Human Attention

THE POSSIBILITY OF MEASURING HUMAN ATTENTION USING OPENCV AND THE VIOLA-JONES FACE DETECTION ALGORITHM

JONATAN CÖSTER & MICHAEL OHLSSON

Abstract

The question of whether an audience is focused and attentive can be of great importance. Research shows that a main concern during lectures is the varying level of attention from the students. Getting real time feedback on the students attention could give the lecturer an insight into what can be improved in terms of the material being presented. One potential way to get this feedback is to have a face detection algorithm to measure when someone is paying attention. The objective of the study is to investigate if it is possible to measure a person's attention in a controlled environment using the OpenCV programming library and the Viola-Jones algorithm. In order to measure if someone was paying attention, a definition of attention was required. It is obvious to humans when someone is paying attention. However, this is not the case for a computer. A data set consisting of pictures of attentive and inattentive subjects was used to evaluate whether the software could be used to measure attention. The results of the study showed that OpenCV had an almost perfect detection rate with few false positives. The conclusion is therefore that the OpenCV programming library could be used to measure attention in a controlled environment. However, due to the limited scope of the study, further investigations are required in order to use it in a real-world application.

Contents

Chapter 1

Introduction

1.1 Context

The question of whether an audience is focused and attentive can be of great importance, not only to the speaker addressing the audience. A company who wants to further educate their employees or an academic institution can spend large sums on hiring lecturers. This means there is a financial incentive for these organisation to find out whether the lecturer is competent and able to engage the audience.

Research shows that a main concern during lectures is the varying level of attention from the students. Some surveys suggest most students attention span is about 10 minutes [1]. Another difficulty is how to develop ways to keep the students focused for longer periods of time. Getting real time feedback on the students attention could give the lecturer an insight into what can be improved in terms of the material being presented, but also it could signal when it is appropriate to switch context or engage in other activities. One potential way to get this feedback is to have a face detection algorithm to measure when someone is paying attention.

1.2 Problem statement

The objective of the study is to investigate if it is possible to measure a person's attention in a controlled environment using OpenCV and the Viola-Jones algorithm. The reason for focusing on an individual's attention and using a controlled environment is that this is the smallest sub problem of the task of measuring attention in an audience.

In order for someone's attention to be measurable, there has to be a clear definition of what attention is. This definition is discussed in detail in the method section of the report. Also, when using face detection algorithms to measure attention, it might not be the case that the most accurate algorithm is best suited. Is the algorithm good enough to detect the faces of people paying attention in the first place? And, if so, is it also sufficiently inaccurate so that it does not detect the faces of people not paying attention? This may be an issue as the very nature of face detection algorithms is to detect faces as accurately as possible.

Chapter 2

Background

This chapter first gives an overview of the computer vision field (section 2.1.1) and its applications (section 2.1.2), followed by a brief history with emphasis on face detection and recognition (section 2.1.3). Then follows a more detailed description of the face detection field and its applications (section 2.2.1) together with the different approaches used (section 2.2.2) and the theory behind the Viola-Jones algorithm (section 2.2.3). The last part of the chapter presents some of the research done in the field of human attention (section 2.3) and studies made on students and learning (section 2.3.1).

2.1 Computer vision

2.1.1 Overview

Face detection is a field which has been studied since the 1970s. It is a subdomain of the field of computer vision which emerged in the 1960s. There are many fields closely related to computer vision such as image processing and computer graphics. While image processing produces images from images and computer graphics translates information into images, computer vision is mainly about recovering the three-dimensional structure of the world from images. In other words, computer vision is the opposite of computer graphics.

Because images are a two dimensional projection of the three-dimensional world, the information is not fully available and has to be recovered in some way [2]. This kind of "inverse" problem of recovering unknowns is part of what makes computer vision a complicated field. The three-dimensional structure of the world with its properties that surrounds us humans is something that we perceive without any great effort [3]. How we recognize objects and interpret shapes, illumination, color distribution, reflecting light comes naturally. However, the processes and the functionality behind the human visual system is complex and is still not completely understood, other than in great levels of detail and generality. Computer vision thus have a very difficult problem to solve; it has to digitally reinvent, with both hardware and software, the abilities of a long developed and evolved biological visual system with the help of physics, geometry, statistics and learning theory [4].

Figure 2.1 gives an example of the problem; two representations of an image,

the bottom one being a normal image and the top one is part of the matrix that holds the different gray scale intensity values of the image. No information is lost between the two, but the top representation only displays information with numbers instead of light and this is what the computer "sees". Also, this data of numbers has a large component of noise and distortion which comes from variations of the real world (lighting, reflections, weather), imperfections in the lens, motion blur, compression effects after image capture, etc [5]. This shows that a task that is undemanding for humans turns into a complicated puzzle for computers [2].
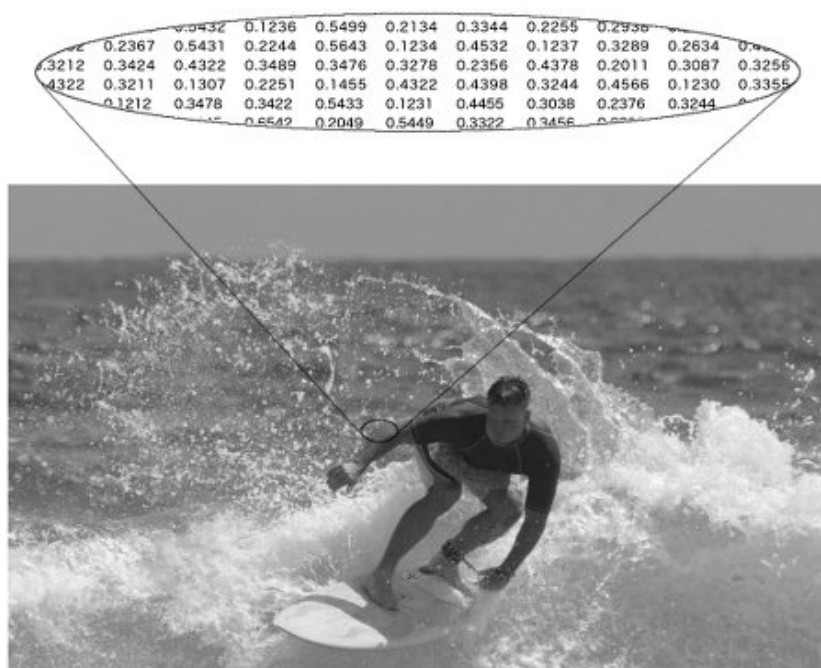

Figure 2.1: Two representations of the same picture.

2.1.2 Applications

Computer vision is a broad field and it is used in a wide variety of real world applications such as "optical character recognition," "fingerprint recognition" and "quality control". Optical character recognition, or OCR, is the process of converting printed text, possibly scanned by a scanner or digital camera, into a machine encoded text. It is a good example of one of the most well-used applications of computer vision; Project Gutenberg has over 48000 digitized books in its collection [6]. However, this is excelled by The Internet Archive which has over 2 million books [7] and Google Books which has more than 30 million books [8].

The use of computer vision in consumer products had a renaissance when Microsoft released the Kinect in 2010; according to Guinness World Records it was the fastest selling consumer electronics device and by 2013 more than 21

million units had been sold. It makes use of computer vision by using a system which can interpret gestures and track objects in three dimensions.

2.1.3   History

One of the first articles ever published on the subject of computer vision was "Machine Perception of Three-Dimensional Solids" [9] which is considered by some to have initiated the computer vision field. It was published by Lawrence G. Roberts at Massachusetts Institute of Technology in 1963, as his Ph.D. thesis. In his thesis he proposed a method to extract 3D geometrical information from 2D views of blocks (polyhedra).

In 1966 researchers became aware of the difficulties in the field of computer vision. Professor Marvin Minsky, a cofounder of what today is known as MIT Computer Science and Artificial Intelligence Laboratory, and his colleague Seymour Papert posed the development of a computer vision system as an undergraduate summer project. The goal was to be able to extract information such as "likely objects" and "likely background areas" from pictures. The abstract of the project reads: "The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system" [10]. This turned out to be a more difficult task than expected.

The first successful study that focused on real world images was a framework for "scene understanding" proposed by David Marr at MIT in 1972 [11]. It handled methods of so-called "low-level" vision tasks such as edge detection and segmentation (the process of locating objects and boundaries). His paper was a major milestone in the field.

In the 1980s, there were more focus on developing advanced mathematical models for executing image and scene analysis. The concept of image pyramid was widely spread, which can be used for example to scale images but is also used in a wide range of other computer vision applications [5].

During the 1990s, the most important progress in the field was the increased interaction with computer graphics, especially in the area of image based modeling [12]. A major contribution to the development of face recognition during this decade was made by Turk and Pentland. They showed that with the use of eigenfaces, which is a method that uses the characteristics of certain patterns in an image, it was possible to detect faces in cluttered images and also to decide the location of the faces [13].

In the 2000s a large part of the research conducted in the area of visual recognition has been about using machine learning techniques to solve computer vision problems [14]. This is much due to the vast amount of labelled data that is available on the Internet, which makes the learning task more effective.

2.2   Face detection

2.2.1   Applications

The task of a face detection algorithm is: Given an arbitrary image, identify all regions that contain a human face. This is not to be confused with a face recognition algorithm which in addition to detecting faces also compares them.

Like computer vision in general, face detection technology is used in many real world applications, from surveillance to marketing and entertainment.

When used in surveillance, face detection is commonly used together with face recognition. One of the earliest uses of this technology was in 1998; the Newham Borough of London used a CCTV-system consisting of a few hundred cameras and a face recognition system where the images were compared to a police database of a hundred known offenders [15], [16], [17]. Since then, the scale of facial recognition systems have increased. The U.S. Department of State has a database with over 75 million photographs. However, this is triumphed by Facebook where more than 250 million images are uploaded daily [18]. According to Transparency Market Research, the global facial recognition market was valued at USD 1.17 billion in 2013 [19].

Face detection is also a very popular feature in digital cameras, used by manufacturers such as Canon and Fuji Film. It is mainly used to improve auto focus, and in conjunction with face recognition to add "tags". Other large companies who use face detection or face recognition technology in their products are: Apple, Adobe, Google, Sony and Microsoft. Despite the technology's wide variety of uses, the use of it to measure attention seems limited as is evident by the lack of published research on the subject.

In the early days of face detection technology, performance was a big issue. In 2001 the so-called Viola-Jones algorithm [20] revolutionised the field of face detection. It was developed by P. Viola at Microsoft Research in Redmond, and M. Jones at Mitsubishi Electric Research Laboratory in Cambridge. This algorithm could detect human faces in an image in real-time, and at the time it was about 15 times faster than state-of-the-art face detectors with comparable accuracy [21].

2.2.2 Methods

While there is a clear definition of a face detection algorithm, the different techniques used are not so clearly defined. Erik Hjelmås and Boon Kee Low [22] categorises the different approaches to face detection as feature-based and image-based, with the feature-based approach having additional sub categories. However, Ming-Hsuan Yang, David J. Kriegman and Narendra Ahuja [23] defines the methods as: knowledge-based, feature invariant approaches, template matching and appearance-based. Figure 2.2 shows an overview of different face detection approaches.

Knowledge-based Knowledge-based methods is a system which uses the knowledge of how different facial attributes relate to one another and derives rules from this anatomical information. These rules are used in a so called cascade of decision making when trying to detect a face. Each step in the cascade can be viewed as a question that needs an answer in the form of a yes or no in order to continue to the next step; Is there a nose? Is there eyes? And so on.

Feature-based Feature-based methods relies on discovering well defined characteristics as regions in an image such as mouths, noses and eyes. The next step is to verify if the detected feature is in an feasible location of the image, this can be achieved by geometrical testing.

6

Template-matching Template-matching methods uses a number of stored standard templates of a face to either describe it as a whole or in separate features. The detection is then based on computing the correlation between an input image and the stored models.

Appearance-based Appearance-based methods also uses templates in a similar way as template-matching. However, instead of letting experts define the templates these models are trained with images of faces and non-faces and should then be able to distinguish between the two. This method rely on techniques from machine learning and statistics.

In a recent study, Farfade [24] et al. propose a method based on deep learning, called Deep Dense Face Detector. Deep learning is a new area of machine learning whose task is to find complex relationships among data using algorithms to learn in multiple levels of representation, each corresponding to different levels of abstraction and is often done with artificial neural networks. The motivation of the study was to find a more accurate and effective way to solve the problem of multi-view face detection; faces from different angles or faces that are not fully visible, this was achieved partly by training the network with a total amount of 200K positive images and another 20 million negative images. Comparisons with previous face detection methods shows that the deep learning technique has similar or better performance, in conjunction with it being less complex.
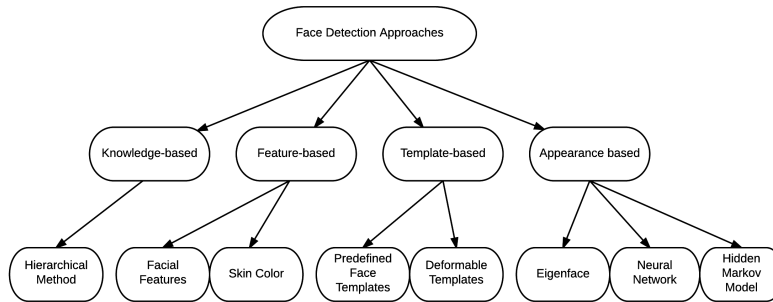

Figure 2.2: Overview of different face detection approaches.

2.2.3 Viola-Jones

The Viola-Jones algorithm is based on previously done research by using a classifier, which is built by using the AdaBoost learning algorithm developed by Yoav Freund and Robert Schapire, [25] for which they were awarded the Gödel Prize in 2003. This classifier is used to select a subset of visual features from a much larger set of potential features. In order to have the algorithm focus on the face-like regions and disregard the background portions of the image, Viola-Jones combines classifiers in what is known as, a cascade. Furthermore, the use of an image representation called the "Integral Image" allows features used by the face detection algorithm to be computed rapidly.

Object recognition, including face detection, relies on the fact that objects of the same type share some common visual features. Some experimental results published by Viola and Jones [20] shows that the first feature selected by AdaBoost uses the property that the region of the nose and cheeks is often brighter than the region of the eyes. The next selected feature focuses on the property that the bridge of the nose is brighter than the eyes.

2.3 Human attention

William James, one of the most influential psychologists and philosophers in the United States defined attention in his major work, *The Principles of Psychology*, in the following way: "Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. ... It implies withdrawal from some things in order to deal effectively with others" [26]. To put this in academic terms; the brain has a limit regarding the ability to process the stimuli that comes from the physical world, instead it depends on the cognitive process of attention to help it allocate resources depending on where it is needed at the time [27].

In the theory of human information processing there is a concept described by psychologists as "serial bottlenecks". These bottlenecks shows that there is a point at which it is no longer possible to process information in parallel. One example of parallelism in the motor system is that most people can perform different tasks simultaneously when these tasks depend on different motor systems (such as chewing gum and walking), but getting one motor system to perform two things at once becomes much more difficult and shows that there are limits to this parallelism. So if there is a bottleneck, the cognitive systems need to choose what information to deal with and what to ignore [28]. Research shows that these bottlenecks also appear in the visual systems, where there is a limitation to how many objects that can be processed at the same time in a visual scene. The procedure of selecting between sources of information, either by intensifying the processing of certain objects or to decrease the information from others, is called "selective attention". The attention mechanism also consists of two independent stages: a pre-attentive stage that does not have any limitations and operates on the whole field of vision, proceeding this is an attentive stage that has a capacity limitation and can only handle one item at a time. The object is selected when it goes from the first stage to the second [29].

2.3.1 Attention and students

Several studies have been conducted on the subject of attention and learning in a classroom environment. McKeachie points out that students attention will vary over the course of a lecture and suggests the use of interactive activities to keep the students focus [30]. Johnstone and Percival [31] measured the attention of students during a lecture by carrying out a classroom study that involved observers recording interruptions in the attention of the students. The study showed that there were lapses in student attention in the beginning of the lecture, 10-18 minutes in and more frequently as the lecture went on. In 2010, Bunce et al. did a study on students attention by having them report their

attention lapses using clickers, a small remote device that registers button-press responses. With this data they examined the relationship between attention lapses and different pedagogical methods used by the instructors. The results showed that student attention is higher during non-lecture segments such as demonstrations and questions, and also that attention was higher immediately after these active learning sections [31]. This shows that it can be of importance for a lecturer to know when to change context during a class.

Chapter 3

Method

3.1 Definition of attention

The first obstacle encountered was how to measure attention. In order to measure if someone is paying attention, a definition of attention was required. It is obvious to humans when someone is paying attention. However, this is not the case for a computer. During this process, numerous questions arose. In the context of a speaker addressing an audience, what if someone is looking away, but is still listening? What if a person is looking at the speaker, but not listening? The decision was made to define attention from a human perspective. This meant that we disregarded whether someone was actually listening and focused on where the person was looking since this is how a human would gauge if someone else is paying attention, i.e. attention was defined as "if someone is looking straight at an object, then their attention is focused at that object". The fact that people tend to pay attention to what they look at is also supported by published research [32].

3.2 Data sets and software

The data set used to evaluate the face detection algorithms consisted of pictures, each one containing a human acting like someone would during a lecture. In each picture, the human subject was in the centre and the background was free of other objects. The choice of using pictures instead of film was made to give better control of the variables affecting the experiment; pictures with insufficient lighting or contrast could be filtered out. Due to the definition of attention, the data set was first reviewed and partitioned into two subsets. One consisting of pictures where the subject was attentive and the other consisting of pictures where the subject was inattentive. This gave a baseline reference to be used in the test. Figure 3.2 shows an example of the data set before the partition. There were 108 pictures of attentive subjects and 127 pictures of inattentive subjects. Since the pictures were used to evaluate the algorithm, as opposed to training it, the difference in the amount of pictures should not affect the results.

A small Java program was written to test the OpenCV library and its face detection algorithm. The program did this by letting the algorithm try to detect a human face in each picture of the data subsets and record the number of

successful attempts. An important distinction to be made is that the successful detection of a face in an image containing an inattentive subject is considered a false positive in the context of measuring attention. The evaluation process is shown in schematic form in figure 3.1.

In order to make a more thorough evaluation, the programming library was used with three of the pretrained object-recognition files included with the library. The used files were: haarcascade_frontalface_alt.xml, haarcascade_frontalface_alt2.xml and haarcascade_frontalface_alt_tree.xml. These were chosen due to the fact that they, theoretically, should be better suited to the task of measuring attention; since they are trained with front view pictures of faces the number of false positives should be reduced.
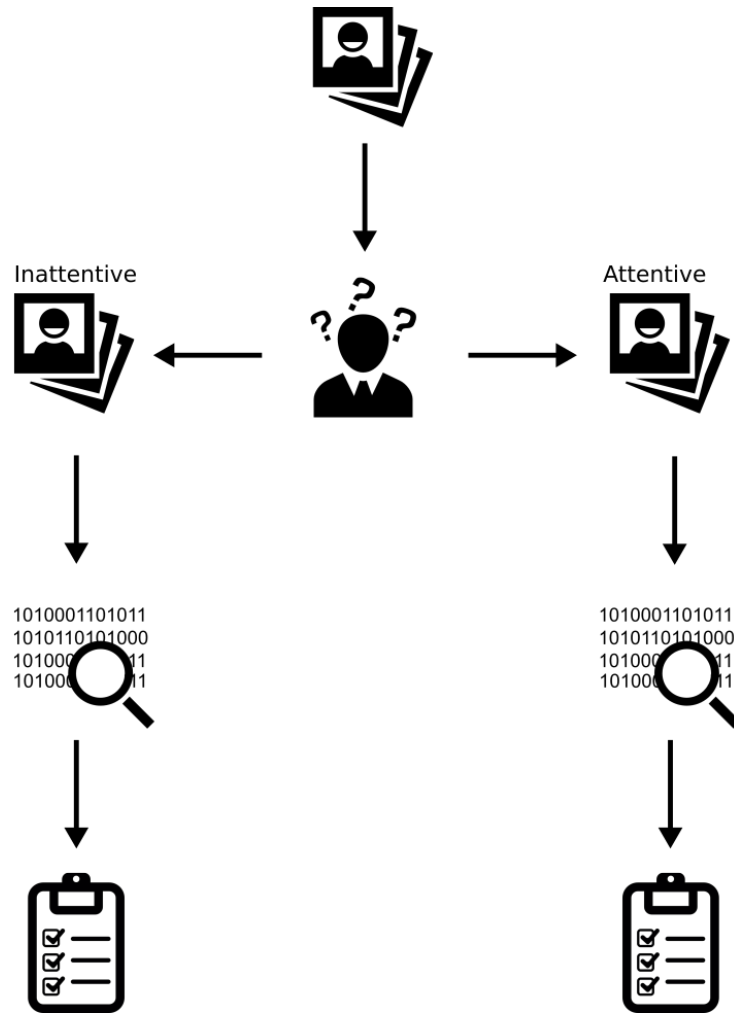

Figure 3.1: Schematic representation of the evaluation process; from the partitioning of the pictures to the recording of the number of pictures with successful face detections. *Icons designed by Freepik*

Figure 3.2: Example of pictures from the data set. The pictures on the first row shows an inattentive subject and the pictures on the second row shows an attentive subject.

Chapter 4

Results

Figure 4.1 presents the detection rates achieved by using different training files. The used training files: haarcascade_frontalface_alt.xml, haarcascade_frontalface_alt2.xml and haarcascade_frontalface_alt_tree.xml are abbreviated as: alt, al2 and tree. Of the 108 pictures where the subject was attentive, OpenCV made a successful face detection in 102, 104 and 104 pictures using tree, alt, and alt2. This yields detection rates of 94%, 96% and 96%.

Using the data set where the subjects were inattentive, there was a more substantial difference in the number of detected faces. Of the 127 pictures where the subject was inattentive, OpenCV made a successful face detection in 19, 32 and 40 pictures using tree, alt and alt2. This equals detection rates of 14%, 25% and 32%. In the context of measuring attention, these are false positives.

This means that the difference between the highest and lowest detection rates, when detecting attentive subjects, is 2%. However, the difference between the highest and lowest number of false positives produced is 56%.
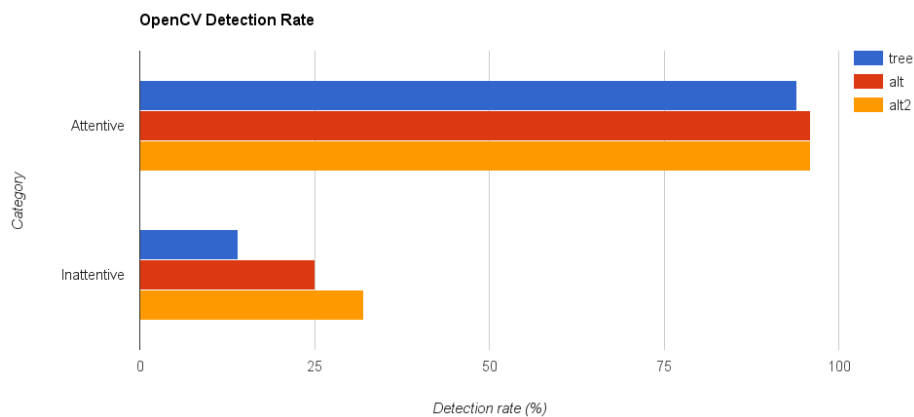

Figure 4.1: Detection rates.

A majority of the false positives, 90%, came from pictures similar to those marked with a square in figure 4.2.

Figure 4.2: The "scale of attention" of the subjects in the pictures of the data sets. Source of pictures: "http://www-prima.inrialpes.fr/Pointing04/data-face.html" [33]

Chapter 5

Discussion

In order for the software to successfully measure attention, it first had to have a high detection rate when used with pictures of people being attentive, e.g looking straight into the camera. Since OpenCV had nearly 100% detection rate regardless of which training file was used, this criterion is considered to have been successfully met. The second criterion was that the software produced a low amount of false positives. OpenCV had 14% false positives when the alt_tree training file was used. In the context of this study and the proposed practical use of the technology, this is considered "low enough".

It is also interesting to note that a majority of the false positives came from images where the subject was "almost attentive", as shown in the square in figure 1. This indicates that an "attention detection system" using OpenCV would be unable to detect minor differences in peoples attention. This means that the system's performance would be close to that of a human, rather than superior. A good example would be a teacher giving a lecture; suppose the teacher is standing at the front of the classroom, looking at the audience in the back of the room, then it might be difficult to detect minor differences in their attention, e.g. if someone is looking just beside the teacher.

As mentioned earlier in this report, the Viola-Jones algorithm can be trained to recognize certain objects in an image. In the case of detecting faces it needs images of faces and images that contain no faces, and to become accurate it has to be a very large dataset of images. For example, the "Deep Dense Face Detector" algorithm, mentioned in section 2.2.2, was trained on 200,000 positive images and a further 20 million negative images. As this study only used training files included with OpenCV the result is limited by how these files were trained. There was an almost negligible difference in the detection rates of the attentive images, but a more substantial difference in the number of false positives. This suggests that if the algorithm had been trained with pictures of attentive and inattentive subjects the result could have been even better.

It is important to bear in mind the limitation of this study and how it affects the outcome, the results would most likely change if it were conducted in a more real life setting. The study is limited mainly by the definition of what attention is, described earlier in chapter three, and do not include all possible scenarios of when a person is really paying attention. For example, it might be the case that a student looks down into the notebook when taking notes, yet paying attention. In this scenario, most likely would a face not be detected, but in order to get

a numerical value when measuring there has to be some benchmark. Also, the data sets used do not include all variations of possible occluded faces which could influence the outcome. Furthermore, other parameters such as lighting issues, backgrounds, poor image quality were ruled out. It must be kept in mind that this as well would have an impact on the result.

It should also be taken into consideration that other face detection implementations could yield different result and these may or may not be better suited for measuring attention. For example, the "Deep Dense Face Detector" algorithm, has a considerably high accuracy for detecting faces at different angles and head rotations, and would probably detect more faces in the category of inattentive faces, making it a less ideal choice of algorithm for our purpose.

Chapter 6

Conclusion

The information about the attention of an audience would be a valuable resource for a speaker in several aspects and the objective of this study was to investigate if it is possible to measure a person's attention in a controlled environment using OpenCV and the Viola-Jones algorithm. Compared to the more general problem of measuring attention in an audience, this introduced some restrictions, and with these restrictions we can now state that it is possible to do so with an accuracy that can be considered adequate. However, as mentioned in the discussion, we cannot conclude that the same results would be obtained if this study were to be made in a more natural environment. In terms of directions for future research, further work could include training the algorithm with images for the specific purpose and also to use a larger data set containing additional images consisting of even more examples of the different ways people behave in an audience.

Bibliography

[1] M. S. Young, S. Robinson, and P. Alberts, "Students pay attention! combating the vigilance decrement to improve learning during lectures," *Active Learning in Higher Education*, Mars 2009.

[2] D. H. Ballard and C. M. Brown, *Computer Vision*, 1st ed. Upper Saddle River, NJ: Prentice Hall, 1982.

[3] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2011th ed. London, United Kingdom: Springer, 2010.

[4] D. A. Forsyth and J. Ponce, *Computer Vision, a Modern Aproach*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2003.

[5] G. Bradski, *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. Sebastopol, CA: O'Reilly Media, 2008.

[6] Project Gutenberg. (2014, October). [Online]. Available: www.gutenberg.org

[7] Internet Archive. (2014, January). [Online]. Available: http://archive.org/scanning

[8] J. Howard. (2012, March). [Online]. Available: http://chronicle.com/article/Google-Begins-to-Scale-Back/131109

[9] L. G. Roberts, "Machine perception of three-dimensional solids," Ph.D. dissertation, Dept. Elect. Eng., Massachusetts Institute of Technology, Cambridge, 1963.

[10] S. Papert, "The summer vision project," Massachusetts Institute of Technology, 1966, memo.

[11] D. Marr and A. Vision, *A computational investigation into the human representation and processing of visual information*, 1st ed. WH, San Francisco: Freeman and Company, 1982.

[12] S. Seitz and R. Szeliski, "Applications of computer vision to computer graphics," *Computer Graphics*, vol. 33, no. 4, pp. 35–37, November 1999.

[13] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, 1991, pp. 586–591.

[14] W. Freeman, P. Perona, and B. Schölkopf, "Guest editorial: Special issue on machine learning for vision," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 1–3, May 2008.

[15] The Parliamentary Office of Science and Technology, *postnote*. London, United Kingdom: The Parliamentary Office of Science and Technology, 2002, no. 175.

[16] S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition*, 2nd ed. London, United Kingdom: Springer, 2011.

[17] S. Balgani and S. Sangeetha, "An efficient approach to improve response time in multibiometric patterns retrieval from large database," *International Journal of Computer Science and Network Security*, vol. 14, no. 5, pp. 102–107, May 2014.

[18] M. Drumwright, Ed., *Ethical Issues in Communication Professions*. London, United Kingdom: Routledge, 2013.

[19] Transparency Market Research. (2014, August 28). [Online]. Available: http://www.transparencymarketresearch.com/pressrelease/facial-recognition-market.htm

[20] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, July 2004.

[21] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl Inf Syst*, vol. 14, no. 2008, pp. 1–37, December 2007, doi:10.1007/s10115-007-0114-2.

[22] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 83, no. 2001, pp. 236–274, October 2001, doi:10.1006/cviu.2001.0921.

[23] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, Januari 2002.

[24] S. S. Farfade, M. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," *arXiv preprint arXiv:1502.02766*, 2015.

[25] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," AT&T Bell Laboratories, Murray Hill, NJ, Tech. Rep., 1995.

[26] W. James, *The principles of psychology*. New York: H. Holt and Company, 1890.

[27] F. Katsuki and C. Constantinidis, "Bottom-up and top-down attention: different processes and overlapping eural systems," *The euroscientist*, vol. 20, no. 5, pp. 509–21, October 2014, doi:10.1177/1073858413514136.

[28] J. R. Anderson, *Cognitive Psychology and Its Implications*, 6th ed. London, United Kingdom: Worth Publishers, 2005.

[29] V. J. Dark and W. A. Johnston, "Selective attention," *Annual Review of Psychology*, vol. 37, pp. 43–75, Februari 1986, doi:10.1146/annurev.ps.37.020186.000355.

[30] W. McKeachie and M. Svinicki, Eds., *McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers*, 13th ed. Boston, MA: Cengage, 2010.

[31] D. M. Bunce, E. A. Flens, and K. Y. Neiles, "How long can students pay attention in class? a study of student attention decline using clickers," *Journal of Chemical Education*, vol. 87, no. 12, pp. 1438–1443, October 2010, doi:10.1021/ed100409p.

[32] R. Stiefelhagen, "Tracking and modeling focus of attention in meetings," Ph.D. dissertation, Fakultät für Informatik, der Universität Karlsruhe, Karlsruhe, Germany, 2002.

[33] N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial features," in *Proceedings of Pointing*. Cambridge, UK: ICPR, 2004. [Online]. Available: http://www-prima.inrialpes.fr/Pointing04/data-face.html