Using Data Visualization to Deduce Faces Expressions

Elizabeth Simão Carvalho* and Paulo Miguel F. C. de Brito+

* *University Aberta/CIAC, Rua da Escola Politécnica 141-147, 1269-001 Lisboa,Portugal*
+ *University of Minho/ALGORITMI, Campus de Azurém, 4800-058 Guimarães, Portugal*

Abstract

Collect and examine in real time multi modal sensor data of a human face, is an important problem in computer vision, with applications in medical and monitoring analysis, entertainment and security. Although its advances, there are still many open issues in terms of the identification of the facial expression. Different algorithms and approaches have been developed to find out patterns and characteristics that can help the automatic expression identification. One way to study data is through data visualizations. Data visualization turns numbers and letters into aesthetically pleasing visuals, making it easy to recognize patterns and find exceptions. In this article, we use information visualization as a tool to analyse data points and find out possible existing patterns in four different facial expressions.

Key Words: Facial expressions, Data visualization, Pattern detection, Kinect.

1 Introduction

Facial expressions are currently used for inferring specific patterns, but they also can be used to indicate mental states. [1] Looked at facial expressions while students worked with an online tutoring system and identified that frustration was associated with activity in the inner and outer brow raiser and dimple; confusion was associated with brow lowered, lip tightened and lip corner puller. Moreover, preliminary results by [2] suggest that high and low stressor situations could be discriminated based on facial activity in mouth and eyebrow regions. In addition, stress can be inferred using multimodal sensor data as stated in a survey. So vast amounts of data from different sensors can be used to try to infer human facial affective states or patterns. Interpreting this data in a meaningful way is challenging.

Patterns or insights may go unnoticed in a data spreadsheet. But if we put the same information on a pie chart, the insights become obvious. Data visualization allows us to quickly interpret the data and adjust different variables to see their effect and technology. It is increasingly making it easier for us to do so. Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

In this article, we will use data visualization as a tool to help us to find out possible existing patterns in the captured sensor data. Sections two and three will give an outline of the scientific background of this work, including a summary of the information visualization area. Section three will describe the system architecture used to capture the data. Section four will present the data analysis and highlight some of the main outcomes. Finally, section five, points out future work and the core conclusions.

2 Facial Tracking

Images containing faces of people are very used in human-computer interaction and in human expression analysis. Because of this, several research groups are focusing on facial recognition, face tracking, assessment of posture and expression recognition. On average, the face detection algorithms can accurately detect faces in relation to their position in an image. A survey of the state of the art on the face detection can be found in [3][4]. Figure 1 illustrates an output given by a Microsoft.Kinect.Face [5] API.

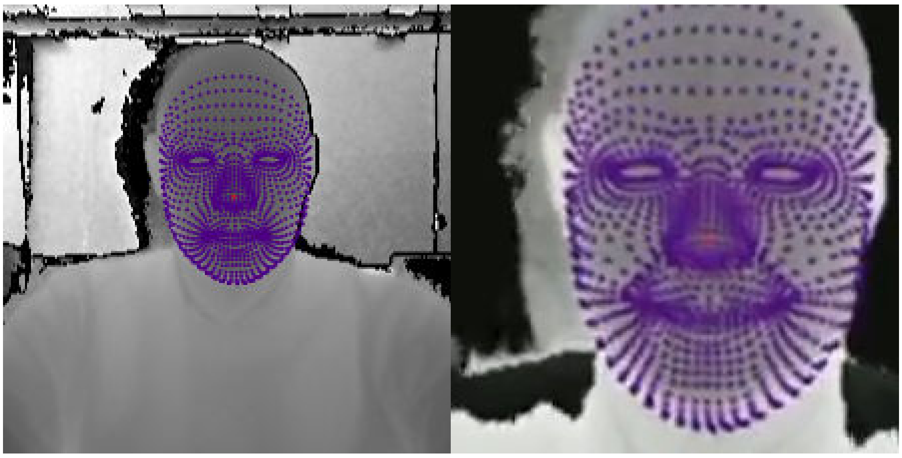

Figure 1. Face expression control points

To develop automated systems that analyse the information in facial images, face detection algorithms have to be efficient and robust. The aim of face detection, given a single image, is to identify all regions in the image that potentially contain a face regardless of their dimensional position, orientation and lighting conditions. This problem is demanding because the faces are not rigid objects and have a high degree of variability in size, shape, colour and texture.

The face detection and tracking is often the first step in applications such as video surveillance and facial recognition. Location and tracking of human faces is a requirement of tracking facial and / or analysis of facial expressions, although often it is assumed that a face is available in the image in question.

Recently, due to improved processing performance of the personal computers, the massive use of webcams, the lower prices and the small size of devices like projectors or even 3d cameras, we can build new Human-Computer Interaction (HCI) systems, combining the strengths and new method approaches of facial tracking.

However, for a real-time tracking, many problems must be solved. A major problem lies in the variety of existing features such as colour of skin or eyes, beard and glasses. Another problem is the system response time. Whether a system is able to recognize and make the tracking of a face is not able to do so within a tolerable time frame. In addition, the tracking has to be accurate and robust for practical use. These problems increase the difficulty for the recognition and tracking of facial features in real time.

Applications that use face tracking algorithms require a fast tracking system, affordable and, most importantly, robust. The confidence in these factors must be high enough to allow a user the convenience and flexibility to perform natural head movements.

Approaches to facial tracking systems can be separated into two classes: based on the images and based on the characteristics approach [6].

Based on images, general face features are used such as skin colour, head geometry and motion. These evidences are robust to rotation, scale, and do not require high image quality. On the other hand, these approaches lack precision and therefore cannot be used to accurately pinpoint specific features.

For a smooth and accurate tracking is used the characteristics approach [7]. These approaches are based on tracking individual facial features. The tracking characteristics can be obtained accurately to the pixel, which allows a direct and precise mapping. The drawback of these approaches is that they usually require expensive cameras and high resolution. Moreover, these are not robust to head movements, especially rotation and scale.

Recently it was showed that the robustness of tracking based on individual characteristics of the face can be significantly improved, if instead of using features such as edges and corners of eyes, mouth and nostrils, they use features based on curvature of the nose [8]. This creates a new range of interesting possibilities for tracking facial features based on face. More compact storage which does not deteriorate.

3 Scope

Visualization is suitable when there is a need to augment [9] human capabilities rather than replace people with computational decision-making methods. The design space of possible vis idioms is huge, and includes the considerations of both how to create and how to interact with visual representations. The visual system provides a very high-bandwidth channel to our brains. A significant amount of visual information processing occurs in parallel at the preconscious level.

Accordingly, from researchers to brand strategists, financial [10] analysts and human resource managers, better understanding and analysis of data/information is becoming an increasingly powerful way for further growth, productivity, and innovation. Moreover, we see average users, including consumers, citizens, and patients, examine public data such as product specifications, blogs, and online communities to choose products to buy, decide issues to vote on, and seek health-related information. Recent advances in InfoVis technologies provide an effective avenue to address the current and future “glut” of information faced by today’s users.

Compared to numerical and textual formats, it is known that data visualizations can highlight relationships in the data, facilitate the recognition of patterns, and reduce cognitive load [11]. As they said data exploration and understanding, it is generally assumed that data visualizations can support better decision making. Based on this intuition, several decision-support systems that rely on interactive data visualization have been developed.

Certain general analysis techniques [12] work well for certain types of questions—scatterplots for finding correlations, bar charts for ranking and comparison, and so on. There are several visualization techniques available to map different kinds of data. Depending on data structure and visualization goals, certain visualization techniques are more suitable than others.

The vast majority of Treemap algorithms [13] have focused on rectangular regions that are subdivided into smaller rectangular cells. However, non-rectangular regions have been examined. The Voronoi Treemap (VT) algorithm differs significantly from its predecessors. It produces non-rectangular spatial subdivisions that exhibit high quality aspect ratios that are close to one. However, it suffers from certain properties that make it unsuitable for animating the display of changing data values. Figure 2 illustrates a Voroni tessellation.

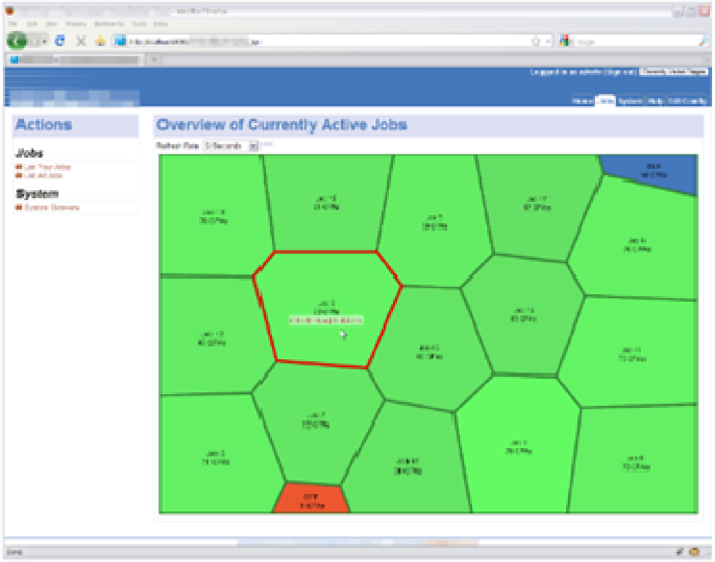

Figure 2. Voroni tessellation [13]

Binning is a technique of data aggregation used for grouping a dataset of N values into less than N discrete groups. There are many reasons for using hexagons instead of squares for binning a 2D surface as a plane. The most evident is that hexagons are more similar to circle than square. This translates in more efficient data aggregation around the bin centre. Huang and Hu used a density plot based on hexagonal binning to examine how people make friends in social networking sites [14]. Figure 3 shows a hexagonal binning plot.

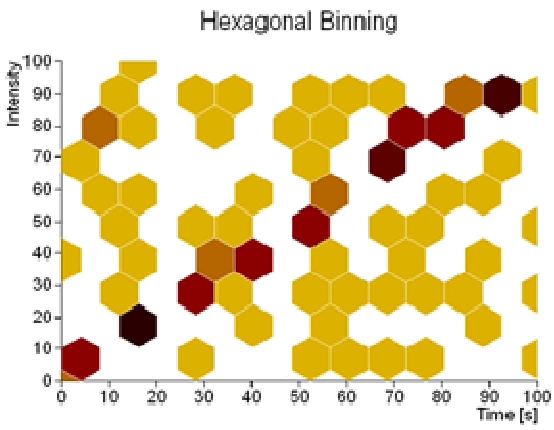

Figure 3. Hexagonal binning

4 System Architecture

In our system for facial pattern recognition, we used Microsoft Kinect for Windows sensor V2, which simplifies the facial feature extraction. Kinect for Windows v2 sensor is a product from Microsoft. It is a device with depth sensing technology, a built-in color camera, an infrared (IR) emitter, and a microphone array, enabling it to sense the location and movements of people. Microsoft provides, with this new Kinect sensor, a development kit (SDK 2.0) with new facilities, drivers, tools, APIs and device interface. As compared to its predecessor, it features enhanced color depth, fidelity, video definition, depth perception, skeletal tracking and improved range of high quality operation (.5 meters near, 4.5 meters far). The sensor additionally includes full-HD video and wider field of view, improved skeletal tracking and new active infra-reduction for better tracking in low light. The sensor is also known to connect to a Windows 10 computer via a power supply and computer interface hub that features a USB3.0 port. The Microsoft Kinect Studio [Microsoft, 17], together with the Kinect for Windows Software Development Kit (Kinect for Windows SDK), enables us to create applications that can track human faces in real time. The Microsoft.Kinect.Face API provides capability for tracking facial feature locations, tracking facial animations, and capturing 3D representations of faces in real time.

Figure 4. System interface

With this technology, we built an application that is able to track up to a max of 1347 3d points of the face that we can map to the 2D image. In addition, the application can evaluate still images and video with different techniques for facial pattern depending on the input requisites, and also export and filter 3D data. User annotation capability is also possible to improve evaluation systems. Figures 4 and 5 illustrate its interface and the system capturing process.

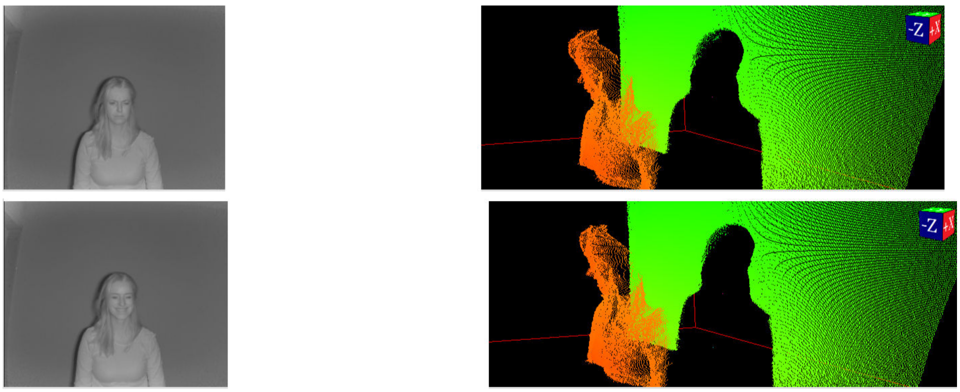

Figure 5. System capturing process

The export data option is able to save to disk the 3D points being filtered and showed in the image in figure 6. The application has also an algorithm that calculates in real time and then exports the values used to evaluate facial expressions and create the charts.

Figure 6. 3D point visualization and depth map with 3D eyebrows and nose points mapped into 6 issues per year.

5 Data Visualization and Analysis

Data capture from the system is basically consisted of 3D spatial points, besides the timestamp. To elaborate this study it was considered a frame rate of 16 fps (frames per second) and only the control points from the subject eyebrows were considered (13 points in each frame). Only one subject face expression is considered at this stage. As a starting point, we considered four facial expressions, to be analysed and evaluated graphically:

- Happiness;
- Angriness;
- Surprise;
- Neutral – base facial expression.

Our goal was to use data visualization to help us infer the potential existing differences between the subject's eyebrows movement, while expressing each of the expressions. Each expression was registered in a separate sequence and recorded file. To simplify, the third dimension was discarded, because the capture was done without any head movement.

The subject was asked to start always from a neutral expression, evolving gradually to another expression. When the expression was completed, the subject was asked to hold it, so that we guarantee, that the end of the record was the peak of the expression. Each file is more or less 5 seconds long.

The data was then normalized to a 500 per 500 pixels window and any anomalous value deleted. It was visually mapped according the following techniques: scatter plot, hexagonal binning and Voroni treemap.

This choice was made because at this stage we are interested in finding out how the spatial position of control points change along the time, detecting their direction, in what regions of the space, they have a higher rate of incidence, besides how close or far they happen. The scatter plot can show clearly both space position and direction. The colour saturation identifies the point timestamp – lighter to darker blue (starting and ending control points).

The hexagonal binning of the points can help us to identify what spatial areas are more weighted than others, having higher rates of incidence. Darker blue bins means high occurrence regions.

Finally, we also wanted to know how near or far point's locations were happening. The Voroni treemap also was an asset to achieve this. More regular or irregular polygons colour distribution indicate the steadiness or not of the points position and thus distances. The following figures /7, 8 and 9) show the visual outputs in each case. In these plots, we can see the right and left eyebrows control points. These plots were constructed using RAWGraphs [15].

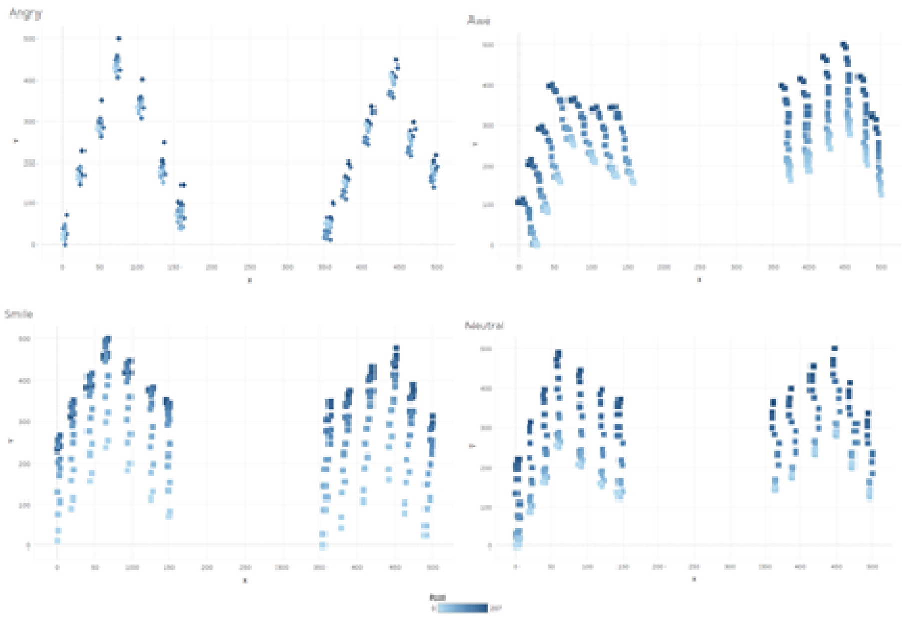

Figure 7. Scatter plots of the control points

Looking at the scatter plots sequence (figure 7) of the four facial expressions, we can notice that there are different directions in the movement of the eyebrows control points. Although all of them seems follow an upward direction, they are not the same, having a change also in their speed. The happiness and neutral facial expressions are very similar, but still, are not the equal. While the first has a steady pace of change, almost to the end, the second one, has always a steady pace of change. The angriness expression conveys to overlapping control points, meaning not a really upward direction, but a downward one.

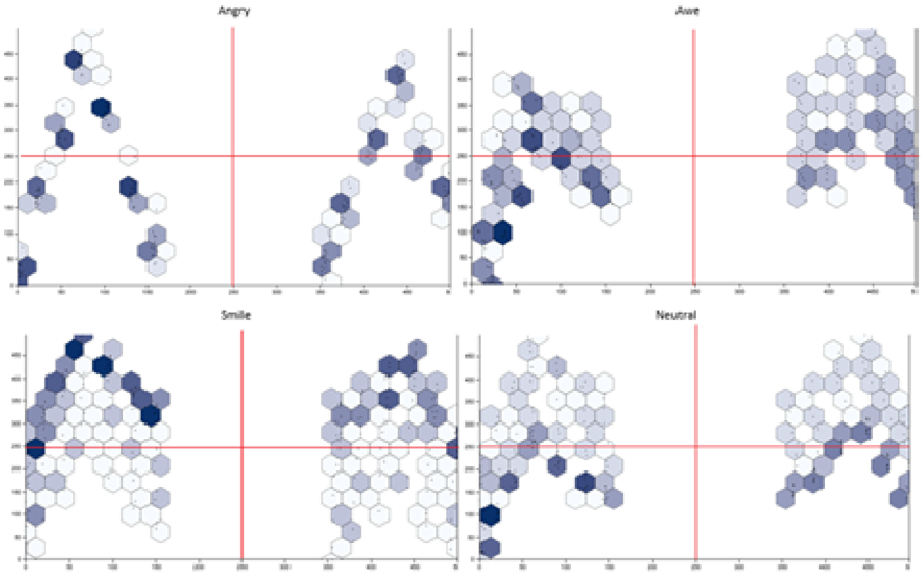

Figure 8. Hexagonal binning of the control points

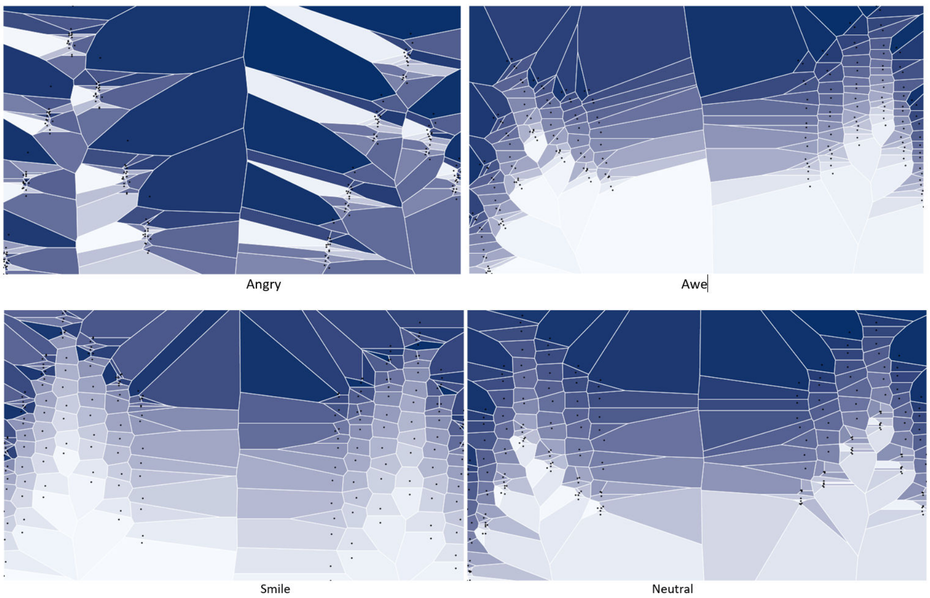

Figure 9. Voroni treemaps of the control points

The hexagonal binning plots (figure 8) depict that the angry facial expression is fully concentrated (darker blue) at a certain spatial position. If we divide the area in four quadrants, the points are concentrated in all of them. The happy expression occupied mostly the two upper quadrants of the plot, while the neutral, the two lower. The awe expression although occupying the four quadrants, does not reveal the same level of concentration as the angry one.

The Voroni treemaps (figure 9) show different spatial distribution between eyebrows control points according the face expression. The angry expression has a lower spatial distribution while the happy and neutral ones have more regular spatial distribution. This means that points are much nearer in the case of the angry expression, presenting lower distances than in other cases.

Expression	**Scatter plot**	**Hexagonal bin**	**Voroni treemap**
Angry	downward	high density in every quadrant, very	very low spatial distribution and

	very fast	concentrated	irregular
Awe	up and right very slowly	high density in every quadrant but more sparse	irregular spatial distribution
Smile	upward slowly	high density in upper quadrants	very regular spatial distribution
Neutral	upward very slowly	high density in lower quadrants	regular spatial distribution

Table 1: Summary of the findings pointed out by the plots of the data.

6 Conclusions and Future Work

Because we considered in this study only one subject, we cannot assume that these findings are actually patterns in human facial expressions. It is only a starting point to elaborate some initial conclusions that will be used to compare against other subject's results as a reference. Table 2 summarizes the comparisons between the findings for each facial expression.

As we can see, according the facial expression, the behaviour of the control points differ. This can give us some starting guidance about what attributes are important to look for in the control points position and possibly, monitor: speed, direction, density and spatial distribution/location.

Future work includes the inclusion of other subjects and facial expressions to the study. We want to compare the differences or similarities between facial expressions, and thus, finding possible existing patterns, general and transversal to everyone. Besides these plots, we will use other visualization techniques. Data will be processed to find out other relevant visual attributes that can help us detect existing patterns.

Expression	**Awe**	**Smile**	**Neutral**
Angry	opposite direction; very different speed; different density and spatial distribution	opposite direction; different speed; high density in opposite locations; very different spatial distribution	opposite direction; very different speed; different density and spatial distribution
Awe		slight the same direction; almost same speed; different quadrants; very different spatial distribution	slight the same direction; the same speed; different quadrants; different spatial distribution
Smile			same direction; different speed; almost the same density but in different quadrants; almost same spatial regularity

Table 2: Comparison between facial expressions behaviours

References

[1] Craig, Scotty D., et al. "Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive–affective states during learning." Cognition and Emotion 22.5 (2008): 777-788.

[2] Dinges, David F., et al. "Optical computer recognition of facial expressions associated with stress induced by performance demands." Aviation, space, and environmental medicine 76.6 (2005): B172-B182.

[3] Hjelmås, Erik, and Boon Kee Low. "Face detection: A survey." Computer vision and image understanding 83.3 (2001): 236-274.

[4] Zhang, Cha, and Zhengyou Zhang. "A survey of recent advances in face detection." (2010).

[5] Microsoft Kinect Studio, https://msdn.microsoft.com/en-us/library/dn785306.aspx (visited on 26/4/2017)

[6] Yang, Ming-Hsuan, David J. Kriegman, and Narendra Ahuja. "Detecting faces in images: A survey." IEEE Transactions on pattern analysis and machine intelligence 24.1 (2002): 34-58.

[7] Toyama, Kentaro. "Look, ma-no hands! hands-free cursor control with real-time 3d face tracking." PUI98 (1998).

[8] Gorodnichy, Dmitry O. "On importance of nose for face tracking." Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on. IEEE, 2002.

[9] Munzner, Tamara. Visualization analysis and design. CRC Press, 2014.

[10] Liu, Shixia, et al. "A survey on information visualization: recent advances and challenges." The Visual Computer 30.12 (2014): 1373-1393.

[11] Dimara, Evanthia, Anastasia Bezerianos, and Pierre Dragicevic. "The attraction effect in information visualization." IEEE Transactions on Visualization and Computer Graphics 23.1 (2017): 471-480.

[12] Kosara, Robert. "Presentation-oriented visualization techniques." IEEE computer graphics and applications 36.1 (2016): 80-85.

[13] Gotz, David. "Dynamic Voronoi Treemaps: a visualization technique for time-varying hierarchical data." Tech. Rep. RC25132 (2011).

[14] Hu, Haibo, and Xiaofan Wang. "How people make friends in social networking sites—A microscopic perspective." Physica A: Statistical Mechanics and its Applications 391.4 (2012): 1877-1886.

[15] RAWGraphs, http://rawgraphs.io/ (visited on 26/4/2017)