UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI INFORMATICA

Impact of memory on periocular features

Corrente Sara, Di Martino Silvio

Abstract

Changes in eyes are connected to our subconscious mind. Investigating eyes and their parts have several advantages and find application in many fields of research (security, psychology and neuroscience). Among the characteristics of a human eye, we decided to consider fixations and saccades for a first analysis. Taking into particular consideration fixations and the data related to them for subsequent developments and final studies with more than one classifier.

The main goal of this work is to analyze data of 45 participants observing 48 images five times, compare and highlight the differences between the first and fifth iteration. More generally, the aim is to understand, depending on the gaze data, whether the image has already been seen or not.

1. Introduction

According to the last major research fields, the cognitive context includes all aspects related to mental processing, such as perception, memory, knowledge and learning.

In earlier works eye movement analysis has been introduced as a new modality for activity and context recognition: the movement patterns the eyes perform during different activities carry information that allows to recognise the activities themselves. A large body of research in experimental psychology has evidenced that, in addition to physical activity, visual behaviour is tightly linked to cognitive processes, such as attention, relational memory and learning.

The development of new technologies now makes possible the recording of gaze data and all its features: fixation points, durations, saccades, areas of interest, pupil width, pupil movement and so on.

These are the most important aspect to describe the human gaze:

- **saccade**: a quick, simultaneous movement of both eyes to shift a peripheral region to the middle of the visual field;

- **fixation**: the moments between one saccade and another during which the eye is still on the target. From the point of view of motor control, these are not "passive" periods, in which the eye, after a saccade, is left to its own: a gaze stabilization system comes into operation at the end of the saccade to maintain the target on the fovea centralis.
  So it indicates how people acquire information: a high number of fixations means difficulties in interpreting the information. A long fixation means a more demanding cognitive processing, more interest in what eyes are focusing on;

- **blink**: a usually involuntary shutting and opening of the eye. This feature is very important in

achieving information about the psychophysiological state of a person. A high rate of blinks usually means drowsiness, tedious task, unpleasant emotional state. Indeed a low rate means attention, pleasure, daydreaming;

- **area of interest**: in a picture, for example, a region that captures the human attention. Usually areas of interest focus on particular details of an image, that help people to interpret, recognize or memorize it.

The aim of our work is to analyze differences in the ocular behavior of 45 participants who were subjected to viewing a series of images repeatedly. Precisely, 45 participants freely observed 48 images in a randomized order and with five repetitions. They consecutively saw 5 blocks of all images, equally covered four categories, namely *Natural*, *Urban*, *Fractal*, and *Pink noise* images. The **final goal** is to interpret and assert if it is possible to exploit the periocular features to distinguish between new images and already seen ones.

2. Related work

Our work started with the evaluation of the state of art: various studies have already been conducted on similar goals, we have analyzed them trying to get ideas about features that can be extrapolated, techniques used and so on.

The study of Beuget, Castagnos, Luxembourger and Boyer tried to analyze the E-education scope and students' behaviour and difficulties in remembering the most important aspects of a lesson. Gaze data interests them more specifically as it has been reported that the gaze behaviour could reflect some cognitive processes. Their goal is to analyze if there could exist correlations between gaze characteristics and the fact to remember some courses items. They consider different features: normalized sum, mean and standard deviation of fixation duration, saccade horizontal amplitude, vertical amplitude, vectorial amplitude, absolute and relative angles, and pupil expansion. Thanks to an ANCOVA test (analysis of covariance), the results prove that there is a direct link between some gaze features and the memorization of images. This validates their main hypothesis: the analysis of those features can reveal which items are recalled, and which are not, during human-computer interactions.

The study of Borkin and al. [2] tried to analyze the eye fixations to examine which elements a person focuses on when visually encoding and retrieving a visualization from memory. The experiment was divided in 3 different phases:

- *Encoding Phase* in which participants examined each image for 10 seconds;
- *Recognition Phase* in which participants pressed the spacebar to indicate recognition of a visualization from the previous experimental phase;
- *Recall Phase* in which each recognized image was presented resized and blurred: the purpose of blurring visualizations was to allow the visualization to be recognizable, but not contain enough visual detail to enable the extraction of any new information. Next to each blurred visualization was an empty textbox with the instruction: "Describe the visualization in as much detail as possible". The goal was to elicit from the participant as much information about the visualization as they could recall from memory.

The results showed differences between the fixation heatmaps of the most and least recognizable visualizations in the recognition phase, where the most

recognizable visualizations have a fixation bias towards the center of the visualization. This indicates that a fixation near the center provides sufficient information to recognize the visualization without requiring further eye movements. In contrast, the least recognizable visualizations have fixation heatmaps that look more like the fixation patterns of visual exploration in the encoding phase.

The study of Bulling and Roggen[3] analyzed the gaze behaviour of 14 participants looking at familiar and unfamiliar pictures from four different categories: abstract, landscapes, faces and buildings. Saccade, blink and fixation detection have been performed. About the techniques used, the 2 researchers used a recognition methodology that combines minimum redundancy maximum relevance feature selection (mRMR) with a support vector machine (SVM) classifier.
The datasets of all but one participant were combined and used for training (the "training set"); the dataset of the remaining participant was used for testing (the "test set"). This was repeated for each participant.
Results showed that looking at familiar and unfamiliar pictures can be recognised from eye movements of participants with decent performance. These initial results are promising as the described approach may soon be applicable to other stationary real-world setups.

The study of Holland and Komogortsev [4] presents an objective evaluation of various eye movement-based biometric features and their ability to accurately and precisely distinguish unique individuals. Considered biometric candidates cover a number of basic eye movements and their aggregated scan path characteristics, including: fixation count, average fixation duration, average, saccade amplitudes, average saccade velocities, average, saccade peak velocities, the velocity waveform, scanpath, length, scanpath area, regions of interest and so on. The ML techniques used were Naive Bayes, Decision Tree and K-Nearest Neighbor, each of them with different accuracy results. However, biometric traits are becoming easier to reproduce, circumventing the purposes of existing biometric identification techniques and leaving gaps in the efficacy of the systems that use them. Scanpath theory presents a unique solution, as eye movements are uniquely counterfeit resistant due to the complex neurological interactions and extraocular muscle properties involved in their generation.
This paper has presented an objective evaluation of a number of scanpath-based biometric features and their ability to accurately and precisely distinguish unique individuals, with equal error rates ranging from 300/0-4 9%. As well, these researchers have presented an information fusion method which allows for the combination of multiple metrics to produce more stable/accurate identification.

The study of Marchal, Castagnos and Boyer [5] tried to to highlight the link between gaze features and visual memory. Our protocol consisted in asking different subjects to remember a large set of images. During this memory test, we collected about 19000 fixation points. Among other results, they show in this paper a strong correlation between the relative path angles and the memorized items. They then applied the Logistic Regression classifier and showed that it is possible to predict the users' memory status by analyzing their gaze data. This is the first step so as to provide recommendations that fit users' learning curve.
Also in this case, results show that there was a strong correlation between some gaze features (number of fixations, sum of the relative angles of the scan path) and the fact to memorize items.

According to these papers, fixations and saccades are the most important features involved in image recognition: many studies highlighted the link between them

and the memory recall and so we would consider just them in our dataset.

3. Proposed section

3.1 Dataset description

For our study we used the dataset [6] Memory I, that consists of (x,y) gaze location entries for individual fixations of 45 participants that freely observed 48 images in a randomized order and with five repetitions. They consecutively saw 5 blocks of all images. The block number is coded as 'iteration'. The images equally covered four categories, namely *Natural*, *Urban*, *Fractal*, and *Pink noise* images: the first two categories can be categorized in turn as *clear images*, just because their content is easily recognizable to the human eye; the second two categories can be categorized as *unclear images*, because their content is noisy and not very clear. Presentation duration was 6s for each image. Before an image appeared, participants had to fixate on a cross presented in the center of the screen. A short 5 minute break after the third presentation block maintained participants' alertness and avoided potential fatigue.

Fixation coordinates were given in pixels with respect to the monitor coordinates (the upper left corner of the screen was (0,0) and down/right was positive). Fixations were labeled with a subject ID, start and end times, image category and image number, the ordinal rank of the fixation within a trial, the trial within an experimental session, and a dataset ID that refers to the source study. In summary:

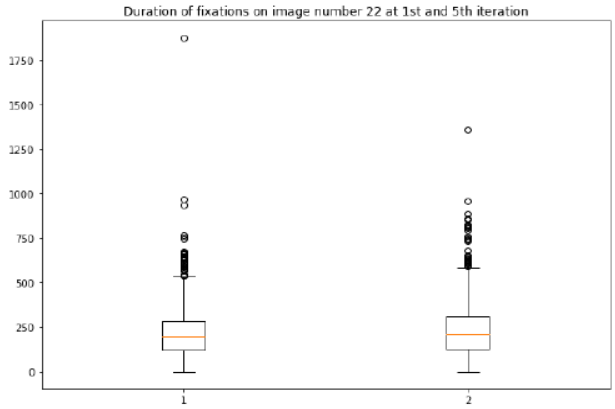
SUBJECT INDEX	Id of the participant
CATEGORY	Image category
EYE	0.0 or 1.0 for left and right pupil
FILENUMBER	Code number of the image within its category
PUPIL	Pupil size
SAMPLE	Ordinal rank of the sample within a trial
TIME	Timestamp in milliseconds
TRIAL	Trial number within experiment
(x,y)	Coordinates in pixels

3.2 Data pre-processing

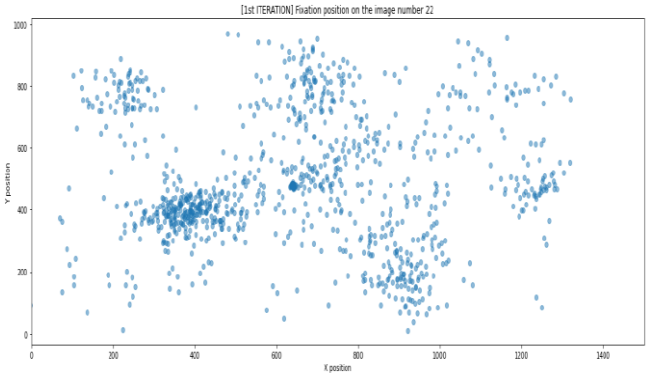
First of all, we are not interested in the whole dataset: in fact, it considered all the data of five iterations.
As we already said, we are interested in 1st iteration data and 5th iteration data.

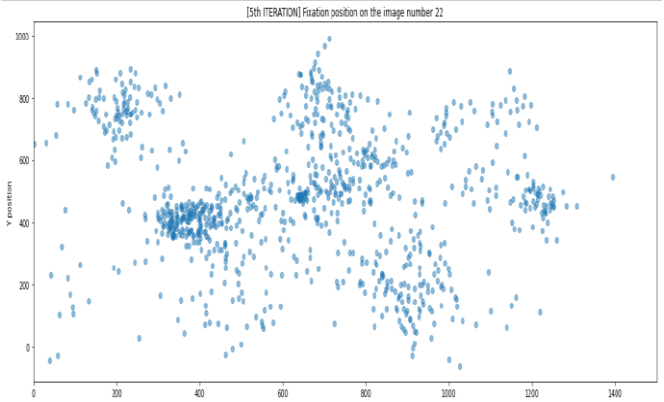
In an earlier analisys we focused on fixations detection on an image with clear content in the 1st and 5th iteration (*filenumber 22, Natural).*
Then we calculated the duration of each fixation and its quartiles, so we used the boxplots to get a clear overview of the results.
Boxplots allow to represent on the same graph five of the most used position measures in statistics. They show the median, upper and lower quartiles, minimum and maximum values, and any outliers.

The height of the box shown is equal to the interquartile range (IQR) and contains the central 50% of the observations made, those between the first and third quartiles. The line inside the box instead represents the median. The two segments that start from the box and extend upwards and downwards are called "whiskers": whiskers indicate the dispersion of values below the first quartile and above the third quartile not classified as outliers. Isolated points indicate possible outliers.



We can see that there are many outliers, in both the boxplots. Analysing them thoroughly, we found that this behaviour was common in all participants, both on the first and fifth iterations. So we deduced that there must be something about that image that captures participants' attention. After duration analysis, we calculated all the positions of the fixations and plotted the results of 1st and 5th iteration.

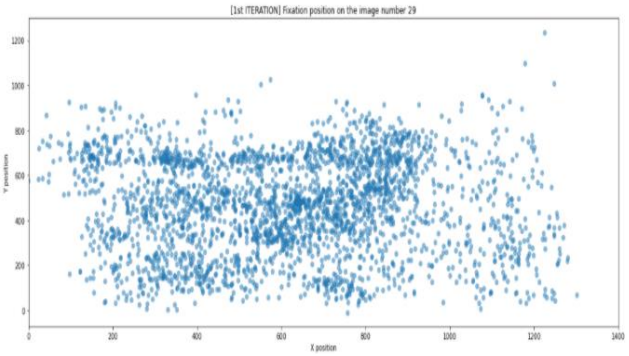


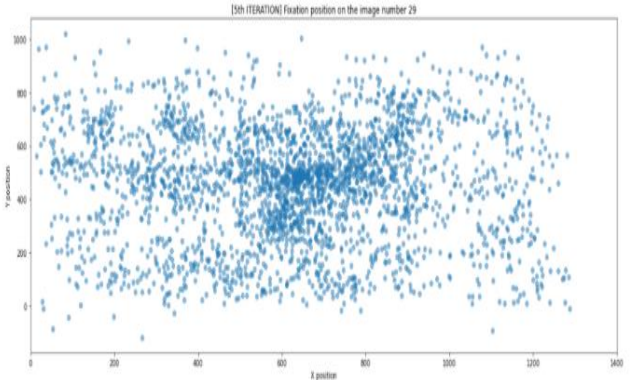


As we can see fixations tend to aggregate mainly in some areas that we will call as "areas of interest". These are roughly identified just plotting the end positions of fixation themselves. Later, we will perform a more precise analysis on these positions, thanks to a clustering alghoritm.

For the sake of completeness we performed the same work on the saccades: since fixations are gaze stabilizations between two different saccades, we did not notice interesting data to report, compared to fixation analysis.

Since nothing new got our attention, we performed only the fixations' end positions analysis on another image of unclear content (*filenumber 29, Pink noise).*





We can certainly and immediately notice a greater number of fixations than the previously analyzed image: these fixations almost concentrate themselves in the whole image! This huge number, in conjunction with the fixations' distribution, indicate that this image is full of details and also not so clear: it requires strong concentration of participants to understand and memorize it.
Clustering will help us and will give us more accurate information about these area of interest.

3.3 Clustering with DBSCAN

At this point, we decided to exploit and investigate position of fixations, as in our opinion it could give us meaningful information for our purpose. So we performed clustering by relying on data of each participant for each image at 1st and 5th iteration.

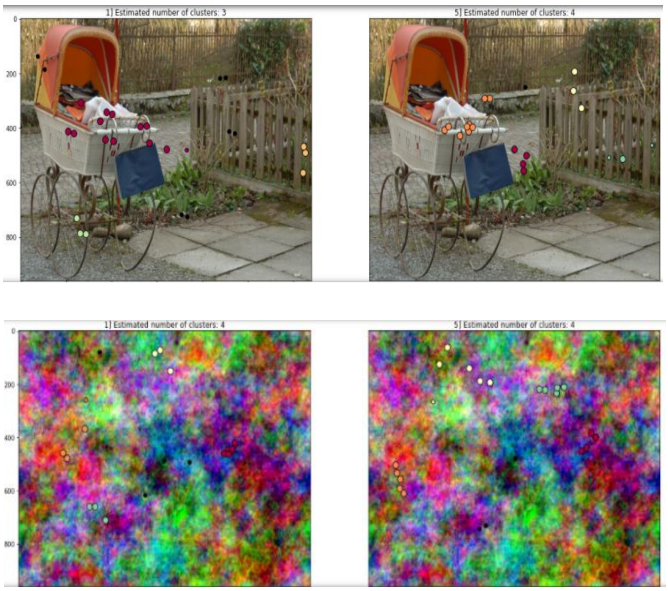
To do this, we used Density-based spatial clustering of applications with noise, also called DBSCAN. This is a data clustering algorithm proposed by Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu in 1996. It is a density-based clustering non-parametric algorithm: given a set of points in some space (our position of fixations), it groups together points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are too far away). DBSCAN is one of the most common clustering algorithms and also most cited in scientific literature. [7]
The DBSCAN algorithm basically requires 2 parameters:

- **eps**: specifies how close points should be to each other to be considered a part of a cluster. It means that if the distance between two points is lower or equal to this value (eps), these points are considered neighbors;
- **minPoints**: the minimum number of points to form a dense region.

The parameter estimation is a problem for every data mining task. To choose good parameters we need to understand how they are used and have at least a basic previous knowledge about the data set that will be used.

Thanks to our previous analysis on fixations, after performing tuning on these parameters, we finished with a good and realistic result in terms of cluster and noise point identification.

To follow, some examples of clustering on some of analysed images:



In addition to the number of clusters and noise points of each image, we calculated also another parameter, called *Silhouette score* or *Silhouette coefficient.*
It is calculated using the mean intra-cluster distance (*a*) and the mean nearest-cluster distance (*b*) for each sample. So the Silhouette Coefficient for a sample is:

(b - a) / max(a, b)

To clarify, *b* is the distance between a sample and the nearest cluster that the sample is not a part of. Silhouette Coefficient or silhouette score is a metric used to calculate the goodness of a clustering technique.

Its value ranges from -1 to 1:

- **1**: Means clusters are well apart from each other and clearly distinguished;
- **0**: Means clusters are indifferent, or we can say that the distance between clusters is not significant;
- **-1**: Means clusters are assigned in the wrong way.

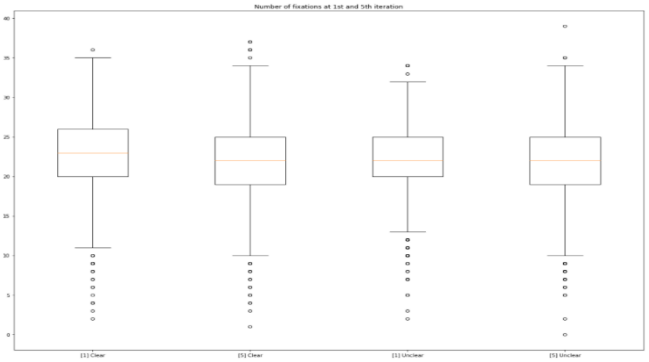
3.4 Comparison between clear and unclear images

At this point, we compared clear and unclear images based on various data we have come up after previous explained analysis.
In particular, we made a comparison with these information calculated for each participant for each image to highlight differences not only relying on the content of the image, but also to point out diversities between the 1st and the 5th iteration. These are considered data:

- **Number of fixation**
- **Mean duration of fixation**
- **Number of clusters**
- **Number of noise points**

For each of these information, we used boxplots to analyse the value distribution and how it changes between clear and unclear image and between the two iterations.
After that we calculated the percentage of how often a certain behaviour occurred (for example, how many times the number of fixation at 1st iteration was greater than the same data at 5th iteration, and vice versa).

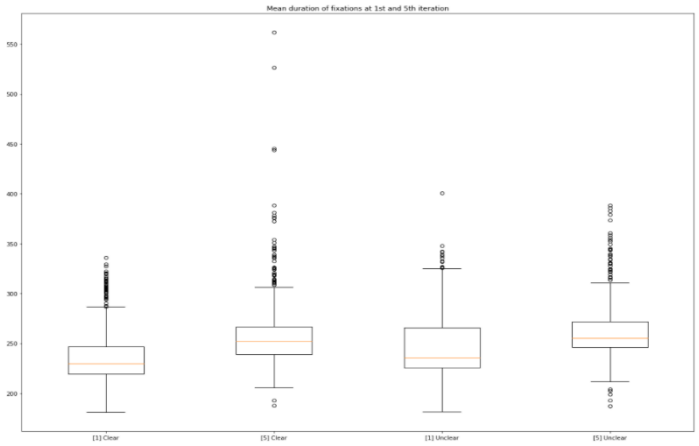
3.4.1. Number of fixation



In percentage:

- (**53, 2**) % of times that number of fixations at 1st iteration is greater than number of fixations at 5th iteration;
- **[CLEAR IMAGES] (54, 2)** % of times that number of fixations at 1st iteration is greater than number of fixations at 5th iteration;
- **[UNCLEAR IMAGES] (51, 2)** % of times that number of fixations at 1st iteration is greater than number of fixations at 5th iteration.

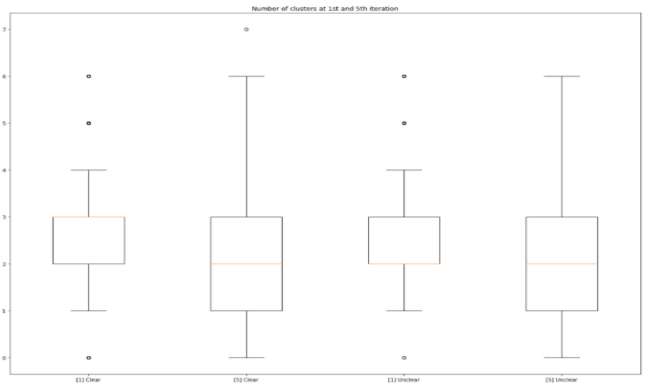
3.4.2 Mean duration of fixation



In percentage:

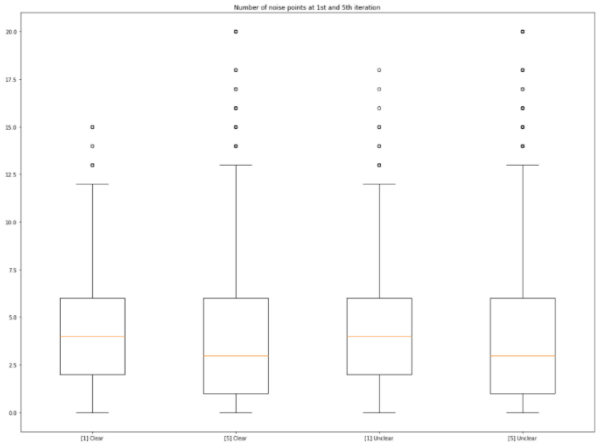
- **(80, 2)** % of times that mean duration of fixations of 5th iteration is greater than mean duration of fixations of 1st iteration;
- **[CLEAR IMAGES] (86, 2)** % of times that mean duration of fixations of 5th iteration is greater than mean duration of fixations of 1st iteration;
- **[UNCLEAR IMAGES] (70, 2)** % of times that mean duration of fixations of 5th iteration is greater than mean duration of fixations of 1st iteration.

3.4.3 Number of clusters



- **(71, 2)** % of times that cluster of 1st iteration are equal to or more than clusters of 5th iteration;
- **[CLEAR IMAGES] (72, 2)** % of times that clusters at 1st iteration are equal to or more than cluster at 5th iteration;
- **[UNCLEAR IMAGES] (69, 2)** % of times that clusters at 1st iteration are equal to or more than cluster at 5th iteration.

3.4.4 Number of noise points



In percentage:

- **(58, 2)** % of times that noise points of 1st iteration are equal to or more than noise points of 5th iteration;
- **[CLEAR IMAGES] (59, 2)** % of times that noise points at 1st iteration are equal to or more than noise points at 5th iteration;
- **[UNCLEAR IMAGES] (58, 2)** % of times that noise points at 1st iteration are equal to or more than noise points at 5th iteration.

3.5 Silhouette score

About this coefficient, given the range *[-1,1]* of possible results, we calculated the number of occurrences for each sub-interval included in the range.

In particular, about the 1st iteration:

	<-0.5	]-0.5,0]	]0,0.5]	]0.5,1]
ALL	3.52 %	0.05 %	53.77 %	42.66 %
CLEAR	3.28 %	0.11 %	54.07 %	42.54 %
UNCLEAR	2.96 %	0.0 %	58.52 %	38.52 %

About the 5th iteration:

	<-0.5	]-0.5,0]	]0,0.5]	]0.5,1]
ALL	9.08 %	0.09 %	42.15 %	48.68 %
CLEAR	8.25 %	0.11 %	43.07 %	48.57 %
UNCLEAR	6.96 %	0.15 %	46.67 %	46.22 %

Let's focus on some data. About the *clear images*:

- at 1st iteration we have 54% of silhouette score between 0 and 0.5 and 42% of silhouette score greater than 0.5;
- At 5th iteration we have 43 of silhouette score between 0 and 0.5 48 % of silhouette score greater than 0.5.

So we have a decrease in terms of clusters quitely distinguished and an increase of clusters well distinguished. What does it changes? Clusters become more distinguished and separated, as participants' fixation are more clustered: they already know which details focusing on, and they don't spread their gaze in the whole picture.

About the *unclear images:*

- at 1st iteration we have 58% of silhouette score between 0 and 0.5 and 38% of silhouette score greater than 0.5;
- At 5th iteration we have 46of silhouette score between 0 and 0.5 46% of silhouette score greater than 0.5.

We have a decrease in terms of clusters quitely distinguished but an increase of clusters well distinguished: different values but same trend. Clusters become more distinguished!

4. Machine Learning strategy

At this point, we have enough data to perform the binary classification between familiar and unfamiliar images. These are the features that have been extracted for our dataset:

- **Image category** (0 for clear images, 1 for unclear)
- **Number of fixation**
- **Mean duration of fixation**
- **Number of clusters**
- **Number of noise points**
- **Silhouette score**
- **CLASS:** 0 for data of 1st iteration (familiar images), 1 otherwise.

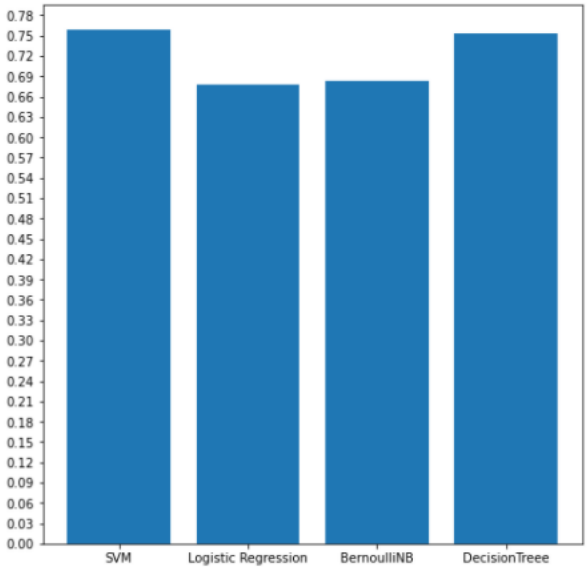
Dataset preparation has been done by standardizing it along axis and splitting it in two different parts: 70% for training and 30% for testing

Various classifiers have been considered, also considering those used in the related works previously analysed: *Support Vector Machine* (**SVM**), *Logistic Regression* (**LR**), *Bernoulli Naïve Bayes* (**BNB**), *Decision Tree* (**DT**).
For all these techniques we started by applying the GridSearchCV function to obtain the best hyperparameters.

5. Result section

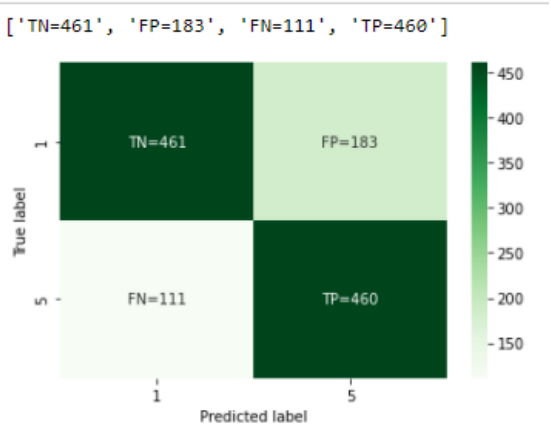
CLASSIFIER	Accuracy	Precision	Recall	F1-Score	Iteration
SVM	76%	81%	72%	76%	1
		72%	81%	76%	5
LR	68%	69%	72%	70%	1
		67%	63%	65%	5
BNB	68%	69%	73%	71%	1
		67%	64%	65%	5
DT	75%	81%	70%	75%	1
		71%	81%	76%	5

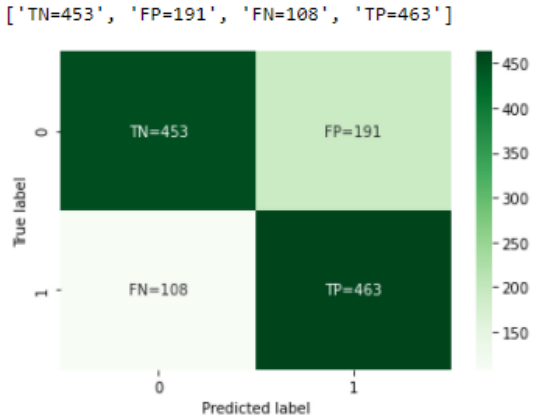
Let's compare further the accuracy results:

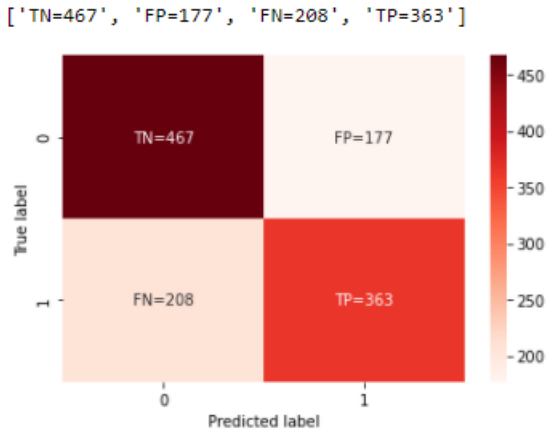


We can see that the best results are obtained by SVM and Decision Tree, meanwhile Logistic Regression has slightly the worst results.

We report the confusion matrix respectively for SVM, DT and BNB:







At this point, we tried to perform the binary classification also with an *Artificial Neural Network*: it has three simple layers.

- **Input layer:** 6 nodes, one for each column of dataset, except the CLASS column;
- **Hidden layer:** 4 nodes, the mean between number of input layer and output layer;
- **Output layer:** 1 node, because our goal is the binary classification that requires only one output.

Also in this case we started by applying the GridSearchCV function to obtain the best hyperparameters, and these are the results:

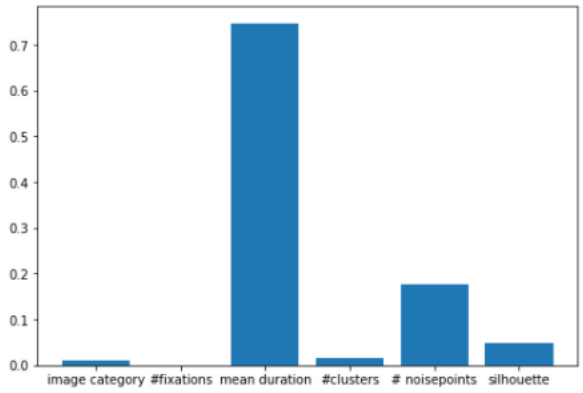
Accuracy	Precision	Recall	F1-Score	Iteration
74%	76%	73%	75%	1
	71%	74%	73%	5

A good result in accuracy terms, better than LR and BNB. However, DT and SVM obtain best results.

6. Discussion section

As stated with the reported results, Support Vector Machine and Decision Tree scored the best performances, but it was not a surprise since these classifiers has been reported to have good results in the considered related works.

This study highlighted differences in gaze behaviour of 45 participants: we can say that the number of noise points and the Silhouette score pointed out important differences between 1st and 5th iteration, but the mean duration mattered most. Here a graph that plots the importance that each feature had in the classification:



It's important to underline that this study has been conducted only on the fixations of each participant: further studies can

be done on the same dataset, including blink, saccades and other periocular features to extract other information about this interesting topic.

REFERENCES

[1] **Eye Gaze Sequence Analysis to Model Memory in E-education,** Mael Beugeti, Sylvain Castagnos, Christophe Luxembourger and Anne Boyer CNRS-LORIA-University of Lorraine, Vandoeuvre-l`es-Nancy, France. 2LPN-University of Lorraine, Nancy, France

[2] **Beyond Memorability: Visualization Recognition and Recall**, Michelle A. Borkin*, *Member, IEEE*, Zoya Bylinskii*, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, *Senior Member, IEEE*, and Aude Oliva

[3] **Recognition of Visual Memory Recall Processes Using Eye Movement Analysis,** Andreas Bulling Computer Laboratory University of Cambridge andreas.bulling@acm.org, Daniel Roggen Wearable Computing Laboratory ETH Zurich droggen@ife.ee.ethz.ch

[4] **Biometric Identification via Eye Movement Scan paths in Reading**
Corey Holland, Oleg V. Komogortsev, Department of Computer Science
Texas State University - San Marcos, TX 78666 USA, ch1570@txstate.edu, okll@txstate.edu

[5] **First Attempt to Predict User Memory from Gaze Data**
Florian Marchal, Sylvain Castagnos and Anne Boyer, *University of Lorraine – CNRS – LORIA, Campus Scientifique B. P. 239*

[6] **An extensive dataset of eye movements during viewing of complex images**
Niklas Wilming, Selim Onat, José P. Ossandón, Alper Açık, Tim C. Kietzmann, Kai Kaspar, Ricardo R. Gameiro, Alexandra Vormberg e Peter König

[7]**DBSCAN**
"https://en.wikipedia.org/wiki/DBSCAN"