedded in high-dimensional eye-tracking data to higher-level domain problems. The total number of combinations of parameters grows exponentially with the dimension of the parameter, and finding the relevant eye-tracking parameters through iterative testing appears to be prohibitively inefficient. It requires a more effective method to measure the relative impact of eye-tracking parameters and detect hidden patterns.

According to Arthur Samuel, machine learning is a field of study that gives computers the ability to learn without being explicitly programmed (Simon, 2013). By virtue of its capability to identify trends and make predictions from multi-dimensional data, machine learning has significant potential for detecting new patterns and verifying existing propositions in eye-tracking studies. Greene, Liu, & Wolfe (2012) and Borji & Itti (2014) applied a classification algorithm to eye-tracking data labeled with task information and investigated the statistical foundation of the observation that the given task affects eye-tracking patterns (Yarbus, 1967). The key advantage of machine learning lies in the order of the process; it first learns from data and then identifies the parameters relevant to the classification, rather than first predicting the potential parameters and then verifying their impact. From this perspective, a classification algorithm is a reverse approach that can identify the relevant parameters more effectively than a forward-based one where the discovery of links between eye-tracking parameters and the target effect depends heavily on the initial choice of candidate parameters (Borji & Itti, 2014).

Motivated by the opportunities that machine learning offers, our study intends to evaluate the impact of three factors associated with viewing architectural scenes: individuality, education, and stimuli. Among the factors exogenous and endogenous to the participating individual, we designed the experiment such that eye-tracking data constituted the combined effect of natural tendency, architectural training, and image content. First, the presence of eye-tracking parameters unique to an individual has been studied extensively (Andrews & Coppola, 1999; Castelhano & Henderson, 2008; Boot, Becic, & Kramer, 2009; Mehoudar, Arizpe, Baker, & Yovel, 2014; Greene et al., 2012; Lee et al., 2015). We explored new eye-tracking parameters that are likely to identify an individual from a larger dataset. Furthermore, the art and design community has been paying significant attention to distinguishing between "trained" and "untrained" eyes (Nodine et al., 1993; Weber et al., 2002; Kukkonen, 2005; Vogt & Magnussen, 2007; Park et al., 2012; Lee et al., 2015). According to the notion that evaluative discrepancy in architecture is particularly expensive (Fawcett, Ellingham, & Platt, 2008), we aimed to identify, quantify, and visualize patterns that distinguish between majors and non-majors of architecture-related disciplines. Finally, it has been reported that the presence of image content indicative of the specified task alters what people attend to (Yarbus, 1967; Castelhano, Mac, & Henderson, 2009; Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010; Greene et al., 2012; Borji & Itti, 2014). One of our primary focuses was the impact of image stimuli in relation to individuality or educational background. When classifications of an individual or major/non-major across all image stimuli failed to predict the identity, we compared the classification accuracy of each image and looked for the key image features that distinguished an individual or educational background.

# Background

In art and design research, eye-tracking data have been used as a quantifiable measure, an objective indicator, and scientific evidence of various aesthetic rules and design heuristics. In art research, one primary question was the manner in which trained artists behave differently from novices. According to Berlyne's (1971) notion of diverse vs. specific exploration, Nodine et al. (1993) assumed that artists shift from specific to diverse exploration when the symmetry of aesthetic composition breaks. The hypothesis was verified by artists' dispersed, shorter gaze durations at asymmetric compositions to which non-artists were less sensitive. In subsequent research, Vogt & Magnussen (2007) found that artists pay more attention to structural aspects than to individual elements. Miall & Tchalenko (2001) focused on the actual process of painting by combining an eye-tracker with a hand-tracker and identified three distinct patterns: initial prolonged attention to the model, rapid alternation of attention between the model and the canvas for sketching, and practice strokes on the canvas. They proposed fixation stability, fixation duration, and targeting efficiency as parameters for defining the artist's eye skills and eye–hand coordination.

In design discipline, the eye-tracking research diversified by its sub-disciplines. In the product design domain, researchers have explored the use of the eye-tracker as a tool for understanding user preference within the entire product development cycle. Using the effectiveness of the eye-tracker for measuring user attention (Hammer & Lengyel, 1991), Kukkonen (2005) explored the connection between the attended area and product preference. Reid et al. (2012) used eye-tracking data to corroborate survey information that investigated the influence of product representation on user selection, and Köhler, Falk, & Schmitt (2015) proposed that eye-tracking aids the extraction of the visual impression and the emotional evaluation as part of the Kansei engineering process. In the visual communication domain, numerous studies have addressed the usability of 2D graphical user interfaces. Nielson & Pernice (2010) used a large set of eye-tracking data to produce design guidelines for webpage design, and Dong & Lee (2008) externalized how cultural background affects webpage reading behavior using eye-tracking scanpath maps. Prats, Garner, Jowers, McKay, & Pedreira (2010) demonstrated that eye-tracking parameters can indicate the moment when shape interpretation occurs, and Ehmke & Wilson (2007) listed the eye-tracking parameters relevant to various web usability problems. Two eye-tracking studies in the architecture domain have investigated the role of architectural elements and the impact of architectural training on viewing architectural scenes (Weber et al., 2002; Lee et al., 2015). A study in the fashion design domain revealed how designers and non-designers view differently in the context of participatory design (Park et al., 2012).

Occasionally, design research that used eye-tracking data exhibited variance in the level of success, i.e., differences in the number of objectively verified hypotheses and proposed hypotheses. The characteristics of these studies are the absence of analysis on the proposed questions, lack of quantitative reasoning, and high rate of unexpected findings. For example, Koivunen et al. (2004) initially intended to reveal the influence of design education and rendering style; however, they observed behaviors during the first impression and different fixation durations per task. Kukkonen (2005) measured gaze data, preference scores, the most favored product, and individual evaluations and concluded that there is negligible correlation among them. Reid et al. (2012) identified that a long fixation duration can indicate either of high and low preferences, but did not provide statistical evidence or an in-depth analysis. Lee et al. (2015) identified potential eye-tracking parameters for differentiating individuals that were not part of their original research questions. Occasionally, an unexpected factor, e.g., image size (Kukkonen, 2005) and presentation order (Reid et al., 2012), was the source of the failure, but a more fundamental cause appears to be the inability to predict the affecting parameters. Whereas high-dimensional eye-tracking data enable a large set of parameter combinations, their connection to high-level design issues is not revealed until we test them.

Recently, two papers have reported controversial opinions on the observation of Yarbus (1967) that the specified task affects the eye-tracking pattern. Greene et al. (2012) displayed 64 images to 16 participants with four tasks but the correct prediction rate was only marginally higher than random chance (27.1%, 95% CI = 24–31%, chance = 25%). Using the same data, Borji & Itti (2014) disputed the conclusion with a significantly higher prediction rate (34.12%). The element that differentiates their methods from previous ones was the adoption of machine learning, in particular a classification algorithm. More traditional approaches would have selected a set of indicative eye-tracking parameters and tested whether they fluctuate by a significant margin as the specified tasks differ. Rather, they constructed a prediction model using training data and analyzed its performance by comparing the predicted task with the actual task using validation data. The prediction model essentially draws boundaries between eye-tracking data with different tasks within the multi-dimensional parameter space. Its performance depends on how clearly the model can detect boundaries among training data and the extent to which the logic for dividing the training data is applicable to the validation data. The key difference between Greene et al.'s (2012) and Borji & Itti's (2014) studies was the selection of the classification method for constructing the prediction model, i.e., a linear discriminant vs. the RUSBoost classifier.

In this study, we aimed to evaluate the impact of three factors, i.e., individuality, educational background, and image stimuli, by using machine learning to explore multi-dimensional data. Regarding individuality, the question has been whether endogenous eye-tracking parameters consistent across different viewing conditions exist. The motivation was to know (1) the extent to which endogenous factors affect eye-tracking patterns, and (2) the potential connection with neural substrates, such as ADHD, dementia, memory (Castelhano & Henderson, 2008), visual search performance (Boot et al., 2009), and visual recognition strategy (Mehouda et al., 2014). Andrews & Coppola (1999) found that the mean fixation duration and saccade amplitudes formed a linear relationship in active and passive viewing tasks. Castelhano & Henderson (2008) demonstrated that these parameters are stable across differing image content, quality, and format. Greene et al. (2012) succeeded in predicting the identities of eye-tracking data using machine learning with significantly higher probability than random chance (26% vs. 6.3%). Recently, Lee et al. (2015) proposed the existence of additional patterns unique to certain individuals based on visual inspection. In this study, we searched for more fingerprinting patterns with higher predictability. Second, previous studies have found that groups of individuals with and without certain educational training differed in exploration patterns or cumulative fixation durations on the designated area of interest. The group with educational training focused more on the background or structural relationships among individual elements (Nodine et al., 1993; Weber et al., 2002; Vogt & Magnussen, 2007) and on image content that was relevant to the focus of their training (Park et al., 2012; Lee et al., 2015). Observing that only a few eye-tracking parameters have been associated with group characteristics (Nodine et al., 1993; Weber et al., 2002; Park et al., 2012; Lee et al., 2015), we explored additional parameters distinguishing between major and non-major students of architecture discipline. Finally, the impact of image stimuli varied in different decoding tasks. Although it was not sufficiently strong to affect individual decoding (Castelhano & Henderson, 2008), image content with diagnostic information relevant to the specified task was crucial in task decoding (Borji & Itti, 2014). Image contents such as symmetry (Nodine et al., 1993), background complexity (Park et al., 2012), and inclusion of architectural elements (Lee et al., 2015) were found to affect the decoding of educational background. Our focus in the case of image stimuli was to identify particular image content that attracts a particular individual or major/non-major.

## Methods

We used the data generated by Lee et al. (2015), which is publicly available (http://bit.ly/2eqb4TV), as input data. To observe the effect of architectural training on visual attention, they recorded 10-s eye-tracking patterns of 71 major/non-major participants (39 majors and 32 non-majors) on 14 images with certain architectural elements (Appendix). The data consist of eye positions sampled with a frequency of 60 Hz in normalized coordinates: a screen space with a size of 1.0 (width) by 0.74 (height). A fixation was defined as a group of sample points whose diameter does not exceed 0.02 in normalized length and 300 ms in time between the first and last one. All other events, such as glissades and smooth pursuits as well as saccades in the traditional sense were collected into single 'saccade' category in our study, following the **I**dentification by **D**ispersion **T**hreshold algorithm (Figure 1, Holmqvist & Nystrom, 2011). Hence, the definition of saccade in this paper is broader than more typical and conventional definition of saccade usually between 30-500 deg/s.

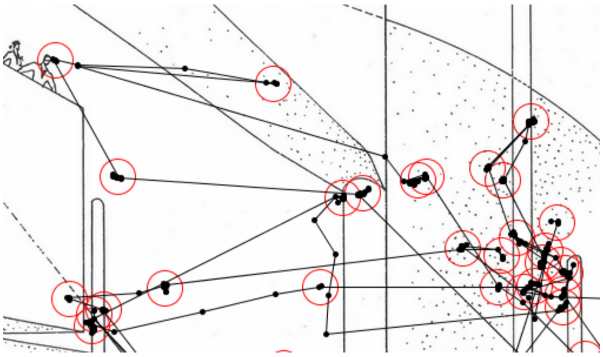


Figure 1. Fixations from eye positions using the I-DT algorithm.

Rather than establishing candidate parameters and verifying their statistical significance, we applied machine learning to identify the distinguishing parameters and their patterns that characterize individuals and majors/non-majors. Regarding the data features for decoding individuals and majors/non-majors, we considered characteristic patterns such as oscillating movements and the extent of fixation over time proposed by Lee et al. (2015), as well as well-established endogenous parameters such as total fixations, mean fixation duration, and mean saccade amplitude (Greene et al., 2012). To understand the impact of image stimuli, we also included a fixation map representing the cumulative fixation durations on each of the 10 × 10 cell grids within the image area. The complete list of features is as follows:

(1) Fixation data comprising average fixation duration, total fixation count, average saccade length, and total saccade count

(2) Fixation histogram data whose bins represent different ranges of fixation duration

(3) Saccade histogram data whose bins indicate ranges of saccade lengths

(4) Velocity histogram data whose bins are ranges of normalized lengths between adjacent points sampled at 60 Hz

(5) Fixation map data representing cumulative fixation durations on a 10 × 10 grid

Velocity histograms store normalized lengths traveled in 1/60 of a second. Because a 24.98 cm × 18.61 cm screen was placed 50 cm from the participant, velocity $v$ ($s^{-1}$) can be converted to a degree in the visual angle by using $2 \tan^{-1} (v/2 \times 24.98/50)$ (deg/s). The histogram consisted of 14 bins, 1 special bin reserved for zero velocity, and 13 bins for the rest. The ranges of 13 bins were determined by first sorting the data and then dividing them into 13 groups of equal sizes. Fixation and the saccade histogram consisted of 10 bins of equal sizes with no special bin.

In our study, we performed the classification by using each individual or major/non-major labels. Because the performance of the prediction model can vary widely according to the selected classifier (Greene et al., 2012; Borji & Itti, 2014), we compared the results of three classifiers: decision tree, a support vector machine (SVM) with a linear kernel (SVM.LinearSVC), and an SVM with a radial basis function kernel (SVM.SVC) implemented with the Python machine learning package (www.scikit-learn.org). While SVM with a linear kernel was the choice of previous eye-tracking research with high-dimensional data (Greene et al., 2012), we tested a radial kernel because it may perform better for lower-dimensional data with feature selection. Decision tree was useful to compare the importance of different features. For SVM classifiers, we applied feature selection by using the extremely randomized tree (Geurts, Ernst, & Wehenkel, 2006) with [1.7, 1.8, 1.9, 2.0, 2.1] × mean as the threshold, [10, 30, 50, 70, 90] as the max depth, and [2, 5, 10, 15, 20] as the min_samples_split. For hyperparameter tuning, [0.25, 0.5, 1, 2, 4] was used as the C value. For decision tree, we used [gini, entropy] as the criterion, [10, 20, … 80] as the max_depth, [auto, sqrt] as the max_features, [1, 2, …, 10] as the min_samples_leaf, and [2,3, .., 10] as the min_samples_split.

To decode the individual identities of the eye-tracking data, we compared the correct prediction rate (accuracy) against the random chance level (1 / 71 individuals = 1.41%). Different classifiers were compared to determine the most optimum result. To split the entire data into training and validation datasets, we adopted the leave-N-out selection scheme; 71 samples from the individual-image pairs formed the validation set and the remaining (14 - 1) × 71 samples formed the training set. With each of the 14 iterations, we chose one image out of 14 images to form a validation set. This folding scheme allowed no sample in the testing set see the target image of the validation set, ensuring that the prediction of validation set is based solely on endogenous factor (individuality) by excluding the effect of exogenous factor (image). An exhaustive alternative would have been to iterate over all the $14^{71}$ training/validation set combinations. To decode majors/non-majors, we divided the eye-tracking data of all the participants (14 images × 71 participants) into 70/30, i.e., 70% for the training dataset and 30% for the validation dataset. As with individual decoding, we compared the performance of different classifiers averaged over 14 iterations. To estimate the statistical significance, we adopted one-way ANOVA using 14 data samples against the chance level (50%) in accordance with Greene et al. (2012). Please note that the histogram bin ranges were recalculated for each iteration by using the data samples in the training set only. This was done to make sure that the validation set had no effect on feature extraction (Table 1).

Finally, in order to investigate the effects of image stimuli, we classified 71 participants' data for each image into major/non-major groups and identified the image content that contributed to high correct prediction rates. We ran 71 iterations per image; in each case, one of the 71 data samples formed the validation dataset and the remaining formed the training dataset. Decoding an individual per image was not feasible because there was only one sample from each participant per image, preventing division into training/validation datasets.

# Results

Table 1. The range of each bin averaged over 14 iterations.

| **Bin** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| Fixation duration histogram | 0.050 ±<br>0 | 0.111 ±<br>2.01e-3 | 0.169 ±<br>5.34e-4 | 0.201 ±<br>6.76e-4 | 0.222 ±<br>1.45e-4 |
| | **6** | **7** | **8** | **9** | **10** |
| | 0.233 ±<br>2.59e-5 | 0.250 ±<br>2.77e-17 | 0.283 ±<br>1.62e-6 | 0.476 ±<br>±4.11e-3 | ∞ |

| **Bin** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| Saccade length histogram | 0.020 ±<br>4.81e-5 | 0.027 ±<br>8.72e-5 | 0.035 ±<br>1.37e-4 | 0.045 ±<br>2.21e-4 | 0.057 ±<br>2.81e-4 |
| | **6** | **7** | **8** | **9** | **10** |
| | 0.072 ±<br>3.97e-4 | 0.094 ±<br>4.82e-4 | 0.125 ±<br>7.6e-4 | 0.184 ±<br>±1.36e-4 | ∞ |

| **Bin** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|
| Velocity histogram | 0.0<br>±<br>0.0 | 7.79e-4<br>±<br>4.37e-6 | 1.27e-3<br>±<br>5.65e-6 | 1.79e-3<br>±<br>7.86e-6 | 2.38e-3<br>±<br>1.36e-5 | 3.08e-3 ±<br>1.63e-5 | 4.03e-3<br>±<br>2.22e-5 |
| | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
| | 5.39e-3<br>±<br>3.46e-5 | 7.62e-3<br>±<br>5.64e-5 | 0.012<br>±<br>9.67e-5 | 0.0224<br>±<br>2.46e-4 | 0.0456<br>±<br>4.68e-4 | 0.0947<br>±<br>7.49e-4 | ∞ |

## Decoding Individuals

Before matching various classifiers with different data features, we applied a decision tree classifier to obtain the relative importance of data features. The overall results indicated that the fixation data and velocity histogram were more effective for individual decoding than the others (Figure 2), with a higher correction prediction rate (967/7100 = 13.62%) than the chance level (1/71 = 1.41%). Figure 2 visualizes that the velocity histogram data (yellow), particularly the earlier bins, exerted the highest importance, followed by the fixation data (light blue). The importance of the fixation histogram (orange) and saccade histogram (gray) were lower than these two features. The fixation map data (dark blue) displays 10 peaks, whose maxima are higher near the middle (third to seventh peaks) than both ends. Each peak represents the longest fixation duration of each row in a 10 × 10 cell grid. The higher peaks near the middle indicate that cells around the center of an image were a better indicator of the identity of the individual.

Table 2. Correct prediction rate (accuracy) of different classifiers averaged over 14 iterations.

| Classifiers | Accracy |
|---|---|
| Decision Tree | 12.47 ± 4.19 |
| SVM.LinearSVC | 25.25 ± 3.82 |
| SVM.SVC | 31.09 ± 8.22 |
| | 65.29 ± 7.21 (top 5) |

Table 2 presents the results of different classifiers obtained by using feature selection and hyper-parameter tuning on all data features. Whereas the decision tree exhibited an average correctness of approximately 14%, the LinearSVC classifier performed better, and the SVC classifier generated the highest correct prediction rate. The features chosen by feature selection over 14 iterations matched well with those having high importance in the preliminary testing run using the histogram ranges averaged over all folds (Figure 2, red dotted line). The top-five correct prediction rate with the SVC classifier was 65.29%, implying that the prediction rate doubles if we permit up to five guesses per individual.

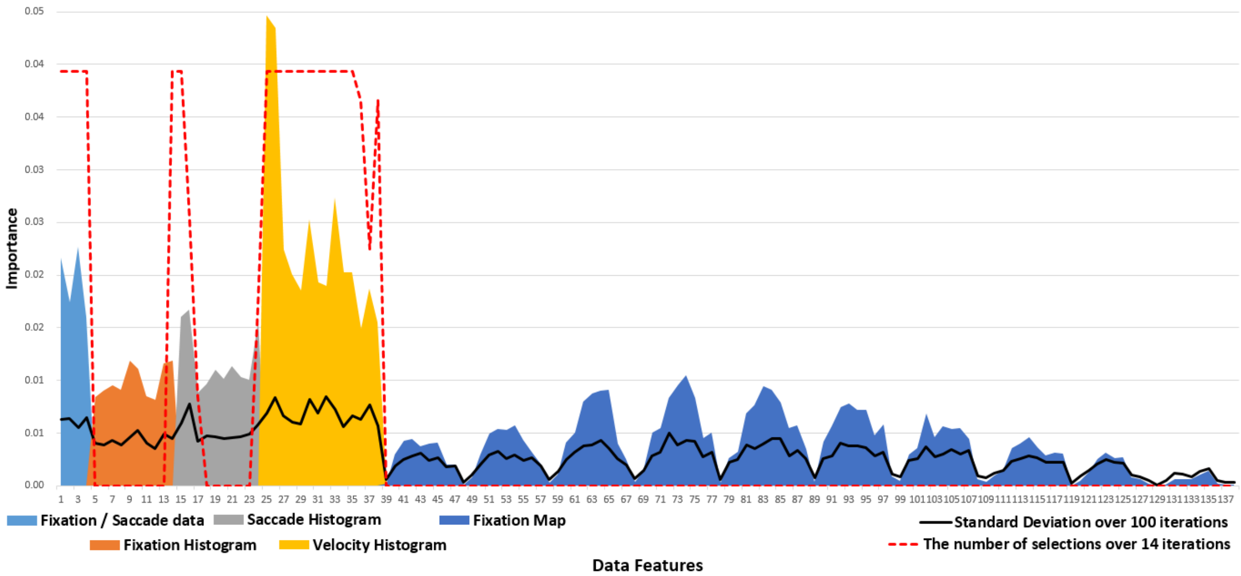


Figure 2. Importance of data features during individual classification. The indices on the x-axis correspond to the fixation data (1-4), fixation histogram (5-14), saccade histogram (15-24), velocity histogram (25-38), and fixation map data (38-138). Red dots indicate the number of selections over 14 iterations by feature selection.

## Difference between Individuals

Using the classification results with the best options (SVM.SVC classifier), we compared the correct prediction rate of each individual. Figure 3 (left) shows that the majority of the participants had a correct prediction rate higher than random chance (1.41%). The average correct prediction rate of the non-majors (33.3%) was higher than that of the majors (29.3%). Figure 3 (right) shows a confusion matrix that visualizes the correct and incorrect predictions. The x- and y-coordinates or each dot represent true and predicted individual for a specified data, and the bright diagonal line indicates the overall success of individual decoding. A few dots off the diagonal line are incorrect predictions, and the higher density of warm dots in the lower right quadrant is indicative of the high prediction rates of non-majors than majors.

## Consistency across Images

Figure 4 is a graph showing the velocity histogram per individual for all the images. It represents eight participants' data with the highest and lowest correct prediction rates. The highest rates were 92.86% (Figure 4(a,b)), 85.71% (Figure 4(c)), and 71.43% (Figure 4(d)); the lowest were 7.14% (Figure 4(f)) and 0% (Figure 4(e,g,h)). Each Figure contains 14 colored lines, representing data from all the 14 images. The most prominent difference between the two groups was the consistency across images; the first four participants, particularly Figure 4(b), tend to have more narrowly clustered lines than the others. Moreover, as the early bins exert larger impact for individual decoding (Figure 2), the level of convergence at bin 1 appears to have contributed to the higher prediction rate of Figure 4(d) than that of Figure 4(f,h) notwithstanding their overall similarity. The highest performance in Figure 4(a) seems to be explained by its uniqueness of pattern among the participants; this is further supported by the fixation data (Figure 5(d)).

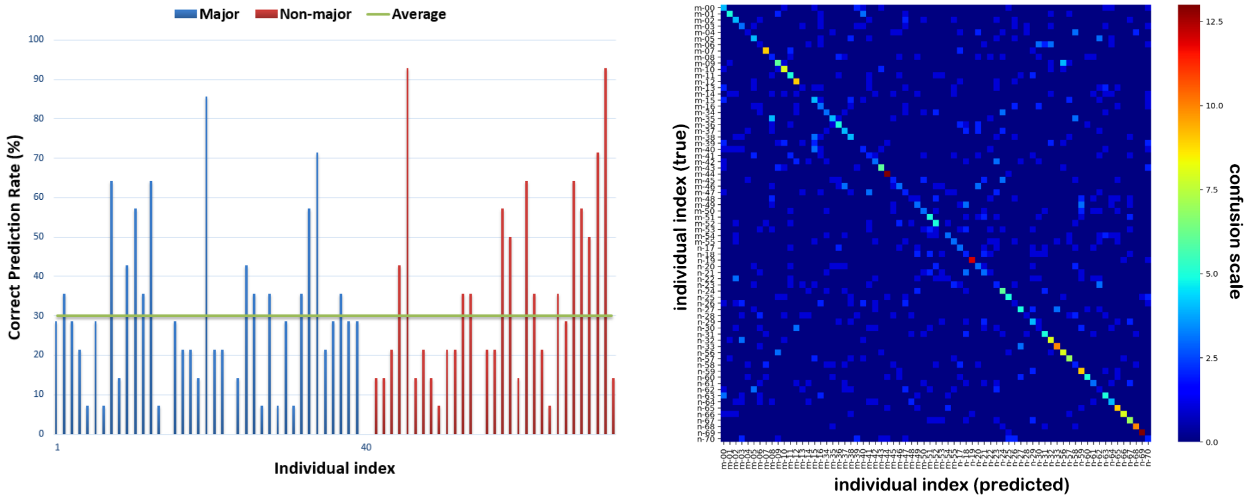
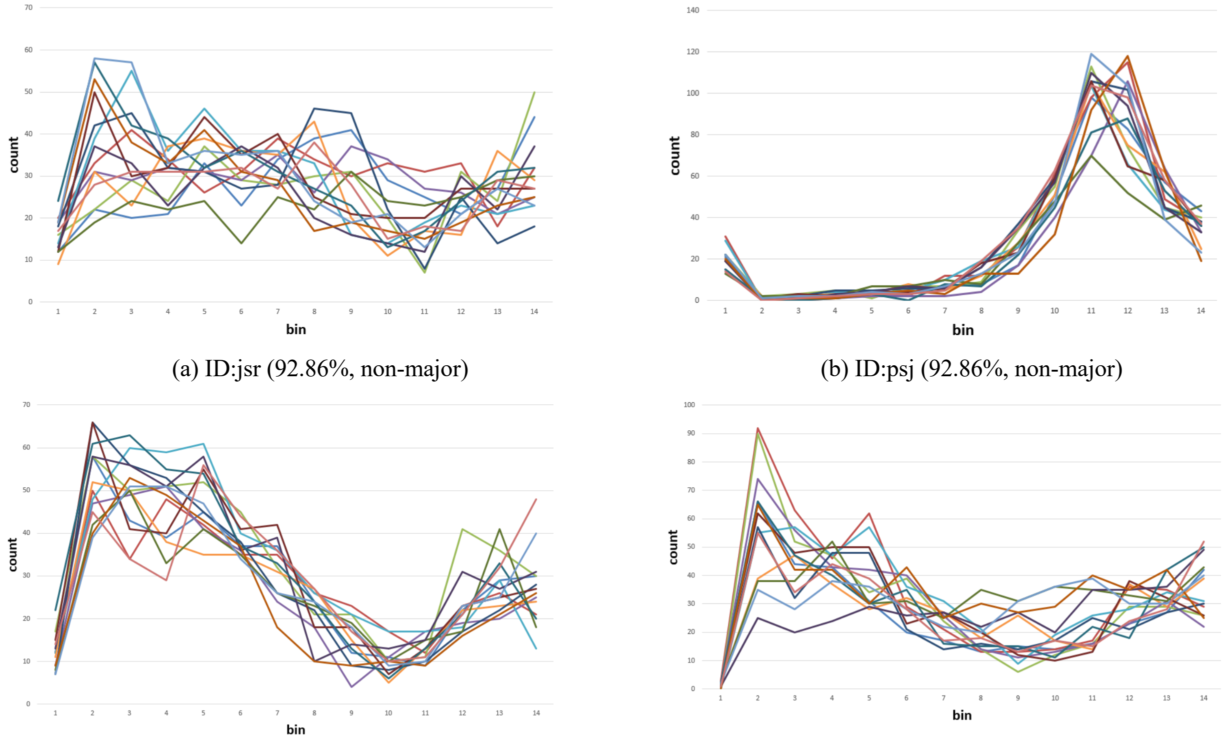


Figure 3. Correct prediction rate per individual (left) and corresponding confusion matrix (right).

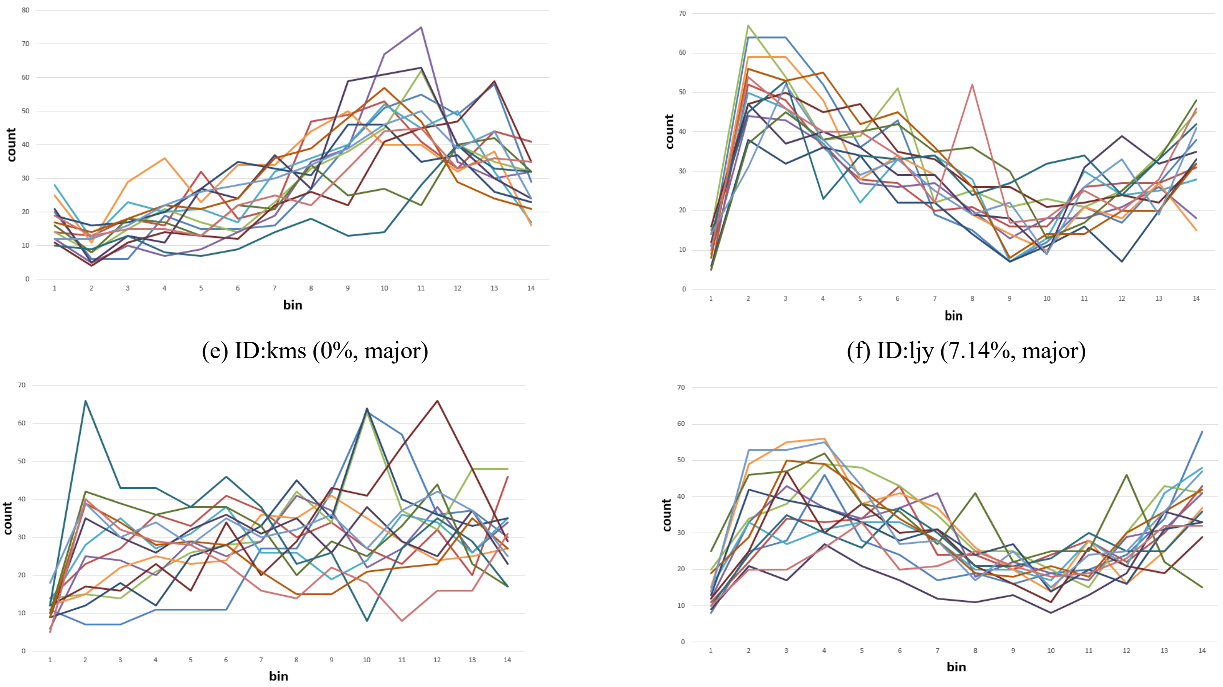


(a) ID:jsr (92.86%, non-major)

(b) ID:psj (92.86%, non-major)

(c) ID:kha (85.71%, major)

(d) ID:pej (71.43%, non-major)



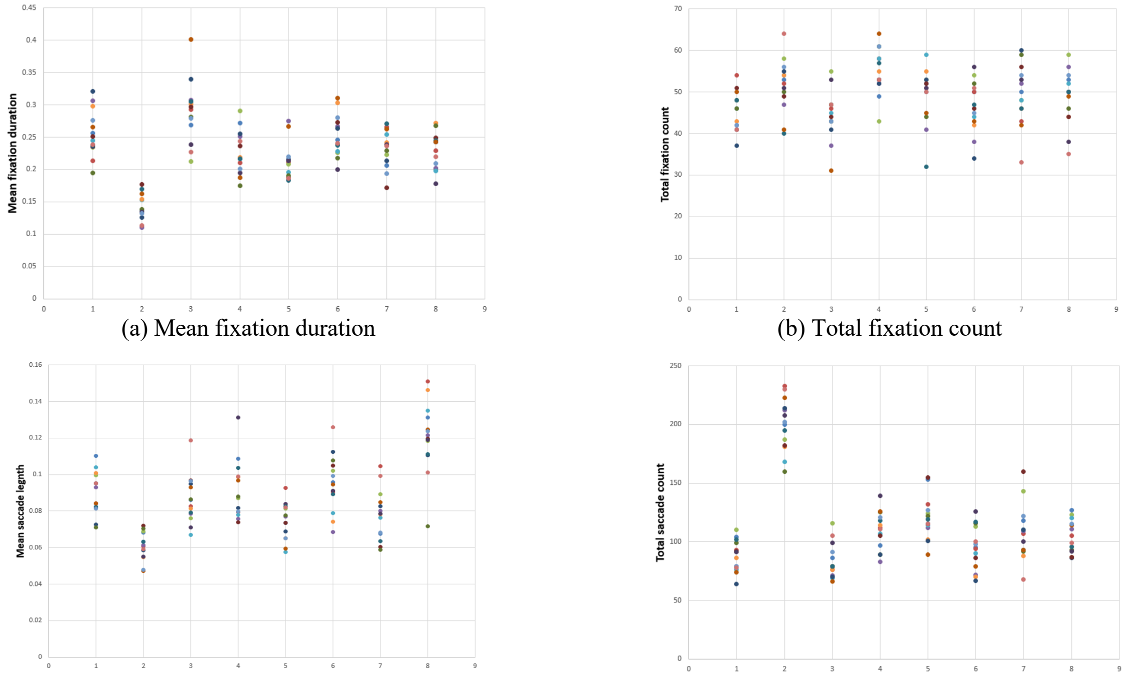
(e) ID:kms (0%, major)

(f) ID:ljy (7.14%, major)

(g) ID:ndk (0%, major)

(h) ID:sny (0%, non-major)

Figure 4. Velocity histogram of the individuals with top four prediction rate ((a)–(d)) and bottom four prediction rate ((e)–(h)), using the average bin ranges shown in Table 1.



(a) Mean fixation duration

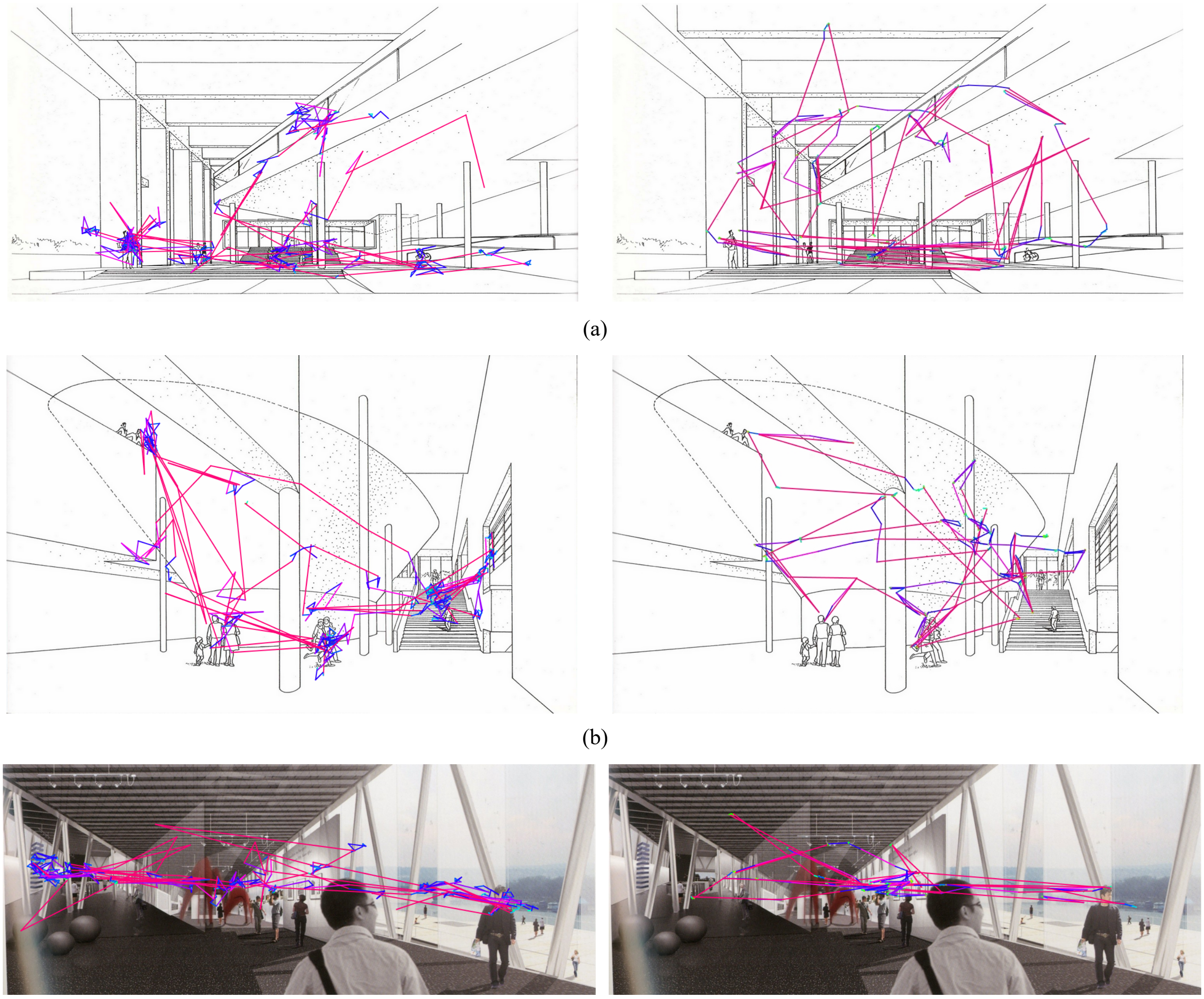
(b) Total fixation count

(c) Mean saccade length

(d) Total saccade count

Figure 5. Fixation and saccade data features of the eight participants in Figure 4. 1–8 on the x-axis corresponds to (a)–(h) in Figure 4. The dots of the different colors represent the samples from the 14 images.

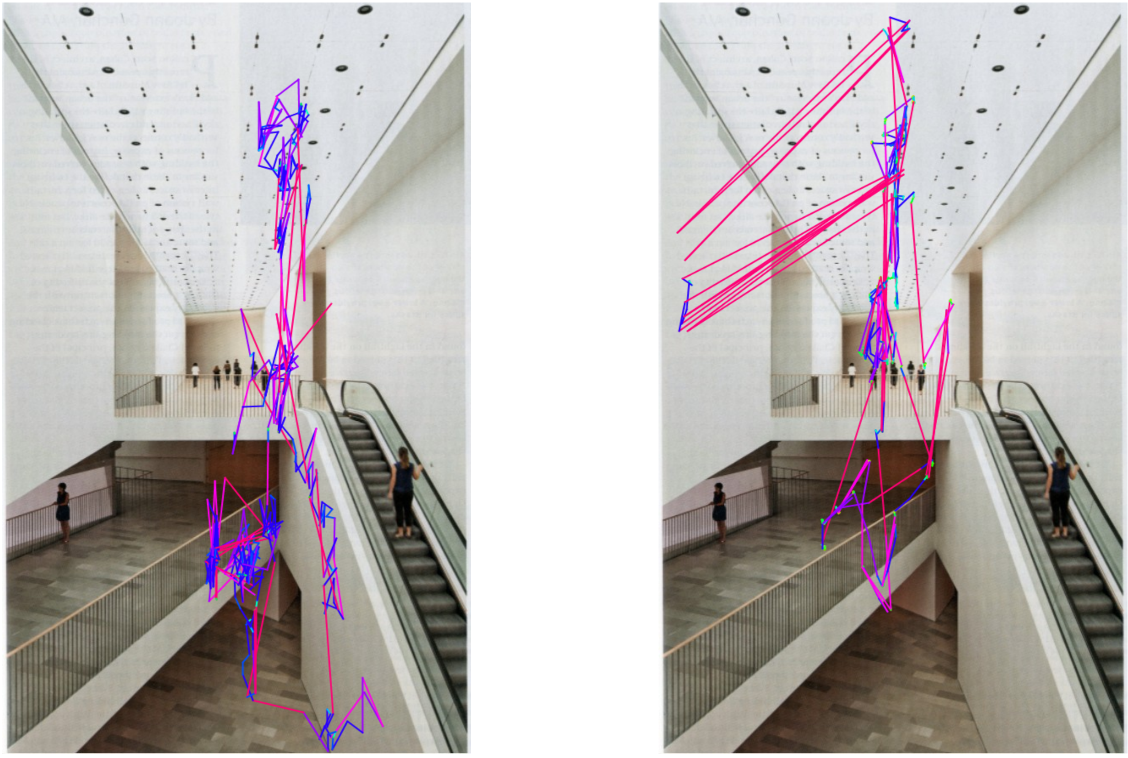
We could observe such consistency from the scanpath visualization as well. Figure 6 shows the actual eye-tracking patterns of the two participants (Figure 5(b,d)), whose lines with identical colors represent the same bin in the velocity histogram. We can infer from this that for each individual, the ratio between the numbers of the lines with the same color is more or less stable across the images. Moreover, as a unique sequence of colored lines repeats, it seems that it is not just the distribution of the lengths of these lines but also the order of their occurrence of them that carries individual character. For example, whereas a long red line and a set of shorter blue lines alternate in Figure 6 (left), there are numerous green dots in conjunction with the adjacent blue lines between the longer red lines in Figure 6 (right). Such an observation implies that we can have a better measure representing an individual character by incorporating both temporal (e.g., the order of lines) and spatial (e.g., orientation) properties.



(a)

(b)

(c)



(d)

Figure 6. Scanpath visualization of individuals in Figure 3(b) (left column) and Figure 3(d) (right column).

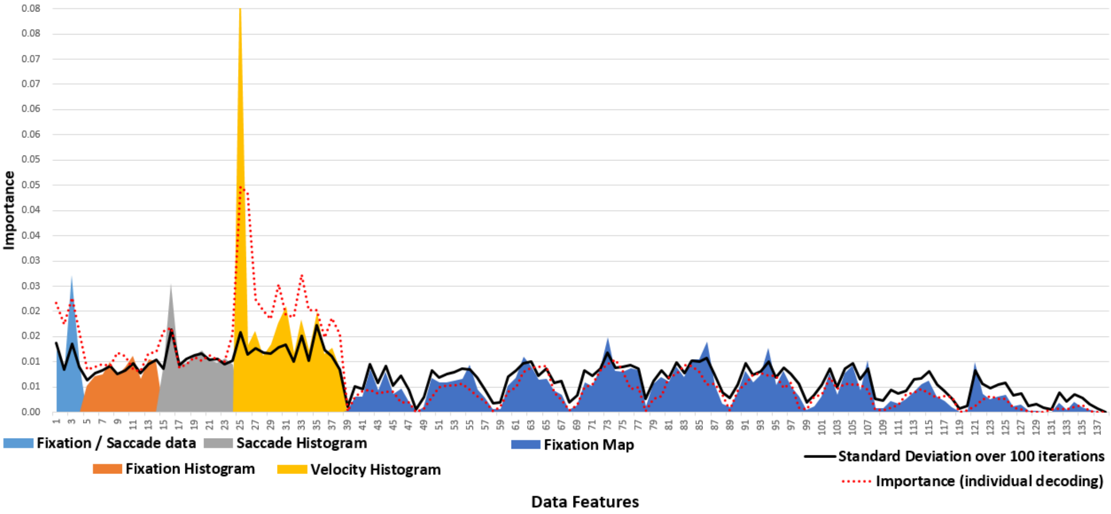


Figure 7. Importance of data features during major/non-major classification.

## Decoding Majors/Non-majors

A preliminary run of the decision tree classifier for major/non-major decoding revealed that the distribution of importance over all the data features is similar to that of individual decoding (Figure 7). However, the standard deviation over 100 iterations exceeds the average values, indicating the unavailability of clear feature(s) containing the unique properties of the majors/non-majors.

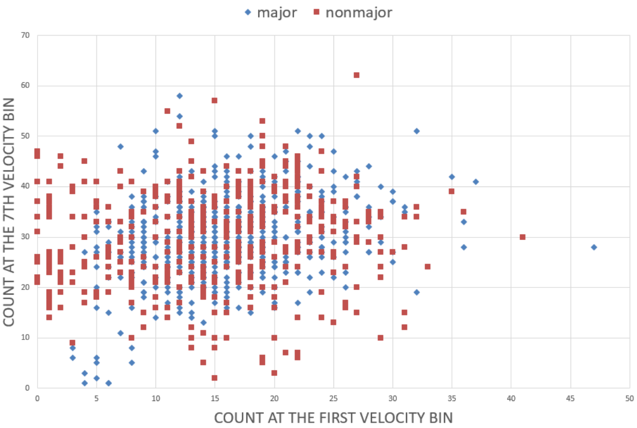


Figure 8. Distribution of the majors/non-majors by data points at the first bin (x-axis) and 7th bin (y-axis) of the velocity histogram.

Next, we matched three classifiers with different data features. The best result (64.5 ± 7.16%) was from the SVM.SVC run with feature selection and hyperparameter tuning on all data features, as in individual decoding. Whereas t-test verified a statistically meaningful difference ($p$ = 2.73e-38) against the chance level (50%), non-majors exhibited significantly lower performance (57.32%) than majors (70.39%). When we plotted a 2D graph using the two features with the highest importance (Figure 8), the majors were more narrowly clustered than the non-majors with substantial overlap. We can infer that the shared area had been labeled as majors' territory, and only the non-majors outside this territory were correctly predicted as non-majors.

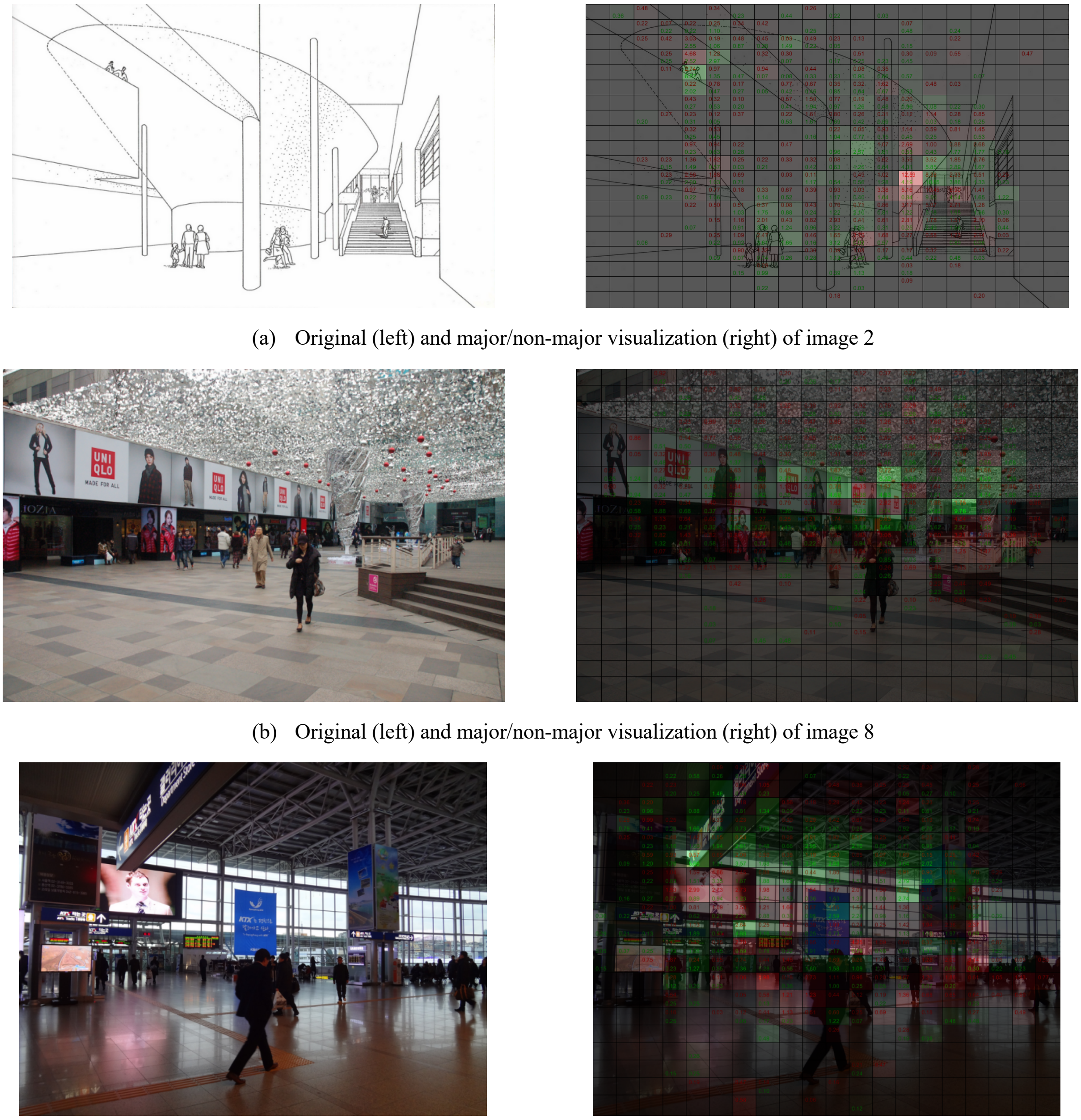
## Influence of Image Stimuli

The major/non-major classification result was statistically meaningful, but not impressive, revealing that endogenous features were not good indicators. We then considered the final factor, i.e., the impact of image stimuli, with a focus on exogenous features as recommended in previous studies (Nodine et al., 1993; Lee et al., 2015). However, when we applied a fixation map data with a 10 × 10 grid, the overall prediction rate was significantly lower owing to the insufficient resolution. We could obtain comparable results by enhancing the 10 × 10 grid to a 20 × 20 grid (Table 3), using LinearSVC as a classifier because of its suitability with high-dimension data.

Table 3 illustrates that certain images are more effective in distinguishing between majors and non-majors than others. To identify the content receiving different level of attention, we focused on those image with a performance of 70% or higher and visualized the cumulative fixation time with red (major dominant) or green (non-major dominant) with transparency (Figure 9). The larger the difference, the brighter and more transparent the color became. We also marked the fixation time on each cell to differentiate between cells with equally high and low attention.

Table 3. Per image classification result with LinearSVC classifier on 20 × 20 fixation map data.

| Img | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Avg |
|---|---|---|---|---|---|---|---|---|
| Rate | 61.8 | 70.6 | 61.8 | 52.9 | 50 | 64.7 | 58.8 | **60.7 ± 7.95 (%)** |
| Img | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
| Rate | 70.1 | 47.1 | 50 | 58.8 | 64.7 | 73.6 | 64.7 | |



(a) Original (left) and major/non-major visualization (right) of image 2

(b) Original (left) and major/non-major visualization (right) of image 8

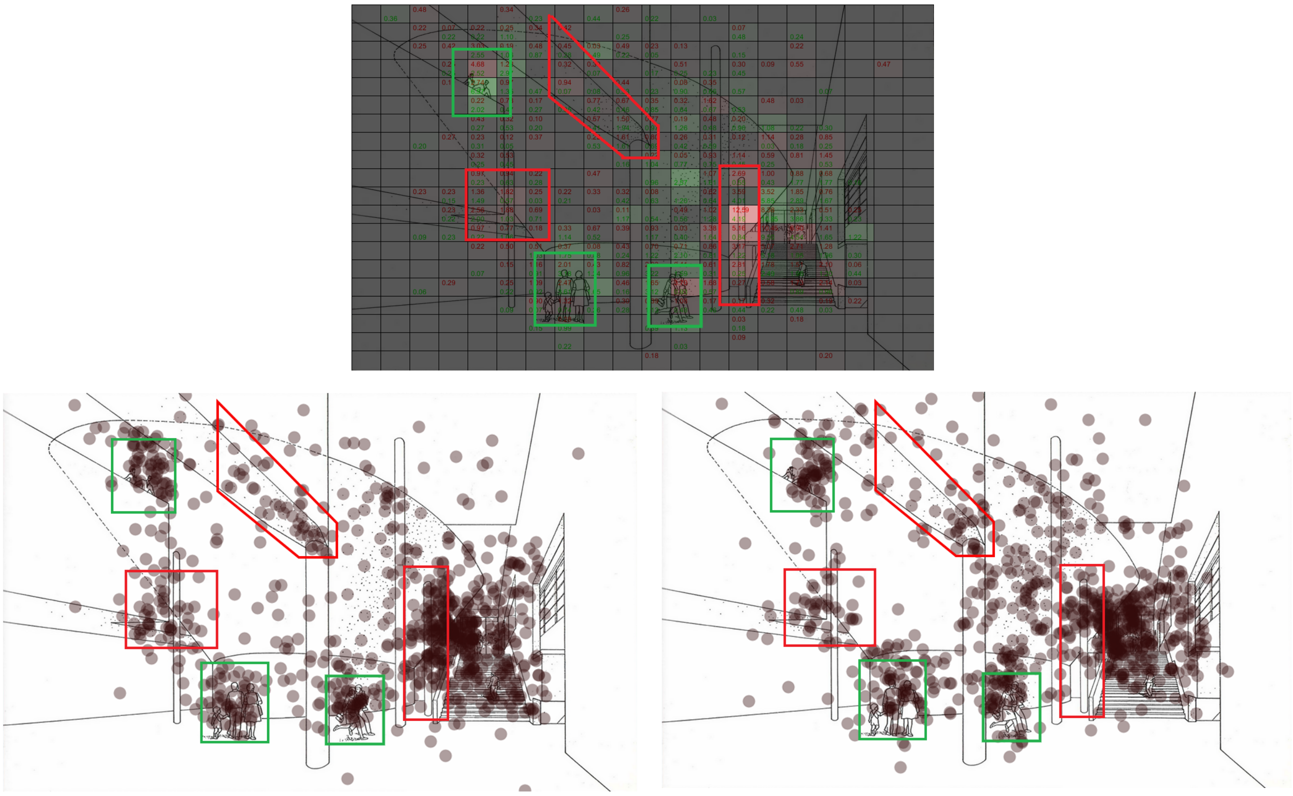
(c) Original (left) and major/non-major visualization (right) of image 13

Figure 9. Original (left column) and major/non-major visualization (right column) of images 2, 8, and 13.

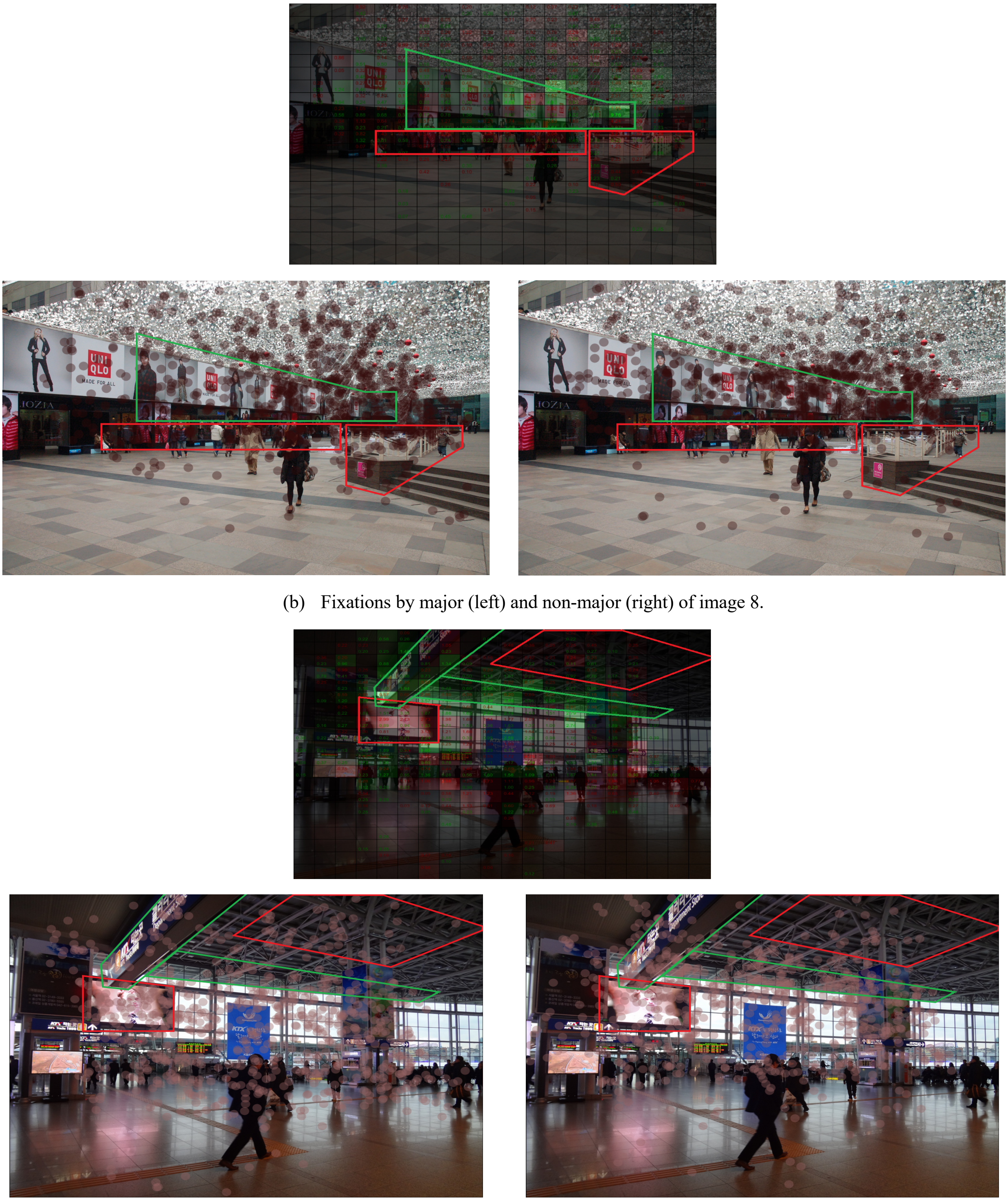
In Figure 10(a), both the majors and non-majors focused on human figures, but the non-majors exhibited marginally higher concentration (green box). We could observe larger differences at architectural elements with complex forms: conjunction between a column and a beam, the space between the stairs and a column, and the setback of the ceiling slab around a column. It appears that majors spend more time processing and interpreting structural ambiguities.

In Figure 10(b), it is noticeable that the non-majors focused more on the commercial signboards, particularly those in the brighter upper area. The strength of focus tended to get intensified toward the vanishing point. On the other hand, the majors spread more to the darker region right below the commercial signboards and paid more attention to the complex shape of the stair rails.

In Figure 10(c), both the majors and non-majors paid attention to the LED display. However, the majors stayed longer near the larger screen, whereas more non-majors tended to read the direction sign. A subtle albeit visible difference was the uneven attention toward the roof truss structure. Whereas the non-majors concentrated most on the brighter side, where the window wall and the ceiling meet, some majors focused on the ill-lighted deeper area.



(a) Fixations by major (left) and non-major (right) of image 2.



(b) Fixations by major (left) and non-major (right) of image 8.

(c) Fixations by major (left) and non-major (right) of image 13.

Figure 10. Visualization of areas receiving uneven attention by majors and non-majors.

Overall, the majors exhibited a tendency to focus more on the structural elements, whereas the non-majors were attracted to signboards or human figures. The lighting condition and complexity of the shape appeared to play a role in the division of attention in that the darker and more complex a target element was, the longer the majors stayed than non-majors. However, an exception was that the majors paid stronger attention to the large LED screen area.

## Discussion and Conclusions

To investigate the impact of different endogenous and exogenous parameters on how we view architectural scenes, we applied a classification algorithm to multi-dimensional eye-tracking data obtained from students of architecture and other disciplines. We verified the effect of three factors, namely individuality, major/non-major, and image stimuli, on visual attention. The individual identity of the eye-tracking data was encoded in the velocity histogram, representing the distribution of the speed of eye movement measured at a fixed frame rate (60 Hz). The separation between the major and non-major groups was enabled using endogenous parameters, although it could be better explained by the differing sensitivities toward structural and symbolic image features.

Regarding individual decoding, the classification was successful using velocity histogram when the impact of other factors such as fixation duration and count was not as much significant. This is inconsistent with previous findings where the fixation duration and saccade length were consistent across different images (Andrews & Coppola, 1999; Castelhano & Henderson, 2008) and the mean fixation duration, count, saccade amplitude, and coverage percent could classify individuals (Greene et al., 2012). An explanation is that our classification algorithm required more explicit distinction between individuals in a larger pool than previous forward- or reverse-based approaches (16 by Greene et al. (2012) vs. 71 in our experiment). Therefore, we recommend using the velocity histogram for better individual decoding, in addition to the mean values of fixations and saccades. The use of the eye movement distances is not completely novel in eye-tracking research; Castelhano & Henderson (2008) presented a profile of saccade distribution by length. However, their purpose was to demonstrate how natural saccade distribution could change according to the image types rather than its effectiveness in individual decoding.

The visual analysis of the velocity histogram revealed that a sequence of spatiotemporal pattern, rather than only the distribution of speed, was unique to each individual. Whereas a histogram could capture an aspect of such a pattern, it is not straightforward to determine which parameter can summarize such a feature more effectively. It appears challenging to (1) define the length of a sequence, (2) determine the tolerance of the variation, and (3) completely accommodate spatial disposition in a smaller parameter space. We consider that this venue of exploration has a potential for future research.

It is not evident why certain individuals exhibit stronger characteristics than others; in particular, the group of non-majors included a higher number of similar individuals than that of majors (Figure 3). The question is whether endogenous eye-tracking parameters are innate or acquired. Previous research has concluded that fixation and saccadic measures are natural properties determined by physical, neural, developmental, and psychological constraints (Castelhano & Henderson, 2008) and that they are consistent across substantially different image contents (Andrews & Coppola, 1999) and tasks (Boot et al., 2009) and an 18-month period (Mehoudar et al., 2014). Whereas our findings imply that a longer period of training involving visual construction affects endogenous parameters, we should also consider the likelihood that individuals with certain characteristics tend to select similar disciplines. An investigation of eye-tracking patterns over a period of educational training will help find the answer to this question.

Regarding decoding majors/non-majors, the classification was statistically significant using similar data features for individual decoding. However, the prediction rate was substantially higher for majors than for non-majors, as revealed by a 2D map whose x- and y-axes are two data features with the highest importance. The areas occupied by the two groups exhibited a large overlap, but because the majors were more narrowly clustered, the shared area had been marked as majors' territory. It resulted in the incorrect prediction of non-majors in that area as majors. The map itself is a discovery of data features characterizing the randomness of a more heterogeneous group, but its level of distinction does not appear significant.

The analysis of image stimuli revealed that the level of attention of the majors and non-majors to certain elements differed. Whereas both groups tended to fixate on visually dense areas, the majors focused more on architectural elements (stairs, columns and beams, and truss structure) and the non-majors focused more on non-structural elements (commercial boards and entourage objects). It is noteworthy that the division was prominent where a feature exhibited a complex shape or was in low lighting condition. To summarize, whereas the majors aimed to resolve structural uncertainty, the non-majors were affected more by direct symbolic cue (Nodine et al., 1993; Park et al., 2012). Its design implication is that architectural design should not only focus on organizing spaces but also consider the effect of symbols relative to visual attention. Considering that the training process is irreversible, user participation and an active use quantification method appears essential.

A limitation of the process here is that the classification performance depends highly on the resolution of the grid. In theory, a higher granularity always yields better classification results because it essentially creates more room for boundaries between different groups. Meanwhile, we also noticed a significant number of image features lying on the boundary. We recommend that a classification analysis based on boundary construction be interpreted and supplemented by other visual inspection methods. Another important point is that the reproducibility of our results will depend on the accuracy or precision of the measurement. We used raw data whose positions in normalized coordinates had a resolution in the order of 10e-4, and a timing in the order of 2.5e-6 ms, well below the frequency of the recording (60 Hz, 16.6667 ms). In our study, the first bin of the velocity histogram represents the “zero” distance between adjacent eye positions, and the existence itself along with its low to high variability across and within an individual is one proof of the soundness of the small scale data (Figure 4 and 5).

In conclusion, the application of machine learning to eye-tracking data revealed more data features unique to an individual and provided objective measures indicating the uneven attention between groups with and without educational training. Unlike previous forward-based approaches that test the effectiveness of the selected parameters, machine learning could automatically identify the distinguishing patterns from the candidate features in high dimensional spaces. However, it is also true that machine learning is not a panacea that can reveal all the hidden eye-tracking parameters. Not only did previous studies show the effectiveness of various parameterizations, but also histogram features in our study depended largely on researchers’ insights rather than blind application of machine learning. The problem proposed as future research – investigation of better methods for capturing the spatiotemporal nature and spatial distribution of eye movement – will also require trial and error of multiple hypotheses. In conclusion, the practice of forward-based searches of eye-tracking parameters will continue to exist in the near future, but the machine learning community will keep offering strong alternatives for exploring eye-tracking parameters more effectively and these alternatives would be worthwhile to consider.

### Ethics and Conflict of Interest

The author(s) declare(s) that the contents of the article are in agreement with the ethics described in http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html and that there is no conflict of interest regarding the publication of this paper.

## Acknowledgements

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014S1A5A8012349).

## References

Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research 39*, 2947-2953. https://doi.org/10.1016/s0042-6989(99)00019-x

Berlyne, D. E. (1971). *Aesthetics and Psychology* (p. 100). New York: Appleton-Century-Crofts.

Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy?: The effect of task demands and motivational factors on scanning strategy in visual search. *Journal of Vision 9*(3), 1-16. https://doi.org/10.1167/9.3.7

Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision 14*(3), 1-22. https://doi.org/10.1167/14.3.29

Castelhano, M. S., & Henderson, J. M. (2008). Stable individual difference across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology 62*(1), 1-14. https://doi.org/10.1037/1196-1961.62.1.1

Castelhano, M. S., Mack, M., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision 9*(3), 1-15. https://doi.org/10.1167/9.3.6

Crilly, N., Moultrie, J., & Clarkson, P. J. (2004). Seeing things: consumer response to the visual domain in product design. *Design Studies 25*(6), 547-577. https://doi.org/10.1016/j.destud.2004.03.001

Dong, Y., & Lee, K. (2008). A cross-cultural comparative study of users' perceptions of a webpage: with a focus on the cognitive styles of Chinese, Koreans and Americans. *International Journal of Design 2*(2), 19-30. http://www.ijdesign.org/index.php/IJDesign/article/view/267/163

Ehmke, C., & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. *In Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1* (BCS-HCI '07), Vol. 1. British Computer Society, Swinton, UK, UK, 119-128. https://dl.acm.org/citation.cfm?id=1531311

Fawcett, W., Ellingham, I., & Platt, S. (2008). Reconciling the architectural preferences of architects and the public: The ordered preference model. *Environment and Behavior 40*(5), 599-618. https://doi.org/10.1177/0013916507304695

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research 62*, 1-8. https://doi.org/10.1016/j.visres.2012.03.019

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning 63*(1), 3-42. https://doi.org/10.1007/s10994-006-6226-1.

Hammer, N., & Lengyel, S. (1991). Identifying semantic markers in design products: the use of eye movement recordings in industrial design. In R. Schmid, and D. Zambarbieri (Eds.), *Oculomotor Control and Cognitive Processes - Normal and Pathological Aspects*. Elsevier Science: Amsterdam, pp. 445-455.

Holmqvist, K., & Nystrom, M. (2011). *Eye Tracking: A comprehensive guide to methods and measures* (p. 23). New York: Oxford.

Just, M. A., & Carpenter, P. A. (1976). Role of eye-fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation 8*(2), 139-143. https://doi.org/10.3758/bf03201761

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixation to comprehension. *Psychol Rev*. 87. 329–354. http://dx.doi.org/10.1037/0033-295X.87.4.329

Kiefer, P., Giannopoulos, I., Raubal, M., & Duchowski, A. (2017). Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation 17*(1-2), 1-19. DOI: 10.1080/13875868.2016.1254634

Koivunen, K., Kukkonen, S., Lahtinen, S., Rantala, H., & Sharmin, S. (2004). Towards deeper understanding of how people perceive design in products, *Proceedings of computers in art and design education*, June 29 – July 1, Sweden: Copenhagen Business School. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.4055

Köhler, M., Falk, B., & Schmitt, R. (2015). Applying eye-tracking in Kansei engineering methodology for design evaluations in product development. *International Journal of Affective Engineering 14*(3), 241-251. https://doi.org/10.5057/ijae.ijae-d-15-00016

Kukkonen, S. (2005). Exploring eye tracking in design evaluation. *Joining Forces: International Conference on Design Research*, September 22-24, Helsinki, Finland. http://www.uiah.fi/joiningforces/papers/Kukkonen.pdf

Lee, S., Cinn, E., Yan, J., & Jung, J. (2015). Using an eyetracker to study three-dimensional environmental aesthetics: The impact of architectural elements and educational training on viewers' visual attention. *Journal of architectural and Planning Research 32*(2), 145-167.

Lorigo, L., Brynjarsdottir, H., Xia, L., Joachims, T., Gay, G., Granka, L., Pellacini, F., & Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology 59*(7), 1041-1052.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction 20*, 167-171.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision 14*(7):6, 1-11. https://doi.org/10.1167/14.7.6

Miall, R. C., & Tchalenko, J. (2001). A painter's eye movements: A study of eye and hand movement during portrait drawing. *Leonardo 34*(1), 35-40. https://doi.org/10.1162/002409401300052488

Nielson, J., & Pernice, K. (2010). *Eyetracking Web Usability*. New riders, Peachpit. Berkeley, California, USA.

Nodine, C. F., Locher, P. J., & Krupinski, E. A. (1993). The role of formal art training on perception and aesthetic judgement of art compositions. *Leonardo 26*(3), 219-227. https://doi.org/10.2307/1575815

Oakes, L. M. (2012). Advances in eye tracking in infancy research. *Infancy 17*(1), 1-8.

Park, J., DeLong, M., & Woods, E. (2012). Exploring product communication between the designer and the user through eye-tracking technology. *International Journal of Fashion Design 5*(1), 67-78. https://doi.org/10.1080/17543266.2011.633566

Prats, M., Garner, S., Jowers, I., McKay, A., & Pedreira, N. (2010). Interpretation of geometric shapes – an eye movement study. In *Symposium on Eye Tracking Research and Applications*, Mar 22-24, 2010, Austin, Texas. https://doi.org/10.1145/1743666.1743723

Reid, T. N., MacDonald, E. F., & Du, P. (2012). Impact of product design representation on customer judgement with associated eye gaze patterns. *Proceedings of the ASME 2012 IDETC/CIE*, August 12-15, 2012, Chicago IL USA. https://doi.org/10.1115/detc2012-70734

Simon, P. (2013*). Too Big to Ignore: The Business Case for Big Data*. Wiley. p. 89. https://doi.org/10.1002/9781119204039

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception 1*(1), 7-27. https://doi.org/10.1068/i0382

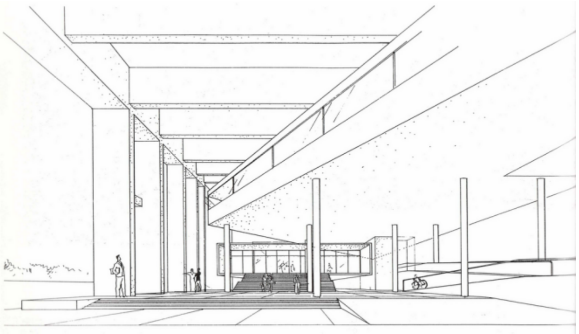
Venjakob, A. C., & Mello-Thoms, C. R. (2016). Review of prospects and challenges of eye tracking in volumetric imaging. *Journal of Medical Imaging 3*(1), 011002-1~011002-7.

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception 36*(1), 91-100. https://doi.org/10.1068/p5262

Weber, R., Choi, Y., & Stark, L. (2002). The impact of formal properties on eye movement during the perception of architecture. *Journal of Architectural and Planning Research 19*(1), 57-69.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press. https://doi.org/10.1007/978-1-4899-5379-7

# Appendix



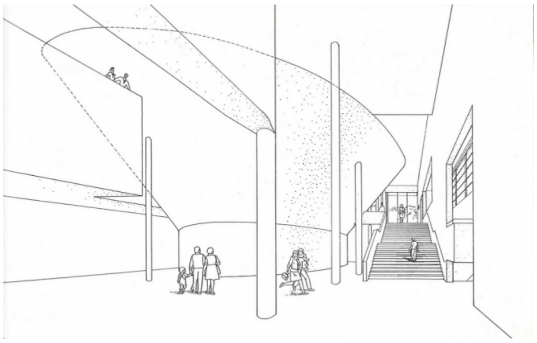
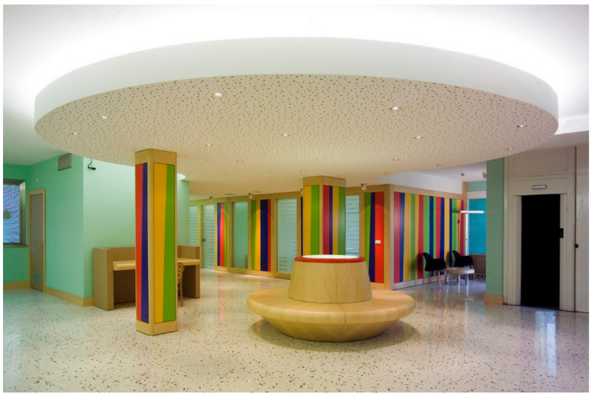


**Figure 1 (left) and 2 (right).** Kim H, Kim Y (1993) *Musee d'architecture*. Seoul: Baleon



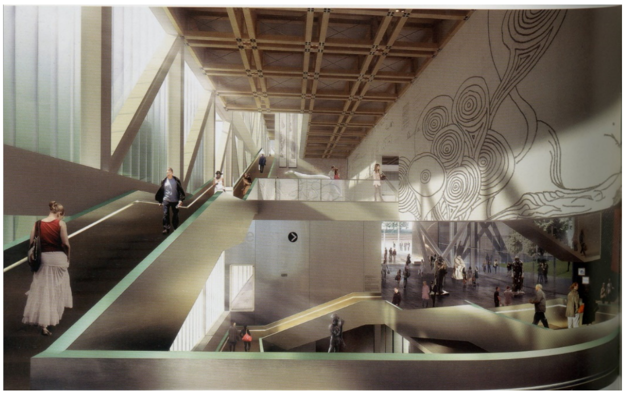
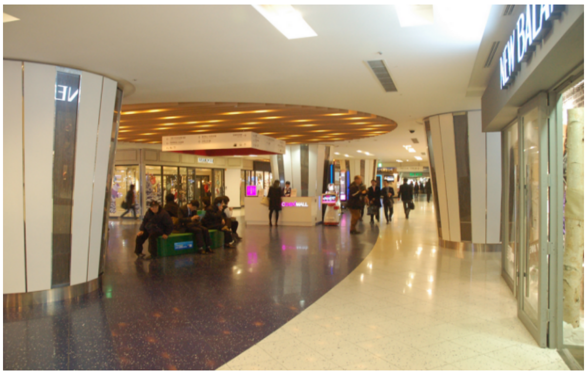


**Figure 3 (left)** Archdaily, May 3, 2009. "Bank in Donoratico / Massimo Mariani"

URL: https://www.archdaily.com/19127/bank-in-donoratico-massimo-mariani

**Figure 4 (right)** Musee National des Beaux-Arts du Quebec, OMA. Rendering by Luxigon

URL: https://www.flickr.com/photos/luxigon/4481481771/



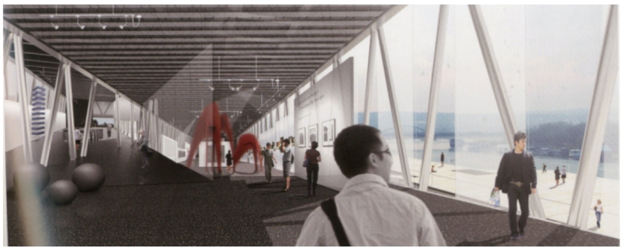
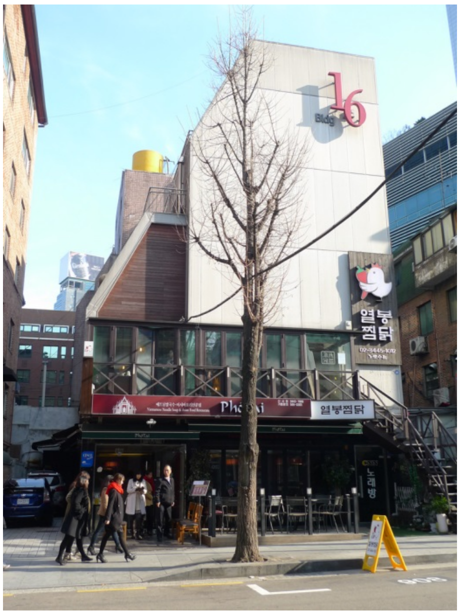


**Figure 5 (left)**

**Figure 6 (right)** Archdaily, June 2, 2011. "Beton Hala Waterfront / dRN Architects" written by Hank Jarz. Architect: Max Nuñez URL: https://www.archdaily.com/139047/beton-hala-waterfront-drn-architects



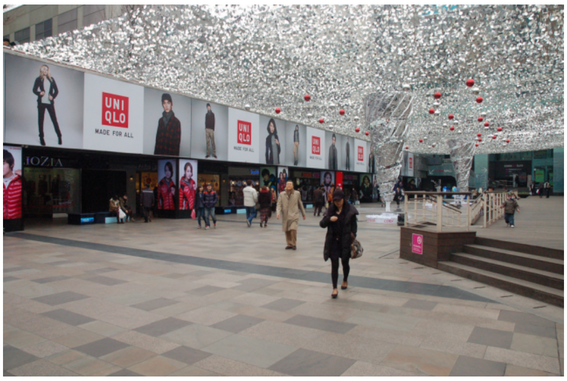
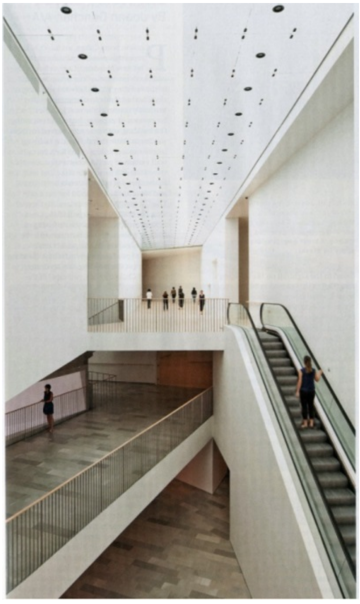


**Figure 7 (left) and 8 (right)**



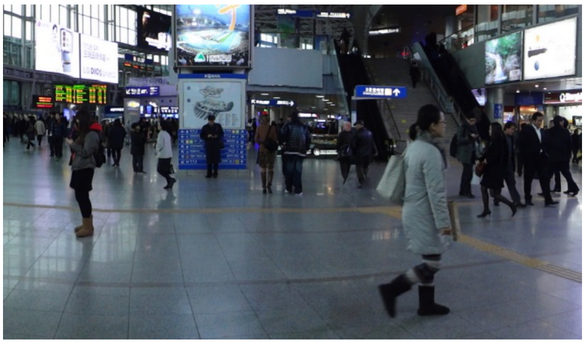
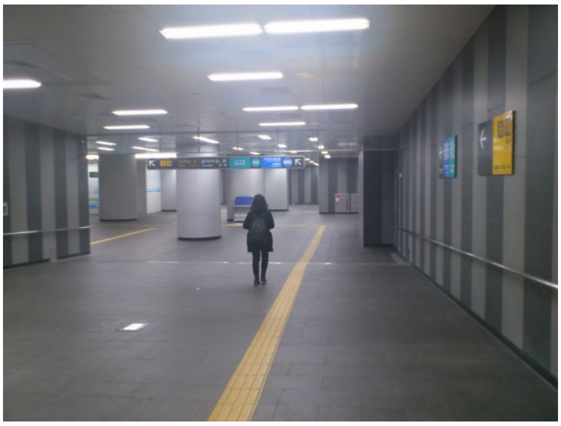


**Figure 9 (left)** Architectural Record, November 2011, Tel Aviv Museum of Art p79, architect: Preston Scott Cohen

**Figure 10 (right)**



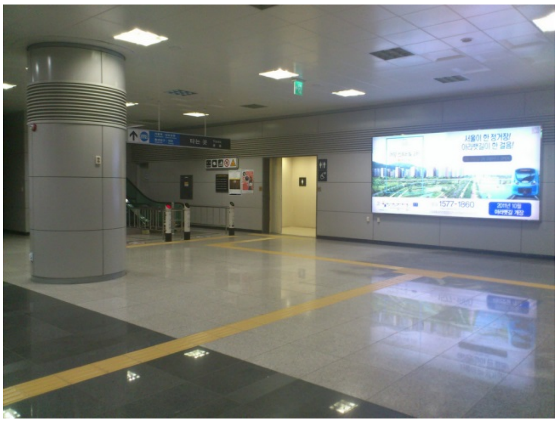
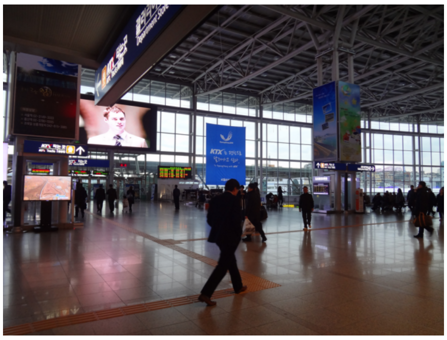


**Figure 11 (left) and 12 (right)**



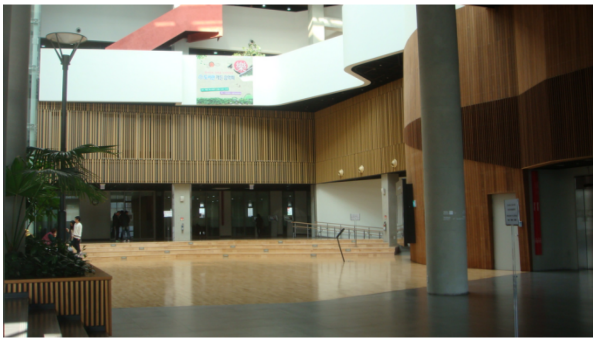


**Figure 13 (left) and 14 (right).**