Lab 2 - Report

TNM098 - Advanced Visual Data Analysis

Isabelle Rosenquist (isaro242)
Sandra Pettersson (sanpe282)

The task

The given data file consist of data from an eye-tracking session. The session lasted for just under 5 minutes. To simplify the process, the data has been pre-processed to only contain the fixations instead of every eye sample.

The data is divided into 6 fields. These fields are a timestamp for the event, a fixation index, the event duration in milliseconds, a gaze point index and the X and Y coordinate of the gaze points in pixels. The task is to identify interesting regions and observe them over time.

The questions we decided to focus on were the following.

- How many regions can be identified?
- How many of the regions are heavily used and when?
- What are the frequent transitions between the areas of interest?

Method

First, we attempted to visualize the coordinates from the eye-tracking data in a scatterplot. The axes has the same dimensions as the screen that was used for the eye-tracking. To visualize the aspect of the total time of screentime, the scatters was divided into time slots of 1 min. Each minute represented by a color in the plot. This was made just to see some initial subdivisions between clusters and time and were not used in the final visualization.

To see the event duration for each data point the size parameter for each dot were set to depend on the event duration. A bigger dot is equivalent to more time spent in that area.

To find and view clusters in the data there are several different methods. From the first plot some noise was discovered, therefore a cluster method which is considering noise would be the best option. The method *DBSCAN*, Density-based spatial clustering of applications with noise was used [1]. In the cluster algorithm *DBSCAN* there are two input parameter needed to be determined before visualizing the clusters. The first parameter is *eps*, the minimum distance between two points. If the distance between two points is smaller than *eps*, those points are considered neighbours and belongs to the same cluster. The second parameter is *MinPts,* the minimum points to form a cluster. If several points are neighbours according to *eps*, but there is less points than *MinPts*, its considered as noise, otherwise it is a cluster.

With clusters determined from DBSCAN, the next interesting thing to see is the time of the event in each cluster. A timeline of each cluster makes it possible to see whether there is some clusters that are heavily used at some point.

To see the different transitions both between the clusters and inside the clusters an animation were made. It shows in which order the coordinates were looked at. The transition speed however was not taken into consideration.

Result

Five clusters were found using the *DBSCAN*. These can be seen in Figure 1. The variables used were *eps* = 60 and *MinPts* = 10.

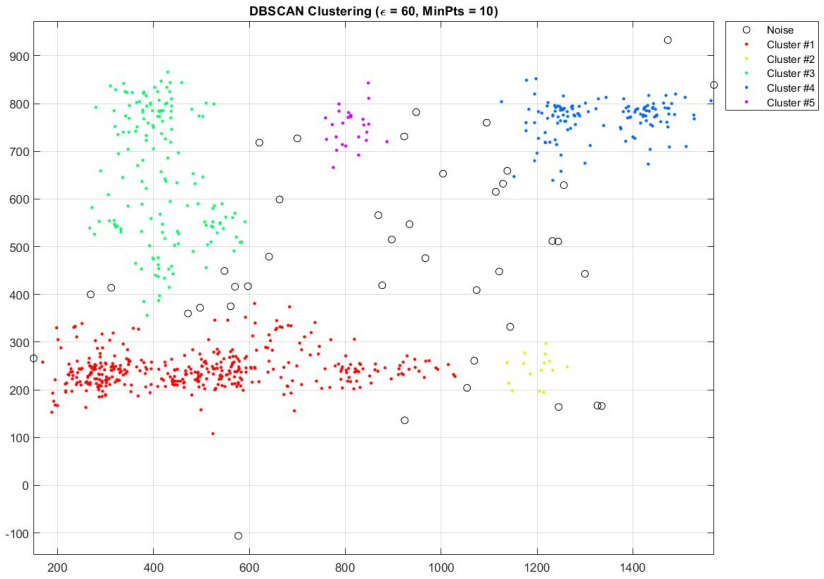

Figure 1: Visualization of the clusters and outliers found using DBSCAN.

Based on the clusters obtained using the *DBSCAN* we can use a timeline to determine when and how long the focus was on a certain cluster. This can be seen in Figure 2.

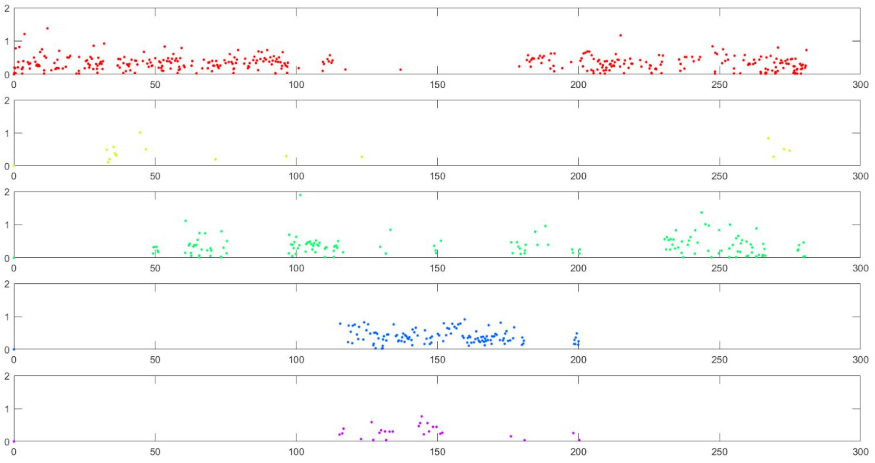

Figure 2: Visualization of when and how long the clusters were used.

When animating the coordinates from the data certain patterns were found. These can be seen in Figure 3.

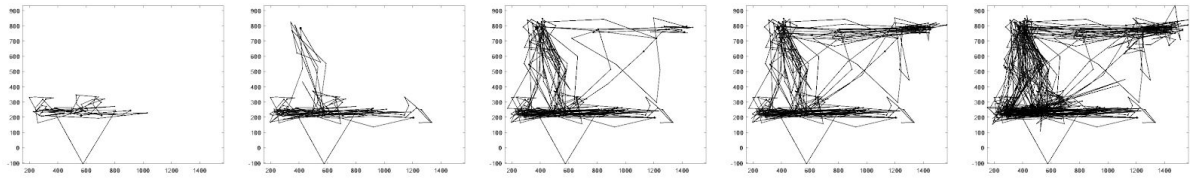

Figure 3: Still frames from an animation displaying when a certain point is looked at.

Conclusion

After these visualization methods were used we should be able to answer the questions we focused on.

How many regions can be identified?

With the variables used five clusters could be found. Two rather large clusters, one medium sized and two small ones. Around 40 outliers were also discovered.

How many of the regions are heavily used and when?

Some of the clusters were more heavily used than others. *Cluster #1* were the most used cluster and were used both in the beginning and the end of the session. *Cluster #4* were used only in the middle of the session. The other clusters had no pattern.

What are the frequent transitions between the areas of interest?

The animation gave some answers to this question. In *Cluster #1* there were mostly movements side to side, the same goes for *Cluster #4.* There were also were many movements back and forth between *Cluster #1* and *Cluster #3*.

References

[1] Towards Data Science. Apr 1, 2017. *How DBSCAN works and why should we use it?* [ONLINE] Available at: https://towardsdatascience.com/how-dbscan-works-and-why-should-i-use-it-443b4a191c80 [Accessed 15 May 2018].