Application for the Analysis of Portrait Composition

María Luisa Ripoll Domínguez

Bachelor's thesis

Application for the Analysis of Portrait Composition

María Luisa Ripoll Domínguez

20. May 2019

Institute for Data Processing
Technische Universität München

María Luisa Ripoll Domínguez. *Application for the Analysis of Portrait Composition.* Bachelor's thesis, Technische Universität München, Munich, Germany, 2019.

Supervised by Prof. Dr.-Ing. K. Diepold and Stefan Röhrl; submitted on 20. May 2019 to the Department of Electrical Engineering and Information Technology of the Technische Universität München.

Institute for Data Processing, Technische Universität München, 80290 München, Germany, http://www.ldv.ei.tum.de/.

Abstract

The visual media and new technologies of today have lead to a photography boom. More and more people are interested in photography and wish to learn about the theory behind it. With this in mind, an application that can help beginners understand the photography rules and how to use them has been developed. This application focuses on portraiture and composition, and it is capable of analyzing an input image based on the *Rule of Thirds* and the *Rule of the Eye-Gaze*. The application then outputs a grade and proposes recommendations on how to improve the photograph.

Contents

1. Introduction

1.1. The Context

The number of photographs taken worldwide every year has been growing exponentially for the last decades, reaching a staggering number of 1.2 trillion pictures in 2017 alone [14]. The particularly notable photography boom of the last decade is – to a greater extent – due to the appearance of the smart-phone. In fact, 85 percent of all 2017 photographs were taken with a mobile phone camera [16]. The overall distribution of pictures taken in 2017 can be seen in Figure 1.1.

Cell-phones are devices that we take everywhere with us and because of that, photography has become available to everyone. This, together with the rising presence of social media has changed our attitude towards pictures and now more and more people have begun to give importance to their quality. However, the many intricacies of the world of photography are not commonly known by the average individual and thus, most people are not aware of how to improve their shots. Furthermore, **what is it that makes a photograph "good" or "bad"?** Is it not an abstract and subjective opinion?

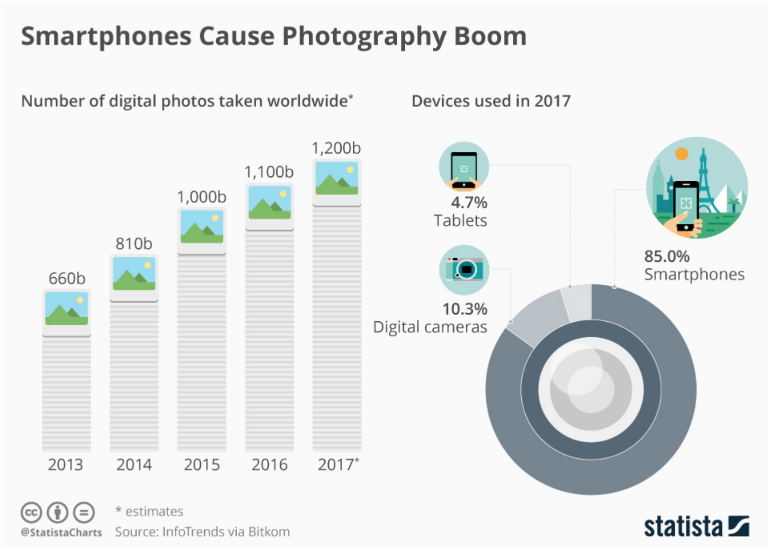


Figure 1.1.: Digital photos taken worldwide per year: *Statista - August 2017*

Partly, yes. But, as a form of art, photography is all about transmitting emotions to the viewer. Photography theory, therefore, boils down to the usage of specific and defined rules that play with the human's psychology [15] [8, pp. 20-22]. These rules can affect the shots in different ways: from creating points of interest and enhancing the sense of reality all the way to inducing an increased feeling of comfort in the viewer. However, applying the rules of photography to a shot does not necessarily mean that the outcome will be “good”, it will just be better than what it could have been [1, p10].

All in all, to improve their photographs, people should apply the photography rules. Yet, to do so they would need to study and learn about them first as well as take them all into account at the same time. This task would be highly simplified if there was an application that could serve as a guideline for all the different rules and that could compute which ones were being followed and which ones weren't – ideally also giving recommendations and explanations to the different rules.

The rules of photography theory depend on the scene and the situation being depicted. Due to the fact that there are countless possible subjects and circumstances for a photograph, it would be very complicated for an application to know which rules to apply in order to accurately analyze the picture. Because of this, the application should focus on only one type of photography, and since the most used form of photography is *Portraiture*, it is also the best choice for the application.

It is very common to find photo editing programs with tools that change the color temperature and contrast of an image. Yet when it comes to composition, these applications usually only offer a cropping tool. Because of the neglected importance of structure and the lack of features that analyze it, the application should focus on the photography rules regarding composition.

1.2. State of the Art

The idea of creating an application to improve photographs in their post-production is not a new one. Image manipulation and retouching is almost as old as photography itself [9, p1], and digital photo editing is likewise as old as personal computers. Raster graphics editors such as Adobe Photoshop have allowed the general public to retouch photographs since the 1980's. Today there are countless mobile and desktop applications with very advanced tools: from Instagram's simple color filters and Snapchat's "fun" features (like face detection and swapping), all the way to advanced desktop programs meant for professional photographers, these applications have given us the ability to do anything we wish with our shots.

However, the fact that everybody has the possibility to edit photographs does not mean that they know how to. Many applications such as Photography 101, Photo Academy or PhotoCaddy offer tips and video tutorials where they teach basic photography theory. Still, I have yet to find an application that directly analyzes the use of the photography rules for a given photograph. Such an application would have educational purposes, would help develop aesthetic criteria and could be used as a complement for already existing photo editing software.

1.3. The Aim

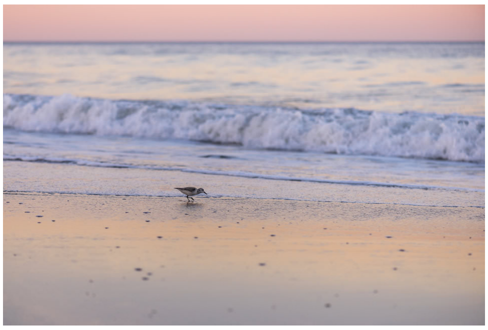
The aim of this thesis is to develop an application that analyzes portraits and examines whether or not they are using two of the most basic rules of composition. Once it has done so, it will give recommendations on what areas to improve on and it will give an "estimated grade". This estimation will not refer to how good or how bad the portrait is but to what extent the photographer was capable of applying the photography rules. The rules that the application inspects are the Rule of Negative Space, that checks on the Eye-Gaze, and the Rule of Thirds that uses Points of Power to drive the attention of the viewer.

2. Theory

2.1. Basic Concepts

The Composition of an image is the photographic feature that has most control over the attention of the viewer [8, pp. 20–22]. To create points of interest and direct the focus, the photographer can manipulate the *negative space* and the *breathing room* until the desired balance is reached.

The *positive space* of a photograph is its main focus. It usually refers to the object with most significant visual weight, but it can also be a striking point that attracts attention. The area surrounding the positive space is the *negative space* - which in most cases corresponds to the background. Photographs where the majority of the space is positive give a feeling of saturation, intensity and frenzy while a photograph with high portions of negative space gives a sense of calm, isolation and clarity [17, pp. 19, 27, 41–43]. Figure 2.1 exemplifies the use of negative and positive space in a composition. The pictures transmit opposite emotions: emptiness vs. saturation.



(a)

(b)

Figure 2.1.: (a) Negative Space Composition [6] (b) Positive Space Composition [7]

The *breathing room* is the negative space surrounding the head of the person in the portrait. If the subject is looking in a particular direction other than towards the camera then the breathing room refers only to the space in that direction [5].

2.2. The Rule of Negative Space

Photography theory has a long list of rules and exceptions when it comes to negative space. Yet when dealing with portraits there is a basic and very important matter to take into account: the *eye-gaze*.

Because of the human psychology, when we see another person looking at something, we want to know what they are looking at, and subconsciously our attention is drawn in the same direction as the subject's glance. This spatial direction is known as a *visual line* [2].

Usually a photographer will use it and add negative space in this area to produce a stable composition. Sometimes, however, the visual line is cut off and the gaze direction is interrupted. This provokes *visual tension*: a small cognitive feeling of unease and instability in the viewer [12]. Visual tension is used in a variety of cases in cinematography to enhance a feeling of stress or fear, and though there are a few ways in which visual tension can be used to produce a great image, the space following the eye-gaze is rarely cut off in an effective way.

The best approach to a strong composition is to add more negative space in the direction of the eye gaze. It is also important to not disregard the space between the figure and the frame [13, chapter 18, p238]. If the subject is not looking towards the center, then the composition should follow one of the rules of object placement such as the *Rule of Thirds*.

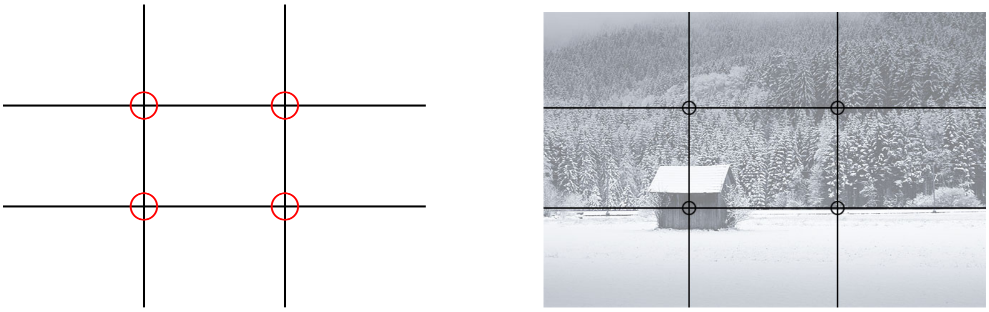
Figure 2.2 shows a portrait of a woman looking left. Yet the negative space between the left frame and the face is too small. The edited version of the original portrait adds breathing room and reduces the visual tension. This is a clear example of how negative space should be implemented in a portrait.

Figure 2.2.: Portrait by british fotographer Julia M. Cameron. Edited by T.L. Glover. [10]

2.3. The Rule of Thirds

The first thing that people are taught about photographic composition is the Rule of Thirds. A basic and useful guideline that helps people that are new to photography with their compositions, namely to avoid putting the subjects always in the center. Though central compositions can be used successfully like for example when using symmetry, they tend to be static and attract less attention from the viewer than other formations [3, chapter 3]. The *Rule of Thirds* responds to the way that humans perceive visual information and it is a standard that guaranties a good and pleasant structure [1].

It must be noted that even though it offers an easy way of creating a good picture it is not exclusive. There are many ways to compose the elements in an image and make a great shot. However, because of its simplicity and the fact that it is almost always valid, the Rule of Thirds is the composition standard that will be used and checked by the program. To apply this rule, one must first imagine 4 lines crossing the image and creating a grid of nine equal cells as shown in (a) of Figure 2.3



(a) **(b)**

Figure 2.3.: (a) The points of power. (b) Example of a Rule of Thirds composition. [11]

The intersection points between the lines are called the *points of power* (or *power points*) and they are the strongest *focal areas* of the photograph. By focal area we are referring to the point or line that the viewers focus most of their attention on when they first see the image[1]. The lines forming the grid are the second strongest focal area. Knowing this, it makes sense to place the main subject of the photograph along one of these lines. This will make the photograph more interesting. For this composition it is also important to place the horizon or any distinguished horizontal line in the background on one of the two thirds and not in the middle as one can observe in (b) of Figure 2.3 It is also helpful to put the image's key points on one of the focal points, be it an eye in the case of a close up or a head in a full shot.

[1]There are many ways to refer to the described "focal area". It is most commonly called a "point of interest" or a "point of emphasis". However, because points of interest can also be lines and areas it is easy to get confused with the terminology. Thus the term "focal area" will be used instead.

3. Method of Implementation

To implement the rule of negative space, the program should be able to recognize the facial features of the person being portrayed – mainly the outline of their head and the direction of their gaze. For this we use facial landmarks, which are facial “key-points” that serve to determine the positions of the nose, eyes, eyebrows, mouth and jawline. These key-points are located within the face region using a shape predictor.

The first step in facial landmark detection is to recognize the face or faces present in the image and set a bounding box around them. These bounding boxes reduce the area to be analyzed by the shape predictor and therefore lower the chance for errors. For this first step, one can choose from a variety of face recognition algorithms and classifiers such as Haar Cascades, Histogram of Oriented Gradients (HOG) or Neural Networks. In this case, we used OpenCV’s Python API[1] and the Dlib[2] machine learning software library. The latter uses a pre-trained HOG together with a Linear SVM(Support-Vector Machine) for object detection.

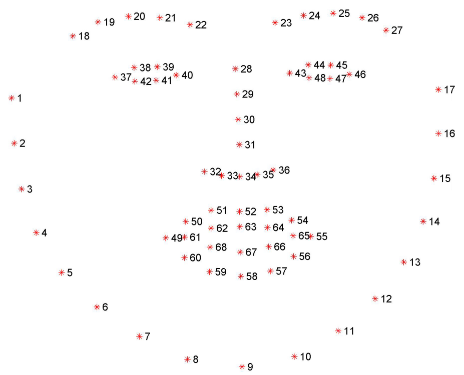


Figure 3.1.: Facial Landmark Indices - trained on the iBUG 300-W dataset [4].

With Dlib’s function "get-frontal-face-detector" the region enclosing the face is determined and then one can proceed to use the shape predictor and extract the landmarks. Dlib’s pre-trained facial landmark detector locates a total of 68 pixel coordinates (x, y) corresponding to the landmarks shown in Figure 3.1[3].

[1] Opencv Python API: https://pypi.org/project/opencv-python/

[2] Dlib Library: http://dlib.net/ , Dlib’s Python API: http://dlib.net/python/index.html

[3] It must be noted that the shape predictor labels the first point 0 and not 1 therefore all the points in this image are actually higher by one than what they should be.

3.1. Determining the Eye-Gaze

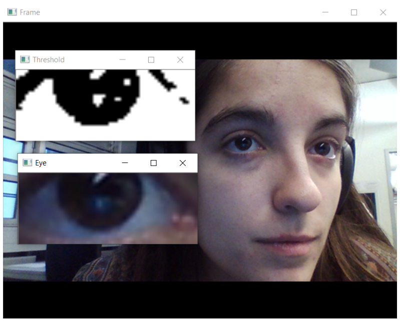


Figure 3.2.: Close-Up of the Left Eye and its binary counterpart

Once the landmarks have been located the program calculates the positions of the eyes of the subject. The width and height between the landmarks that conform the eye-shape are used to create a cut out image of each eye. This image is then transformed to gray-scale to reduce it from a three dimensional color set of data to a one dimensional one. This way the image can be converted to binary using a threshold, as can be observed from Figure 3.2.

The program then calculates the position of the pupil within the eye by analyzing the contours found in the binary image. Of course, what the value of the threshold should depend on the portrait's illumination and on how clear the difference is between the sclera and the iris. Therefore, the code to extract the pupil's position is looped with an increasing threshold until it has a contour that is considered to have an adequate size and circularity. This discards eyelashes and other possible sources of mis-classification. The circularity of the contour is measured with the isoperimetric quotient[4]

$$Q = \frac{4\pi A}{P^2},$$

which compares the area A of a given contour to the area that a perfect circle with the same perimeter P would have. When the contour of the pupil has been detected, the program calculates its center using image moments. (See Appendix A.1.1) Knowing the position of the pupils center, the program can now compare it to the position of the center of the eye. The latter is calculated using eye coordinates. The eye axes are created with the landmarks. For example: for the left eye, the outer side points (36 and 39) form the x axis, while the mean position of the upper points (37 and 38) and the mean of the lower points (40 and 41) create the y axis.

[4] Isoperimetric Inequality on a Plane: https://en.wikipedia.org/wiki/Isoperimetric_inequality

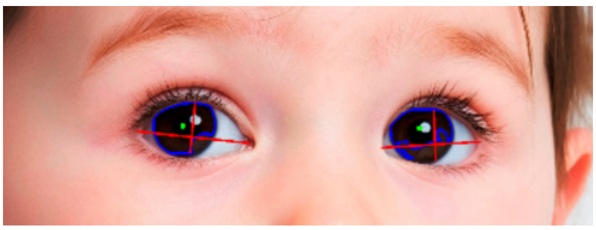


Figure 3.3.: Eye-Axes and calculated pupil center

For each eye, a resulting vector between the origin of the eye coordinates (the center of the eye) and the center of the pupil is calculated. The mean of both these vectors give the final vector for the eye-gaze. This vector is then classified into one of 9 categories shown in Figure 3.4: Up (1), up-right (2), right (3), right-down (4), down (5), down-left (6), left (7), left-up (8) and center (9). The program also calculates the angle of the eye-gaze vector from the negative x axis in a clockwise direction. Figure 3.4 shows the delimiting angles for each area.

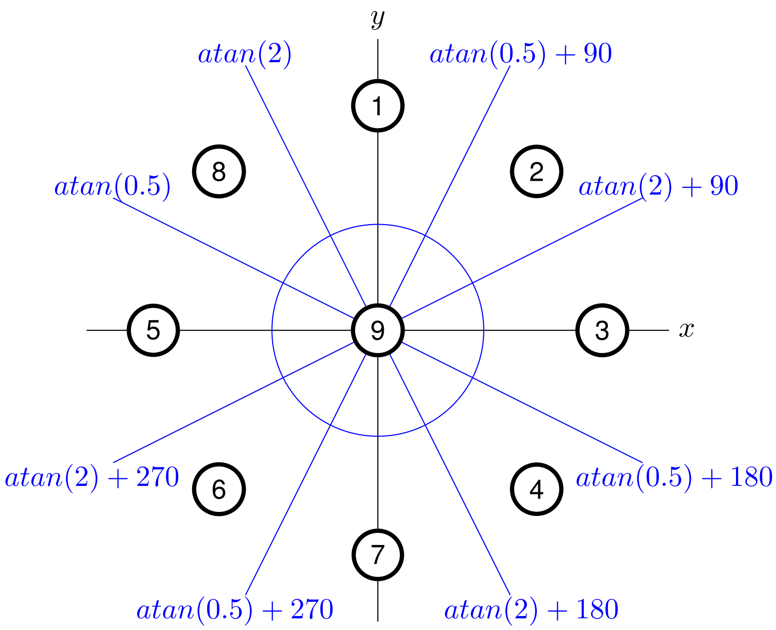


Figure 3.4.: The 9 eye-gazing directions

The 9 categories correspond to the 9 areas in the x- y- plane, which are formed by a circle and four dividing lines: $y = \frac{1}{2}x$, $y = 2x$, $y = -\frac{1}{2}x$, $y = -2x$. The area in which the center of the pupil is positioned will be its class of gaze. The radius of the circle delimits the classification for subjects looking towards the camera. The choice of this radius will be discussed in the area of results.

3.2. Checking the Rule of Thirds

As explained in the theory chapter, the rule of thirds determines the placement of the photograph's main subject. And because we are dealing with portraits, the main subject is always going to be a person. This means that the major issue of object detection is already solved: the algorithm uses Dlib's library to find the face of the person in the portrait - just as it does for the previous composition rule. With this, many of our tasks will be simplified.

To determine how well the program follows the rule of thirds and to specify a grade for it, the following factors will be checked:

- The average eye-height
- The positioning of key facial features on the image's points of power
- The placing of the face

Each of these factors outputs an independent "grade" for how well they are implemented. The final grade for the Rule of Thirds is reached combining the factor grades. The following subsections will deal with these factors and their respective grading systems independently.

3.2.1. The Average Eye-Height

The eyes are the features of a face that we focus most of our attention on, and thus the placement of the eyes in a portrait is key. A basic and commonly used rule is to place the eyes along the upper third of the image [1]. This is one of the factors that the program checks on.

The average height for the eyes of the subject was calculated using the y-coordinates of the centers of both eyes. These centers could be easily computed using the same functions that were used to determine the pupil and eye-gaze.

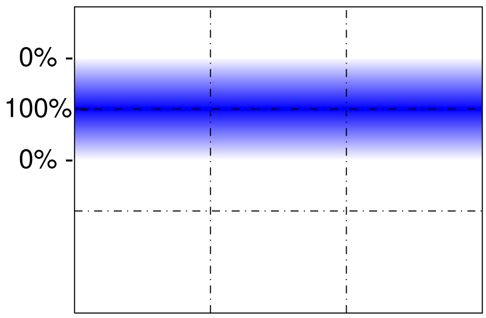


Figure 3.5.: The Eye Height Grading System. The stronger the color, the higher the grade.

The mean of these 2 y-values was then compared to the height of the upper third of the image. To allow space for a margin of error, the grade for this factor was not binary (on the third/ not on the third) but was given as a percentage were 100% referred to a perfect placement of the eyes on the third all the way to 0% for images where the distance between the eye height and the upper third was bigger or equal to a sixth of the total image height, as shown in Figure 3.5.

3.2.2. The Positioning of Features on Points of Power

As explained in the theory chapter for the Rule of Thirds, the points of power are the strongest focal areas of the photograph. They draw the attention of the viewer, and so do the eyes. Therefore, placing an eye on a point of power creates a strong composition. The program checks on the positioning of at least one eye with respect to a point of power.

Due to the rule of the eye height, only the two upper points of power are taken into account for the eye factor. To determine the specific grade, the program calculates the distances between the center of each eye and the two relevant points of power. The closest distance is then turned into a percentage using an arbitrary radius as observed in Figure 3.6. This radius acts as a margin of error and it is equal to the ninth of the longest dimension of the image - width or height. As the distance increases, the grade decreases until a 0% on the radius and beyond it.

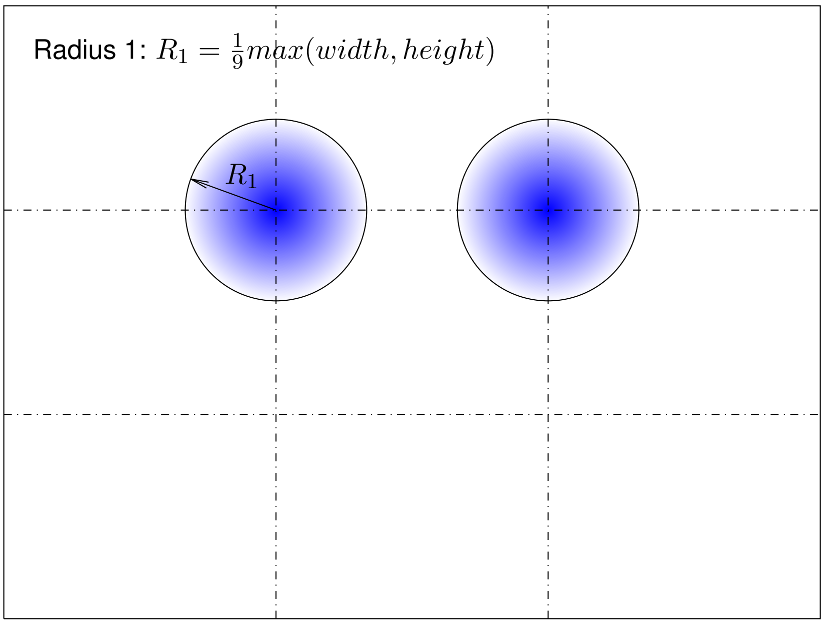


Figure 3.6.: The Points of Power Grading system for Eyes

When the subject is facing in a direction other than center, there exists the possibility to place key facial features along the middle section of the vertical thirds. As explained in the theory chapter, these lines are the second focal area of the image and placing important parts of the composition on them will draw attention from the viewer. This is why the program will check if the tip of the nose can be found within this focal area. Figure 3.7 illustrates the grading system for this extra feature, where the vertical lines that join the power points are graded 100% and the shape surrounding this line at a distance of R_2 is graded 0%.

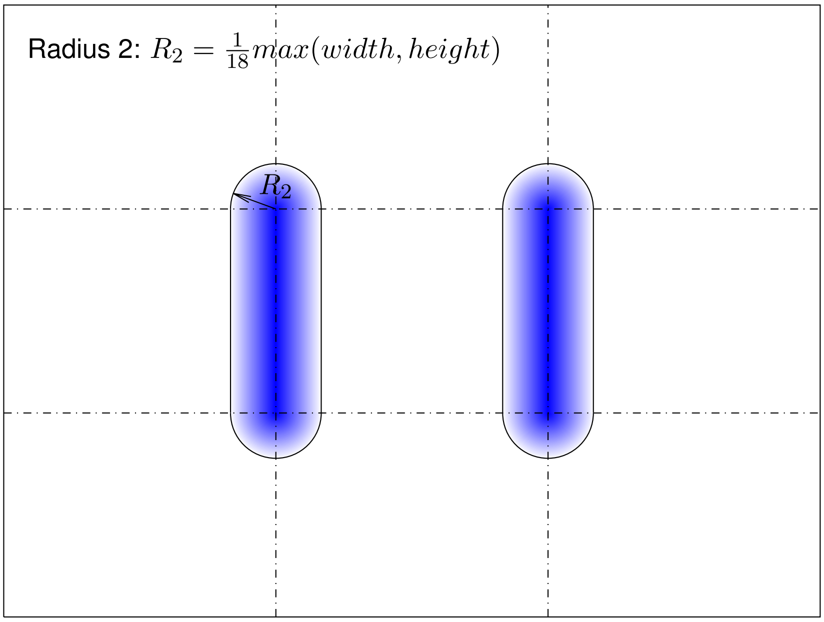


Figure 3.7.: Other Facial Features Grading system

The final grade for the positioning of features on the points of power is computed using Algorithm 1.

Algorithm 1 Grade Computation for "Features on Points of Power"

Known Variables: left and right eye grades, feature grade and facial direction.
Goal: to compute the Grade for this Factor.

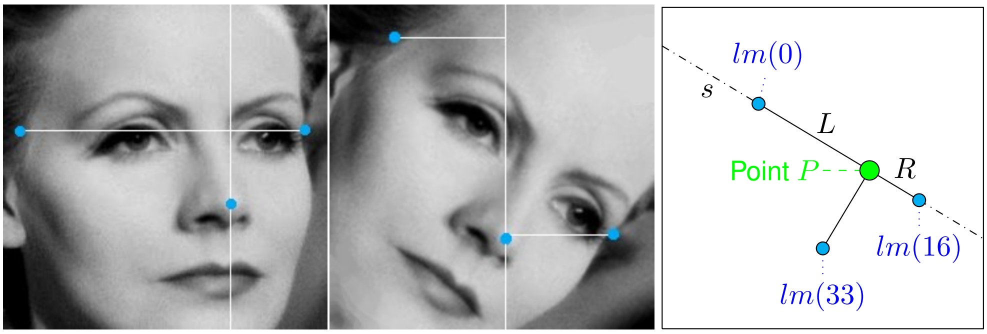
k = max(grade of left eye, grade of right eye)
if subject is facing toward the center **then**
    Grade = k
else
    Grade = (feature grade + k)/2
end if

However, to be able to run this Algorithm, one needs the facial direction of the subject, which to avoid confusion, has nothing to do with the eye-gaze direction.

Determining the facial direction

One can see from Algorithm 1 that the only real thing that needs to be determined regarding the facial position is whether the subject is facing the center or not. This means that we can make the program recognize just that and avoid complications.

A very simple way to tackle this problem would be to compare the x-coordinates of the landmarks 0 and 16 (the sides of the face) with the x-coordinate of landmark 33 (the tip of the nose). If the x-difference between the landmark of the left side of the face and the tip of the nose is considerably bigger to the x-difference between the landmark of the right side of the face and the nose tip, then we can deduce that the person is looking toward the right, and vice versa as shown in (a) of Figure 3.8. Unfortunately, when the face is tilted at certain angles, this procedure may not work. In (b) of Figure 3.8 the subject is facing right but the difference in x between landmarks 0 and 33 and between 33 and 16 is more or less the same, which would give an incorrect assessment of facial direction: center.



(a) **(b)** **(c)**

Figure 3.8.: (a) and (b): Edited images of Greta Garbo in *Queen Christina* 1933, (c): Example sketch of distances L and R

Thus to make sure that the tilting of the face does not cause misclassifications, the distances that are compared to each other are the distance L from the left point (landmark 0: lm(0)) to point P and the distance R from the right point (lm(16)) to point P, where P is the projection of the tip of the nose (lm(33)) on the line s that crosses landmarks (0) and (16). These measures are depicted in (c) of Figure 3.8.

Finally, the threshold radio between L and R, was estimated to be 1:2 or bigger for subjects facing left and 2:1 for subjects facing right. All other photographs were classified as facing forward.

3.2.3. The Placement of the Face

Due to the Rule of the Eye Gaze, a portrait where the subject is looking in a specific direction, should have a wider negative space on that side of the frame. The Rule of Thirds complements the Rule of the Eye Gaze by placing the subject on one of the two vertical thirds ensuring a distinction between the size of the spaces on either side of the person in the image. As stated in the theory chapter this is not the only way to compose a portrait but it is an effective structure that always works. This is why, for subjects looking in directions other than center, the grade for the placement of the face will use the vertical thirds as reference. The grading for this case is shown in Figure 3.9.

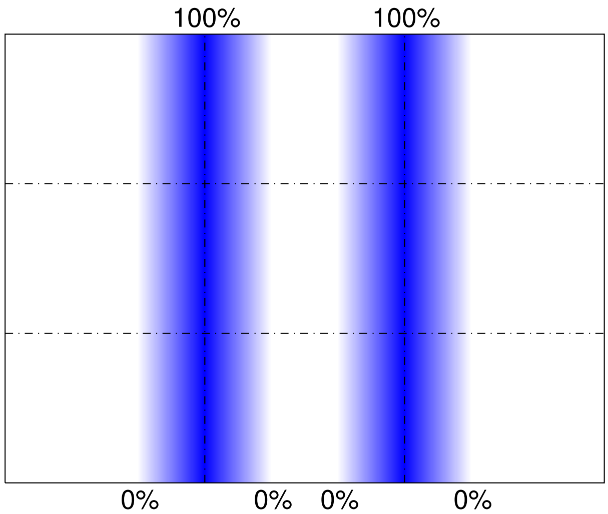


Figure 3.9.: Face Placement Grading System (eye gaze: not center)

For subjects that are not looking towards the side, the placing of the face is almost solely up to the photographer. The subject can be positioned at the center, on one of the vertical thirds, to the side or even so far to the side that the image only shows part of the face. Figure 3.10 shows 4 movie posters that exemplify different possible portrait compositions.

(a) (b) (c) (d)

Figure 3.10.: (a) *Fight Club* 1999; (b) *The Godfather* 1972; (c) *The Greatest Showman* 2017; (d) *The Curious Case of Benjamin Button* 2008.

Because of this, the grade for this factor when the subject is looking to the center (or up or down), will be 100%. There is, however, one exception. When the photographer is aiming for a central composition and slightly decenters the subject, it results in an unbalanced structure that is not pleasant for the eye [2, p15][8, pp. 20-22]. Figure 3.11 shows the grading system for Face Placement with a subject looking in the direction of the camera. The positions halfway through the middle line and the vertical thirds are the only parts of the image graded below 100%.

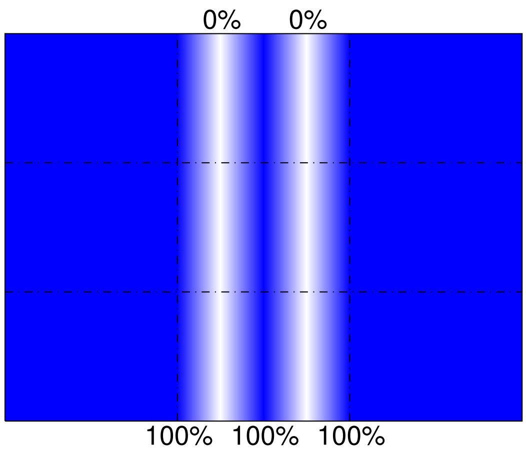


Figure 3.11.: Face Placement Grading System (eye gaze: center)

The Grade for the Rule of Thirds

The final grade for the Rule of Thirds is computed using the grades of the 3 established factors. However, if the portrait being taken is a *close-up*[5] or an image where the face occupies almost the whole frame, then the factor of the placement of the face is clearly not taken into account.

Algorithm 2 Grade Computation for "Rule of Thirds"

Known Variables: 3 factor grades, percentage area of image occupied by face
Goal: to compute the Grade for the Rule of Thirds.

k = percentage area occupied by face
if k < threshold **then**
    Grade = sum(3 factor grades)/3
else
    Grade = (Eye-Height Grade + Features on power points)/2
end if

---

[5] A close-up is a type of shot where the person or object being depicted is tightly framed.

3.3. Application Comments

The structure of the comments for the Rule of Negative Space and Eye-Gaze is always the same. First the Rule itself is explained and then one of two possible assessments is given (along with its grade) depending on whether the rule was followed or not.

As for the Rule of Thirds, one can observe its structure from Figure 3.12.

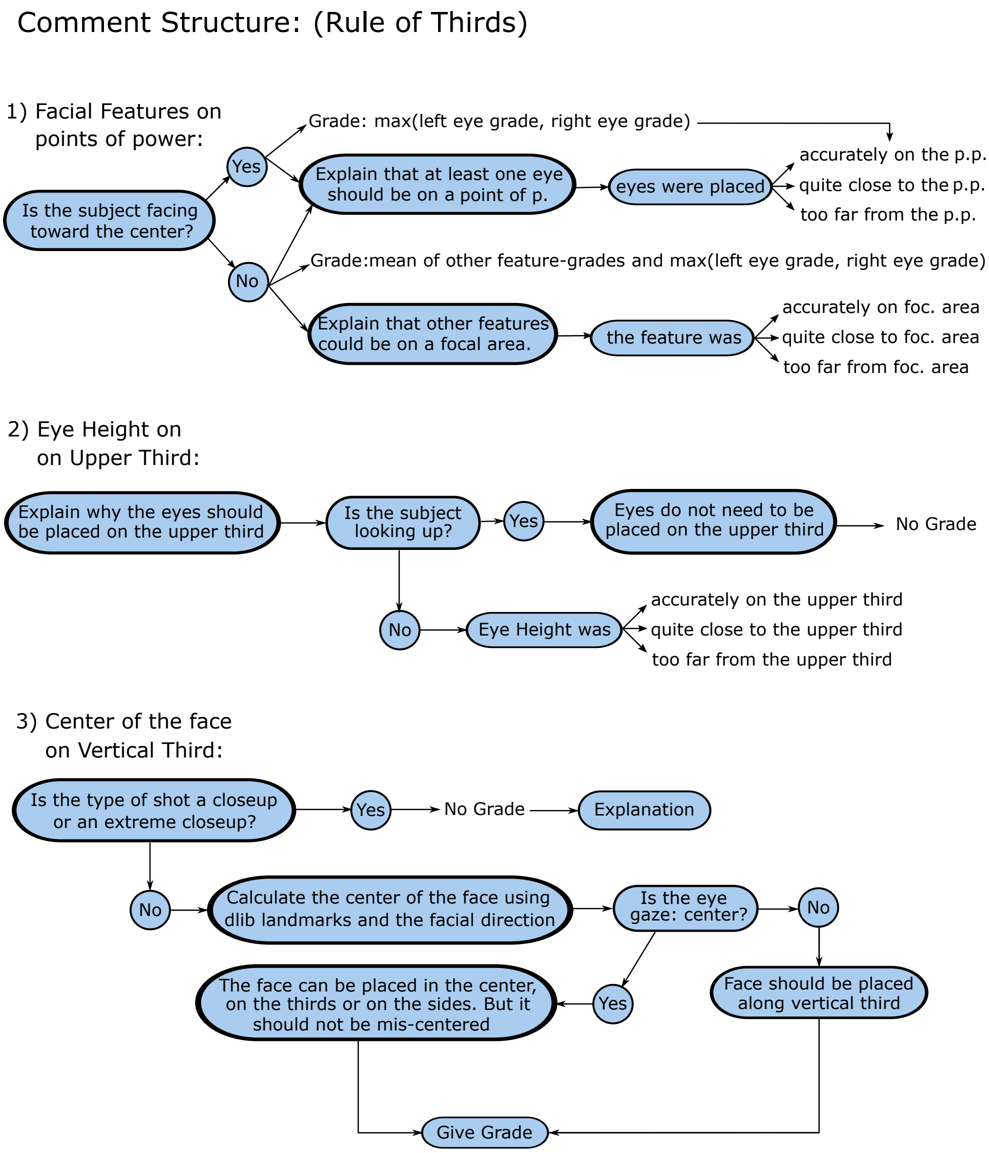


Figure 3.12.: The comment structure for the Rule of Thirds.

4. Results

4.1. Eye-Gaze Algorithm Evaluation and Results

To check the accuracy of the program in calculating the eye gaze direction, a test was carried out. The aim of the test was to run the program with images of different subjects looking in 4 different directions and at 19 different face angles.

4.1.1. Setup

The test equipment, which is displayed in Figure 4.1, consisted of the following:

- A remotely controlled motorized swivel chair. The chair had a an adjustable metallic holder that served as a guideline for the head.
- A computer with a terminal
- A photography lighting kit of 2 light-panels
- And a computer camera.

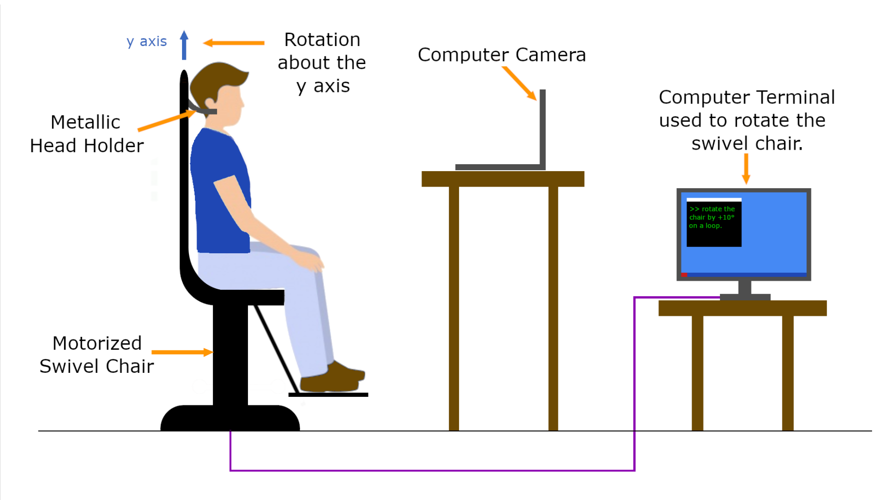


Figure 4.1.: Test setup.

The participants of the test were sat on the swivel chair and told to keep their head fixed on the holder and to face forward. A camera was placed in front of the subject and the zero angle was set to be the position facing towards the camera. The -90 degrees facial position was right and the +90 degrees left.

Through the terminal a command was given to rotate the chair to a starting angle of -90 degrees and loop a rotation of +10 a total of 18 times (until the angle +90 was reached). Between the 10 degree rotations there was a wait of 5 seconds to allow for a photograph to be taken. All the while, having the lighting panels illuminating the subject evenly to have an equal light factor from all angles.

The test was run 4 times per subject. In each run they were told to look left, right, up and down.

4.1.2. Classification and Calculation Errors

General Error

Once the images were taken they were labeled and run through the program, which gave a percentage error based on the output eye-gaze vector. Figure 4.2 shows how an eye labeled as looking left is evaluated. If the angle α between the negative x eye-axis and the eye-gaze vector is in the dotted area delimited by $atan(0.5)$ and $atan(-0.5)$ then the error is equivalent to zero. From the angle $atan(0.5)$ (with error = 0) to the angle 180 (with error = 100) the error grows constantly. The formula to calculate the error for an eye labeled as looking left is therefore

$$f_{Left}(\alpha) = \begin{cases} 0 & \text{for } \{\alpha \leq atan(0.5)\} \\ \frac{\alpha - atan(0.5)}{180 - atan(0.5)} * 100 & \text{for } \{atan(0.5) < \alpha \leq 180\} \\ \frac{atan(2) - 270 - \alpha}{180 - atan(0.5)} * 100 & \text{for } \{180 < \alpha < atan(2) + 270\} \\ 0 & \text{for } \{\alpha \geq atan(2) + 270\} \end{cases} .$$

The angle α is always calculated clockwise from the negative x axis as we can observe from Figure 4.2. However, the error computation is different for each labeled direction. The respective formulas for right, up and down gaze are

$$f_{Right}(\alpha) = \begin{cases} 0 & \text{for } \{atan(2) + 90 \leq \alpha \leq atan(0.5) + 180\} \\ \frac{atan(2) + 90 - \alpha}{180 - atan(0.5)} * 100 & \text{for } \{\alpha < atan(2) + 90\} \\ \frac{\alpha - atan(0.5) - 180}{180 - atan(0.5)} * 100 & \text{for } \{\alpha > atan(0.5) + 180\} \end{cases} ,$$

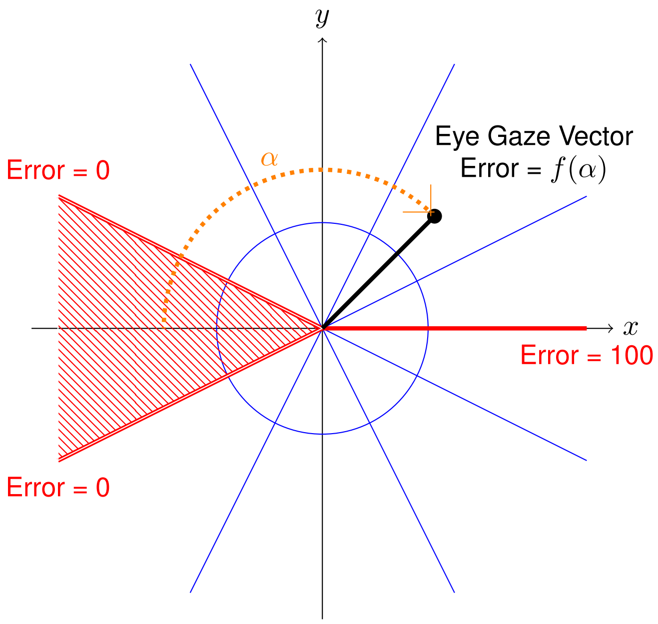


Figure 4.2.: Error evaluation for a photograph with a subject looking left. In the case scenario that the program misclassifies the eye gaze, it first calculates the angle alpha and then uses the error formulas to compute the final error.

$$f_{Up}(\alpha) = \begin{cases} 0 & \text{for } \{atan(2) \leq \alpha \leq atan(0.5) + 90\} \\ \frac{\alpha - atan(0.5) - 90}{180 - atan(0.5)} * 100 & \text{for } \{atan(0.5) + 90 < \alpha \leq 270\} \\ \frac{atan(2) - \alpha}{180 - atan(0.5)} * 100 & \text{for } \{\alpha < atan(2)\} \\ \frac{\alpha - atan(0.5) - 180}{180 - atan(0.5)} * 100 & \text{for } \{\alpha > 270\} \end{cases},$$

$$f_D(\alpha) = \begin{cases} 0 & \text{for } \{atan(2) + 180 \leq \alpha \leq atan(0.5) + 270\} \\ \frac{atan(2) + 180 - \alpha}{180 - atan(0.5)} * 100 & \text{for } \{90 < alpha < atan(2) + 180\} \\ \frac{\alpha - atan(0.5) - 270}{180 - atan(0.5)} * 100 & \text{for } \{\alpha \leq 90\} \\ \frac{90 - atan(0.5) - \alpha}{180 - atan(0.5)} * 100 & \text{for } \{\alpha > atan(0.5) + 270\} \end{cases}.$$

Center Classification and Error

These formulas were used for all except center misclassifications. The categorization problem of the center area does not have to do with the angle of the eye-gaze vector but with the distance of the center of the eye to the center of the pupil. The way to reduce the misclassifications of eyes looking towards the center, is to vary the radius of the circle delimiting area 9 of Figure 3.4. Many different radius values were tested with a looped code and the one that resulted with the least amount of errors was a radius of 30 percent of the mean between the size of the two eye-axes.

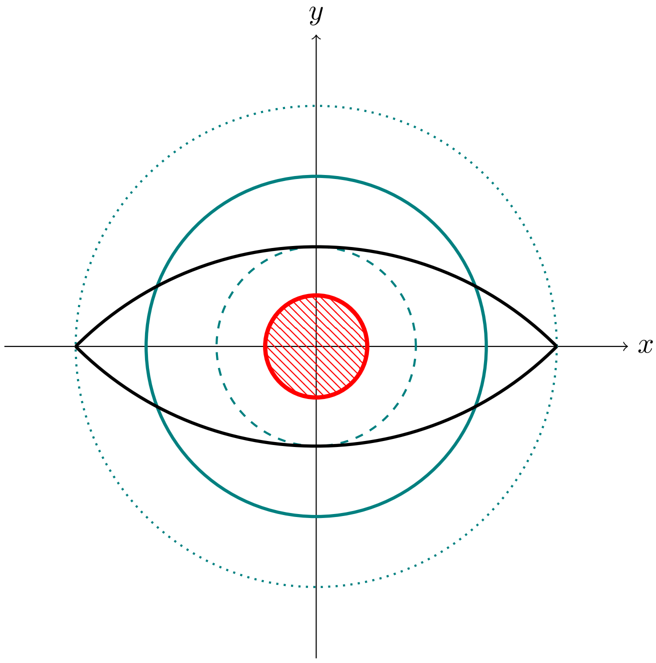


Figure 4.3.: The center classification area is shown in red. It is 30% of the mean between the sizes of the eye axes x and y. Here this mean is represented by the thick blue circle.

In Figure 4.3, the dotted and dashed circles have a diameter equivalent to each of the x- and y- eye-axes respectively. The mean of both their diameters is shown by the thick blue circle. Finally, 30 percent of the latter's radius conforms the area for center classification which is shown in red. The program will classify the eye-gaze as "center" as long as the pupil's centroid resides within this delimited area. To avoid confusion: this does not mean that parts of the pupil can't be found out of it.

As mentioned before, the formulas $f_{Left,Right,Up,D}(\alpha)$ can't be used to compute the error percentage for center misclassification. This is calculated using the distance d from the pupil center to the perimeter of area 9, and the distance a between the perimeter and the external point of the x eye-axis, which is considered to be the furthest possible distance. Then the value is finally multiplied by 100 to get the percentage center error

$$E_{Center} = \frac{d}{a}100.$$

Figure 4.4 shows an example for distances d and a where the center of the pupil is marked as a blue circle. In most cases the distance between the center and each of the outer eye-axes points is not equivalent. Because of this, the distances are normalized before computing the center error.

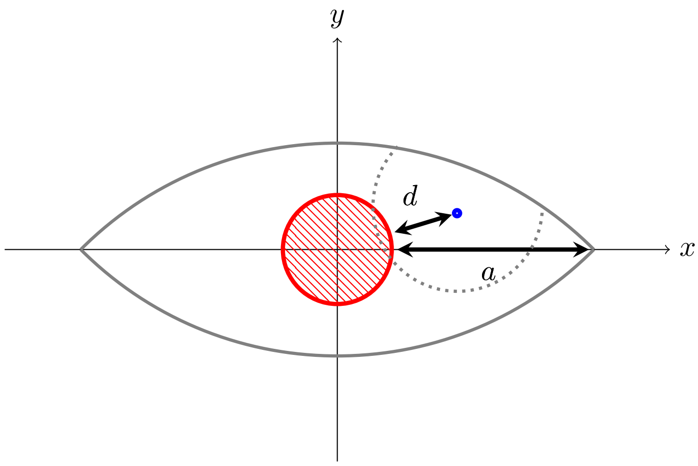


Figure 4.4.: Example of a center misclassification and its respective distances a and d.

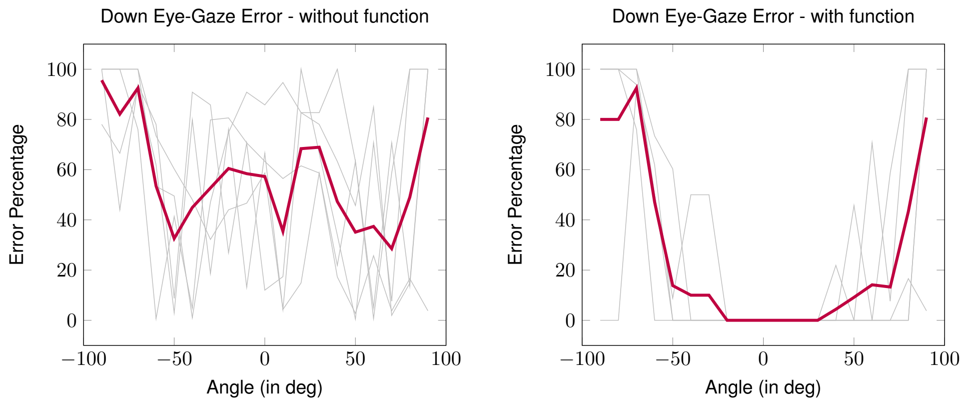
Down Classification

It became clear very early on that the classification of eye-gaze "down" was giving problems. The reason for this is that unlike with directions left, right and up, when a person looks down, the center of its pupil is not necessarily below the center of the eye. In fact, the upper eyelid tends to lower itself, lowering the eye-coordinate center and in many cases resulting in a wrong "center" or even "up" eye-gaze classification. Furthermore, when looking down, the area of the eye shown in the image is significantly reduced and it tends to have more shadows within it - both of these things are sources of misclassification. To try to fix this issue a function was added to the program. This function checked the distance between the upper y eye-axis point and the x eye-axis. If this was below a certain threshold the eye-gaze would be automatically set to be down.

(a) (b)

Figure 4.5.: In this example the highest points of both y eye-axes are very close to the x eye-axes. Thus, the force-down-gaze function will set the eye-gaze to: "down".

In (a) of Figure 4.5 we observe a female subject looking down with a facial angle of 0 degrees. The image next to it (b) is an approximate sketch of the eye-axes points. The upper point on the y eye-axis is almost at the same height as the x eye-axis. Because of this the program automatically sets the eye-gaze as "down". Figure 4.6 shows that the program's percentage error and misclassification when dealing with images labeled as down improved significantly while using this function.



(a) (b)

Figure 4.6.: (a) Without the force-down-gaze function. (b) With force-down-gaze function.

Only 2 of the 250 valid images that were not labeled as "down" were affected by the force-down-gaze function. Giving it a false positive rate[1] of 0.008 and a specificity[2] of 0.992.

Face Recognition Failure

The Dlib facial landmarks and the shape predictor used were meant to be used for images with faces that were looking - in some degree - towards the camera. Because of this the face recognition failed when the divergence angle from the frontal position was too high. For the cases were no face was recognized by the program, the error was automatically set to be 100.

4.1.3. Results

The Results of the test showed that the percentage error of the classification of eye-gaze increased with the angle divergence. The closer the face was to the frontal position of angle zero, the smaller the chance for errors. The x axes of the diagrams

[1] https://en.wikipedia.org/wiki/False_positive_rate
[2] https://en.wikipedia.org/wiki/Sensitivity_and_specificity

in Figure 4.7 represent the angle divergence of the face normal to the camera, while the y axes are the error percentage calculated using the methods shown in section 4.1.2. (e) of Figure 4.7 shows the final error curve.

Sets 1 to 5 show the results for each independent run of the test when the subjects were looking upwards. The mean of the sets shows that the higher the angle divergence, the higher the chance for a wrong eye-gaze estimation. It is significantly more accurate than both left and right graph curves.

The results shown in these figures are based on the images taken from 5 subjects (one woman and 4 men). The test was run 4 times per subject - one for each eye-gaze direction that was to be tested and since each test run produced 19 images, the test set contained 380 photographs. However, the code failed to recognize a face in 51 of them (13.4%) for angle divergences between 70 and 90 degrees. Because of the fact that the program being tested is not responsible for the face recognition part of the procedure, these images were considered invalid when calculating the accuracy and features of the eye-gaze classification algorithm. The tables below show the accuracy and mean error of the program for each direction when using an angle divergence of $\pm 90^o$ and $\pm 50^o$.

Table 4.1.: Results for an angle divergence of $\pm 90^o$

Direction	Left	Right	Up	Down	Total
Correct Est.	30	37	63	61	191
Valid images	85	85	80	79	329
N^o of images	95	95	95	95	380
Accuracy	0.35	0.44	0.79	0.77	0.58
Mean Error	0.150	0.172	0.169	0.262	0.188

Table 4.2.: Results for an angle divergence of $\pm 50^o$

Direction	Left	Right	Up	Down	Total
Correct Est.	26	38	45	49	158
Valid images	55	55	55	55	220
N^o of images	55	55	55	55	220
Accuracy	0.47	0.69	0.82	0.89	0.72
Mean Error	0.039	0.044	0.012	0.043	0.034

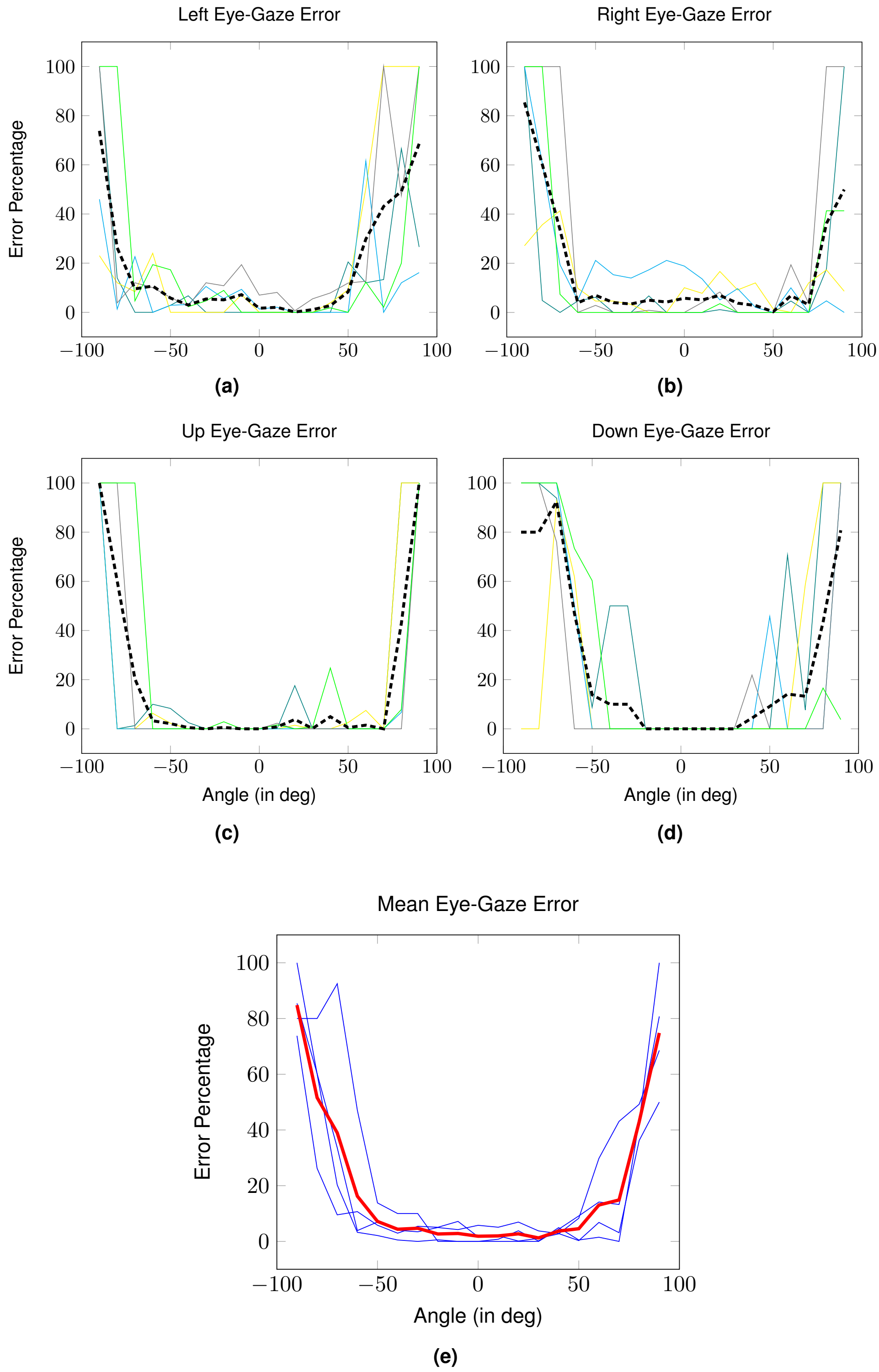


Figure 4.7.: (a),(b),(c) and (d) are the results for each gazing direction, and (e) is the average of these means.

4.2. Rule of Thirds Evaluation

The code for the rule of thirds uses the Dlib landmarks and simple mathematical formulas to compute a grade. All sources of misclasification errors in this case can, therefore, only be due to the accuracy of the Dlib functions and not the code of this program. What needs to be tested is how well the mathematical representation of the rule of thirds that has been developed in this thesis works. This is measured together with the application itself in the following section.

4.3. Application Evaluation and Results

Since the application only analyses composition rules, the photograph grading system should be solely used to compare images with the same depiction and lighting. To find the accuracy of the application, a test was carried out and 20 portrait photographs were preselected for this purpose. Each of these images was then cropped in seven different ways, creating independent sets of images with the same depiction and different compositions (140 captures).

The images were then run through the program. The photographs within each set were given a score of one to seven that reflected the order of grades with respect to each other. Separately, a photography expert[3] analyzed the sets of images and gave them a score in the same manner (1-7). The scores given by the expert were considered to be the ground truth. The table with the difference in results between the images analyzed by the machine and by the photography expert can be found in Appendix A.2. On average, the difference between the program's output scores and the ones given by the expert was 1,15.

Given the distribution, the maximum average difference possible between the scores was of 3,43. Making this equivalent to 100% error would make the programs average error equivalent to 33%. Once the scores had been compared, the images with the highest difference between the programs output and the expert's output, which were the ones where the program performed worse, were added to a list. Then, the photography expert was asked to give explanations for the choice of scores in the images on the list. The evaluation of the rule of thirds and the eye gaze were correct in 70% of the cases and coincided with the comments offered by the program. However, the negative score was due to the fact that in these images there had been an important factor that had not been taken into account. Another rule of photography deals with the cropping of the subjects, their faces and their body-parts. The incorrect cropping of the images on the list was the reason behind their negative scores. Unfortunately, because the program did not implement this rule, it could not use it.

[3]Teacher of photography in the professional training center for audiovisual media and new technologies: *CIFP José Luis Garci* in Madrid

To have an idea of how a further implementation of this rule could improve the program, the rule was manually applied on all 140 images. If the cropping on a photograph was incorrect, the score i that the application had output for that particular version of the image, would be subtracted by two. Then, the images with scores $i-1$ and $i-2$ of the same set would be increased by one. With this procedure, the average difference between the application's scores and the scores of the expert was reduced to 0,91 or 26%

4.4. Conclusion

The availability of cameras and the ease to take pictures have powered a newfound interest in photography. People want to improve their photographs but they do not know how. To help beginners understand how to implement basic photography theory into their portraits and to achieve higher aesthetics regarding composition, an application was to be developed that could give recommendations and explanations of photography rules.

Combining OpenCV, the python API and the Dlib library, it was possible to create a function that could determine the eye-gaze of the portrait's subject: a fundamental feature in the analysis of a photograph and especially in the usage of the *Rule of the Eye-Gaze*. The tests for this function showed that it has an accuracy of 0.72 for faces on a frontal pose or with a divergence of up to 50^{o}. The facial landmarks from Dlib could be used further for the analysis of the *Rule of Thirds*, which deals with the placement of the subject within the frame. This rule could output a grade from 0 to 100 based on the analysis of three factors: the height of the eyes, the proximity of key facial features to the *points of power* and the positioning of the center of the face. Finally the application was made. Using a python API one can run any photograph through the program. The application will output the same image with added guiding lines to help the visualization of the rules, together with a set of explanations and recommendations for the given portrait. The error of the application was analyzed by comparing the photograph scores that the program produced as output with the scores given by a photography expert. The difference between the two sets of scores was on average: 33%. The application outputs that diverged most from the photography experts scores were found to have in common the wrong implementation of a photography rule that was not implemented within the program.

It is worth noting that photography is a very broad and multifaceted form of art with countless rules and even more exceptions. This application focuses on one type of photography - namely portraiture, and only on the composition thereof. Even with just a focused direction such as this one, there were several rules that could not be implemented. Regardless of this, the resulting application is perfectly valid to help beginners to understand the implementation of the eye-gaze, eye-height, points of power and the Rule of Thirds.

Bibliography

1. P. Andrews and M. Langford. *Langford's starting photography: the guide to great images with digital or film*. Routledge, 2007.
2. R. Arnheim. *Art and visual perception: A psychology of the creative eye*. Univ of California Press, 1965.
3. J. Aumont. *L'image*. Armand Colin, 2011.
4. S.Z. C. Sagonas, G. Tzimiropoulos and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. December 2013.
5. J. Campbell and P. Pillai. Leveraging limited autonomous mobility to frame attractive group photos. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 3396–3401. IEEE, 2005.
6. S. Cox. Negative space composition. In *Photography Life*, 2018. https://photographylife.com/elements-of-composition-photography, last accessed on 12/03/19.
7. S. Cox. Positive space composition. In *Photography Life*, 2018. https://photographylife.com/landscapes/positive-and-negative-space-in-photography, last accessed on 12/03/19.
8. D.A. Dondis. *A primer of visual literacy*. MIT Press, 1974.
9. H. Farid. *Photo Forensics*. MIT Press, 2016.
10. T.L. Glover. Mr. robot creates visual tension with composition techniques. In *iPOX Studios and Canon of Design*, 2016. https://ipoxstudios.com/mr-robot-creates-visual-tension-with-composition-techniques/, last accessed on 12/03/19.
11. M. Gommel. Rule of thirds. In *Photography Mad*, 2018. https://www.photographymad.com/pages/view/rule-of-thirds, last accessed on 12/03/19.
12. R. Heinstein. *Visual Tension in Graphic Design*. Master's thesis, Western Kentucky University, 2014.
13. S. Johnson. *On digital photography*. O'Reilly, 2006.

14. C. Linder. Building science narratives using still photography. In *AGU (American Geophysical Union) Fall Meeting Abstracts*. 2018.

15. E. Mavridaki and V. Mezaris. A comprehensive aesthetic quality assessment method for natural images using basic rules of photography. In *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 887–891. IEEE, 2015.

16. F. Richter. Smartphones cause photography boom. In *Statista: The Statistics Portal*, 2017. `https://www.statista.com/chart/10913/number-of-photos-taken-worldwide/`, last accessed on 04/03/19.

17. A.W. White. *The elements of graphic design: space, unity, page architecture, and type*. Skyhorse Publishing, Inc., 2011.

Appendices

A. Appendix

A.1. Mathematical Concepts

A.1.1. Image Moments

An image moment, also known as a raw moment, is a weighted average for the intensities of the pixels of an image. The formula to calculate the raw moment of a gray scale image where the number of x and y variables is discrete (number of pixels) is

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y).$$

$I(x, y)$ refers to a function of intensity that serves to weigh the pixels. In the case of a gray-scale image, $I(x, y)$ will take the value within 0-1 that is relative to the 0-255 value of the pixel. However if the image is a binary one, $I(x, y)$ can only be 1 or 0. Now suppose we have a binary image of 9 pixels as shown below, where the ones indicate a black pixel of intensity one and the zeros indicate a white pixel of intensity zero:

	x = 1	x = 0	x =-1
y = 1	1	1	1
y = 0	1	1	1
y =-1	0	1	0

In this case the moment M_{00} is equivalent to the sum of all the black pixels in the image, M_{01} is the sum of all of the y values of the black pixels and M_{10} is the sum of all x values of the black pixels.

$$M_{00} = \sum_x \sum_y x^0 y^0 I(x, y) = \sum_x \sum_y I(x, y) = 7 * 1 + 2 * 0 = 7$$
$$M_{01} = \sum_x \sum_y x^0 y^1 I(x, y) = \sum_x \sum_y yI(x, y) = 3 * 1 + 3 * 0 + (-1) = 2$$
$$M_{10} = \sum_x \sum_y x^1 y^0 I(x, y) = \sum_x \sum_y xI(x, y) = 2 * 1 + 3 * 0 + 2 * (-1) = 0$$

Image Moments can be used to extract different image properties. The one used in this thesis is the centroid of the image. The centroid of the binary 9 pixel image example is calculated below.

$$Centroid : \{\bar{x},\ \bar{y}\} = \left\{\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right\} = \left\{\frac{0}{7}, \frac{2}{7}\right\} = \{0, 0.286\}$$

A.2. Table Results

Table A.1.: Left 1 - 5 are the error percentages for each angle divergence when the 5 subjects were looking left. The error percentage average is shown under Total L.

Angle	Left 1	Left 2	Left 3	Left 4	Left 5	Total L
-90	100	23.05	45.99	100	100	73.81
-80	13.98	12.01	1.33	3.85	100	26.23
-70	0	8.41	22.65	12.01	4.64	9.54
-60	0	24.03	0	10.05	19.39	10.69
-50	2.85	0	2.87	6.04	17.31	5.81
-40	6.72	0	3.17	2.62	2.28	2.96
-30	0	0	10.52	12.01	4.64	5.43
-20	0	0	5.64	10.81	8.91	5.07
-10	0	7.17	9.35	19.39	0	7.18
0	0	0	1.33	7	0	1.67
10	0	0	2.07	8.1	0	2.03
20	0	0	0	0.64	0	0.13
30	0	0	0	5.49	0	1.1
40	0	3.85	0	7.88	1.82	2.71
50	20.48	9.15	0	11.9	0	8.31
60	12.01	50.49	61.52	12.65	12.3	29.79
70	13.33	100	0	100	2.07	43.08
80	66.54	100	12.01	47.51	20	49.21
90	26.59	100	16.15	100	100	68.55

Table A.2.: Right 1 - 5 are the error percentages for each angle divergence when the 5 subjects were looking right. The error percentage average is shown under Total R.

Angle	Right 1	Right 2	Right 3	Right 4	Right 5	Total R
-90	100	27.13	100	100	100	85.43
-80	4.86	35.64	60.73	100	100	60.25
-70	0	41.34	19.39	100	7.37	33.62
-60	4.64	9.82	5	0	0	3.89
-50	6.17	4.64	21.16	2.87	0	6.97
-40	0	4.64	15.34	0	0	4
-30	0	3.29	13.98	0	0	3.45
-20	6.72	0	17.31	0.87	0	4.98
-10	0	0	21.16	0	0	4.23
0	0	10.05	18.85	0	0	5.78
10	0	7.78	13.64	4.02	0	5.09
20	1.13	16.66	5.09	8.29	3.55	6.94
30	0	9.15	9.68	0	0	3.77
40	0	12.01	2.07	0	0	2.82
50	0	1.46	0	0	0	0.29
60	4.64	0	10	19.39	0	6.81
70	0	12.01	0	3.65	0	3.13
80	17.65	17.31	4.7	100	41.34	36.2
90	100	8.63	0	100	41.34	49.99

Table A.3.: Down 1 - 5 are the error percentages for each angle divergence when the 5 subjects were looking down and the force-down-gaze function was used. The error percentage average is shown under Total D.

Angle	Down 1	Down 2	Down 3	Down 4	Down 5	Total D
-90	0	100	100	100	80	95.608
-80	0	100	100	100	80	82.11
-70	92.63	100	75.97	100	92.486	92.486
-60	61.94	50	0	73.42	47.072	53.482
-50	0	0	0	60.28	13.814	32.576
-40	0	0	0	0	10	44.916
-30	0	0	0	0	10	52.692
-20	0	0	0	0	0	60.458
-10	0	0	0	0	0	58.4
0	0	0	0	0	0	57.26
10	0	0	0	0	0	35.44
20	0	0	0	0	0	68.356
30	0	0	0	0	0	68.918
40	0	0	21.88	0	4.376	47.29
50	0	45.72	0	0	9.144	35.108
60	0	0	0	0	14.134	37.392
70	58.66	0	0	0	13.26	28.54
80	100	0	0	16.6	43.32	48.992
90	100	100	100	3.78	80.756	80.756

Table A.4.: Down 1 - 5 are the error percentages for each angle divergence when the 5 subjects were looking down and the force-down-gaze function was not used. The error percentage average is shown under Total D.

Angle	Down 1	Down 2	Down 3	Down 4	Down 5	Total D
-90	100	78.04	100	100	100	95.608
-80	44.01	66.54	100	100	100	82.11
-70	93.83	92.63	100	75.97	100	92.486
-60	78.04	61.94	53.36	0.65	73.42	53.482
-50	8.79	2.96	49.51	41.34	60.28	32.576
-40	90.85	79.82	1.02	4.83	48.06	44.916
-30	85.79	18.61	47.04	79.82	32.2	52.692
-20	26.9	75.97	74.8	80.61	44.01	60.458
-10	70.67	13.17	90.85	70.67	46.64	58.4
0	12.01	66.54	85.79	63.3	58.66	57.26
10	17.39	4.32	94.7	56.46	4.33	35.44
20	100	82.69	82.69	61.52	14.88	68.356
30	66.54	82.69	78.04	58.66	58.66	68.918
40	33.97	100	63.3	21.88	17.3	47.29
50	0.46	63.3	45.72	63.3	2.76	35.108
60	70.67	0.7	84.88	4.98	25.73	37.392
70	7.64	58.66	1.97	70.67	3.76	28.54
80	100	100	14.78	13.58	16.6	48.992
90	100	100	100	100	3.78	80.756

Table A.5.: Up 1 - 5 are the error percentages for each angle divergence when the 5 subjects were looking up. The error percentage average is shown under Total U.

Angle	Up 1	Up 2	Up 3	Up 4	Up 5	Total U
-90	100	100	100	100	100	100
-80	0	0	0	100	100	40
-70	1.33	0	0	0	100	20.27
-60	10	6.31	0	0	0	3.26
-50	8.29	2.28	0	0	0	2.11
-40	2.54	0	0	0	0	0.51
-30	0	0	0	0	0	0
-20	0	0	0	0	2.87	0.57
-10	0	0	0	0	0	0
0	0	0	0	0	0	0
10	0	0	0	2.28	1.62	0.78
20	17.58	1.46	0	0	0	3.81
30	0	0	0	0	0.7	0.14
40	0	0	0	0	24.75	4.95
50	0	2.54	0	0	0	0.51
60	0	7.49	0	0	0	1.5
70	0	0	0	0	0	0
80	100	100	6.72	0	7.88	42.92
90	100	100	100	100	100	100

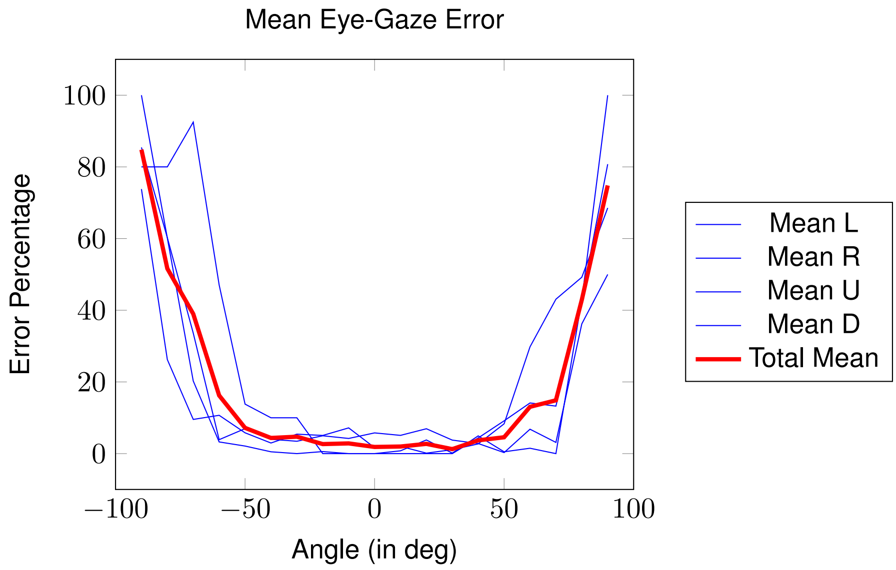


Figure A.1.: Mean Graph: The mean results for the test where subjects where looking left, right and up are shown in blue while the mean of all results is shown in red. This is equivalent to the general accuracy of the algorithm.

	V1	V2	V3	V4	V5	V6	V7	Mean per Img
Img 1	4	2	0	2	2	2	0	1,71
Img 2	4	1	2	3	2	2	1	2,14
Img 3	3	4	1	0	0	1	3	1,71
Img 4	2	nan	1	2	0	0	1	1,31
Img 5	2	2	1	3	0	4	1	1,86
Img 6	1	3	nan	1	5	0	0	1,67
Img 7	2	2	2	2	1	1	2	1,71
Img 8	0	0	0	3	0	1	2	0,86
Img 9	0	0	2	0	0	0	0	0,29
Img 10	0	0	nan	nan	nan	nan	nan	0,00
Img 11	2	2	3	0	3	0	1	1,57
Img 12	0	0	1	0	3	0	0	0,57
Img 13	2	0	3	2	4	2	nan	2,17
Img 14	0	1	0	0	0	1	0	0,29
Img 15	2	1	nan	0	0	0	0	0,50
Img 16	0	2	0	0	3	1	0	0,86
Img 17	0	2	4	0	0	2	0	1,14
Img 18	3	0	2	0	0	1	5	1,57
Img 19	0	0	0	1	0	0	1	0,29
Img 20	0	0	4	2	0	0	0	0,86
Total Mean								1,15

Figure A.2.: This table shows the difference between the scores given as output by the machine and the expert. V1 to V7 are the 7 different cropped versions of each image. The absolute error is 1,15 which gives a relative error of 33%

	V1	V2	V3	V4	V5	V6	V7	Mean per Img
Img 1	4	2	0	2	2	2	0	1,71
Img 2	3	0	2	3	0	1	1	1,43
Img 3	2	2	1	0	0	2	1	1,14
Img 4	2	nan	1	0	2	1	1	1,17
Img 5	2	1	1	2	2	2	0	1,43
Img 6	3	0	nan	1	3	0	2	1,50
Img 7	2	0	2	2	0	0	2	1,14
Img 8	0	0	0	3	0	1	2	0,86
Img 9	1	0	1	0	0	0	0	0,29
Img 10	0	0	nan	nan	nan	nan	nan	0,00
Img 11	1	2	3	0	1	0	0	1,00
Img 12	0	0	1	0	2	0	0	0,43
Img 13	0	0	1	3	2	1	nan	1,17
Img 14	0	0	0	0	0	0	0	0,00
Img 15	0	3	nan	0	2	0	0	0,83
Img 16	1	1	0	0	1	1	0	0,57
Img 17	0	2	4	1	0	1	0	1,14
Img 18	3	0	2	0	0	1	4	1,43
Img 19	0	0	0	1	0	0	1	0,29
Img 20	1	0	3	0	0	0	1	0,71
Total Mean								0,91

Figure A.3.: This table shows the difference between the scores given as output by the machine and the expert. V1 to V7 are the 7 different cropped versions of each image. In this table a new rule was manually implemented. The machine scores in photographs that did not implement this new rule correctly, received a "punishment" of -2. The new absolute error is 0.91 which gives a relative error of 26%