Identifying characteristics of passers-by to provide dynamic advertising in public spaces using Computer Vision

By

Daniel Hodge

HOD15589698



A dissertation submitted in partial fulfilment of the requirements for the degree of

MSc (Hons) Computer Science

School of Computer Science

University of Lincoln

2019

# Abstract

# Table of contents

[Abstract 2](#_Toc17302774)

[Table of contents 3](#_Toc17302775)

[Introduction 4](#_Toc17302776)

[Rationale and Benefits 4](#_Toc17302777)

[Aims and Objectives 4](#_Toc17302778)

[Background 5](#_Toc17302779)

[Report Structure 6](#_Toc17302780)

[Literature Review 7](#_Toc17302781)

[Methodology 10](#_Toc17302782)

[Project Management 10](#_Toc17302783)

[Gantt chart 10](#_Toc17302784)

[Risk Matrix 12](#_Toc17302785)

[Software Development 13](#_Toc17302786)

[Toolsets and Machine Environments 15](#_Toc17302787)

[Research methods 16](#_Toc17302788)

[Design, Development and Evaluation 17](#_Toc17302789)

[Requirements 17](#_Toc17302790)

[Design 17](#_Toc17302791)

[Development 18](#_Toc17302792)

[Evaluation 21](#_Toc17302793)

[Findings and Conclusion 25](#_Toc17302794)

[Reflective analysis 26](#_Toc17302795)

[References 26](#_Toc17302796)

[Appendix 30](#_Toc17302797)

# Introduction

## Rationale and Benefits

This has been undertaken for two distinct reasons. The first is improving and assessing the implementation of state-of-the-art Computer Vision techniques, specifically when determining the context of a situation, such as the type of people in an area, and the relationship they have to others in their group. Furthermore, the quality of these techniques in a public space will also be considered.

The second reason is to illustrate and assess the applications of these Computer Vision techniques, by using the example of tailored digital advertising in public spaces. This is an ambitious goal and solves a real problem that already exists within advertising, which is how to convert all the success of tailored online advertisements to those that can be found in public places such as shopping centres, airports and high streets.

The primary benefit of this project is to provide a blueprint for others to follow when approach similar problems, such as an analysis of current research, and an artefact showing how these algorithms and techniques can be implemented in future work.

## Aims and Objectives

The aim of this project is to develop a system which integrates tailored advertisements with digital signage to increase the return on investment of adverts in public spaces. To achieve this aim, the following objectives have been laid out:

Objective 1: Develop a system which can find faces on a photo.

Objective 2: Identify the individuals’ features, such as age, gender and hair.

Objective 3: Gather secondary data, such as location and time

Objective 4: Infer the context of groups, such as family, friends or couple.

Objective 5: Use the gathered information to recommend an advert

Objective 6: Run this system in real time

## Background

Online advertising has seen exponential increases in effectiveness over recent years, as a result of advertising services (spearheaded by Google) creating profiles of each individual’s online identity, such as their basic details (gender, age etc) and their interests, for example if they watch sports or search for cooking recipes. Advertisers are then encouraged to use this information to scope into their specific demographic, such as beer companies advertising to young American males interested in sports around the time of major sporting events such as the Super Bowl. This is a much more efficient approach than the old method, which consisted of placing advertisements on a site for a fixed period of time (or until a certain budget was hit), and just hoping that the correct audience would see it. Small amounts of targeting could still be achieved however, by choosing a website which matched potential user interests.

Advertising on digital screens in public spaces is still vulnerable to these issues, as these devices have no means of discerning the characteristics of passers-by, and therefore wouldn’t be able to change the advertisement as a result. The primary issue is that, unless the user can be persuaded to stop and sign into an online account (such as social media), there is no easy way of collecting the users’ data. Therefore, more complicated means are required, such as using Bluetooth to scrape data from the phone or using Computer Vision to analyse the passer-by. This can’t provide the wealth of data that online advertising can, however it is not unreasonable to expect it to be able to build a basic profile, such as age, gender and facial features. This minimal amount of information would allow ads to be tailored to the profile of the person in front of them, meaning adverts could be significantly more cost effective.

The opportunity to begin targeted advertising in public spaces has recently surfaced due to two factors. The first is the increased number of digital screens used for advertising. While they aren’t completely widespread yet, they have seen a large increase, potentially due to the fact that they require less human involvement than the typical paper advertisements, which must be replaced manually, as opposed to digital, internet connected screens, which can simply have the adverts be updated on a server. These provide the opportunity to be more interactive with adverts in many different ways, which can have a large array of positive impacts, such as getting user responses, and providing a more interesting experience for the user, which could increase the chance of them actually buying the product. The second factor is the improvement in methods to gather and analyse user data in a short period of time, due to improved algorithms, and fields like Big Data and Internet of Things, which provide the framework for these more connected public spaces.

## Report Structure

There are 5 distinct sections used to explain this project. The first is a literature review, which is used to survey the field of research being targeting. This can be used to highlight areas which require more research, or perhaps to find the current state of the art techniques being used in the field. Ultimately, completing this process allows this project to help place itself within its relevant field.

The second section is methodology. This refers to the general approaches and tools used to tackle the predefined problem. This section is split into 4 subsections. The first of which is Project Management, which consists of a Gantt Chart and Risk Matrix in order to plan the scale and timeline of the project, as well as mitigate as much risk as possible in relation to this. The second is Software Development, referring primarily to the framework that was used to produce the artefact, such as Waterfall or Agile. Third is toolsets; describing the tools, software, programming languages, libraries and datasets used in the development of the entire project, whether they be for producing the artefact, or for managing the project as a whole. The last subsection that forms Methodologies is Research Methods, which focuses on the way in which the projects aims and objectives would be evaluated.

The third section is Design, Development and Evaluation. This encompasses the development of the artefact, from start (such as requirement gathering and design) to finish (such as development and evaluation). This will include any methods or documents of particular interest, such as algorithms used, design documents, and a series of ways to evaluate the results of the project.

Following on from the evaluation of the project is the Findings and Conclusion section. This will analyse the results of the project, such as accuracy ratings, and attempt to provide a consensus on the entire project, such as the validity of the initial hypothesis, and how well the aims and objectives have been met.

Finally, a reflective analysis will be provided, to give a retrospective on the project, such as what could have been done differently, as well as some insight into what went well for the project, and what held it back. This will be followed by the references and appendices.

# Literature Review

Targeted advertising has significantly improved cost effectiveness of online ads, as discussed in (Farahat and Bailey, 2012), with adverts targeting users based on browsing history, demographics, user profile and more. This paper goes on to analyse the impact of targeted advertising, finding that targeted ads have the ability to generate clicks increasing by around 4.5 times, and discussing the cost of targeting to the advertiser. They also analyse the fact that targeted adverts are significantly more cost effective (a third) for niche companies, due to the ease of accessing their customer base. However, as this paper was released in 2012, the results may have changed dramatically, especially as a result of more awareness of online tracking and privacy, with people using tools such as VPNs to remain anonymous online.

This has led to more research being conducted on how to advertise in a more highly connected world, with (Aksu et al, 2018) giving a smart car as an example of how this could be done, advancing targeted advertisements to a ubiquitous level that follows us wherever we go, due to the 'digital signature' that is created about us, and the fact that connected technology follows us everywhere we go now, such as wearable technology and smart screens in public. However, this brings up a myriad of privacy and security concerns, which is why (Alt et al, 2012) proposes the use of interactive public displays, which provide useful information such as maps or weather to entice a passer-by, and then show ads once they have started using the system. This doesn't personalise the ads straight away, although it provides the opportunity to do so in several ways, such as requiring the user to sign into an account (and using data from social media for example) or providing more time for a computer vision algorithm to come to an accurate conclusion on the profile of the user. This could even be verified by the user, as they are already most likely using the system, so can answer some simple questions to verify estimates. However, as with many pervasive technologies like this, transparency is key, as people will start to distrust public systems if they aren't sure what is being tracked and stored, and the ways in which that data is used.

Another approach to bringing new technology to advertising (Lyons et al, 1998) uses the camera not only draw attention, but then to enhance the advertising experience, such as placing the image of the product onto the targeted customer to show how it would look. This approach has two appeals; the first is that it allows the user to 'try before you buy', without actually going through the effort of going to the changing rooms, or even having to find the product. Furthermore, this system could be expanded on to target adverts to their specific audience. For example, makeup companies may be interesting in showing their makeup applied to passers-by, but may want to specifically target skin tones in which the product is aimed at. The second appeal is that this is a fun gimmick, which is likely to draw a lot of attention before. This may only be short term, but it provides the opportunity for people to become comfortable with the technology, which is an important part of tracking user habits, especially when it is making it as obvious as returning an edited image of the passer-by.

Contrary to this, (Exeler et al, 2009) considered the ways in which public displays could attract attention whilst remaining non-intrusive, an important step to consider when looking using vision systems to watch individuals. The system has a scanned face on a screen, which reacts to passer-by's emotions, and attempts to emulate them. This is successful, and encourages further interaction with the screen, especially when done by people with similar characteristics. The downside to this is that some were deterred from the system as a result of distrust, which again brings in privacy concerns to the conversation. These concerns are further discussed in (Tucker, 2014), with their findings suggesting that giving users control over how their data is used can have a positive impact on the success of personalised adverts as a whole, most likely a result of the transparency of the system, as users understandably want to be informed about how information about them will be used.

A system was developed to change advertisements based on demographic information (Tian et al, 2012), by using Anonymous Viewer Analytics and Data Mining to collect data, which showed an increase in targeting accuracy over context based targeting, specifically when using Decision trees. Furthermore, this was a paper released by Intel in 2012, showing that this field has been receiving commercial interest for a long time. The system developed in this paper could be expanded upon to provide further improvements in accuracy, using newer technologies such as recent advancements in Machine Learning, and provides useful information on the evaluation of such a system.

The remainder of this literature review will be focused on the different studies and systems attempting to improve the way we gather specific demographics from images of faces. The first example of this is (Huerta et al, 2014), which fused texture and local appearance-based descriptors to achieve fast and accurate results when estimating age, producing a Mean Average Error of 4.25 years, which is sufficient when trying to recommend products or advertisements to potential customers based on age, as it would usually be age groups targeted instead of specific ages anyway. Another benefit of this approach is that it is a robust technique, requiring no additional cues. An early implementation of automatic age estimation (Geng et al, 2007), completed in 2007, used aging patterns as samples, instead of individual facial images, by first modelling the aging pattern, and consequently estimating the age of the face by finding its position in the pattern. This has become more of an industry standard, with a significant amount of other implementations doing this too.

In another study (Zhang et al, 2017), the age of participants is gauged by comparing between two people, and choosing who is older, as a way of training a model to make more accurate age estimations. The results of this are used in a deep convolutional network to produce estimations of overall age. This more modern paper (2017) used these advanced techniques to achieve a MAE of 2.87 on the MORPH dataset, and 2.52 on MORPH2, which places it amongst the top performers, due to the benefits of deep convolutional networks. Finally, (Guo et al, 2009) takes advantage of Biologically Inspired Features to train a Kernel Partial Least Squares regression model to estimate age, due to its ability to reduce feature dimensionality and learn the aging function simultaneously in a single learning framework, a factor that places this algorithm above traditional SVM algorithms in accuracy. Being an early study (2009), this was another trendsetter, with many future studies and implementations using Biologically Inspired Features to train more complex models, increasing overall accuracy.

When trying to recognise gender, (Ng et al, 2012) identified the primary challenges as a combination of human factors (such as age, ethnicity and accessories) and the image capture process, e.g. camera angle, lighting or image quality. This paper also helps classify gender classification problems into two groups; geometric based and appearance based methods. Geometric based methods of feature extraction use the information about the distance between facial features, such as the distance between eyes, or the distance from the nose to the lips. This method was used in multiple papers, such as (Shakhnarovich et al, 2002), which was an early paper which explored geometric based methods based on the now popular Viola Jones algorithm. An interesting technique used in this paper was to combine estimates from many facial detections in order to reduce error rate as a result of noise.

Local Binary Pattern histograms, as presented by (Ojala et al, 2002) are used in (Lian and Lu, 2006) to generate a single vector which represents the face. This LBPH is found by dividing the face into small regions, and taking both shape and texture information. Support Vector Machines are then used as the classification model, which outputs an average accuracy of 94%. Although this appears to be a high accuracy, incorrectly identifying gender by 6% could cause significant issues, due to people potentially insulted by the insinuations. An alternate approach is taken by (Li et al, 2012), which uses both facial features (forehead, eyes, nose, mouth and chin) and external information such as hair and clothing, to classify the image into a given gender, in order to overcome the issues of occlusion (hair, glasses etc. covering the face). This model was slightly different to others, as it classified the images separately based on each type of feature, then combined them afterwards using various strategies, such as Fuzzy integral. As this was an early approach to something like this, being published in 2012, the accuracy of 95% is a significant point, as this wasn't able to use more recent advancements in Machine Learning. Similarly, (Kalam and Guttikonda, 2014) uses facial distance measures as a progenitor for gender classification, such as the distance between the midpoint of the right eye and the midpoint of left eye, and the distance between the lips and the nose. Classification is then applied using this data. This paper also explores the types of pre-processing used, such as converting the RGB image into a two dimensional grey scale image instead. Another step was to perform noise reduction, and this paper weighed up the benefits of different filters, ultimately choosing the median filter due to it's ability to preserve image quality. This paper returns the highest accuracy seen yet, at 95.6%.

Many studies have observed the benefits of gathering groups of demographics at the same time, due to the way they each impact each other, as explored by (Guo et al, 2009), who states that gender recognition accuracies can be 10% higher on adult faces than young or old. This paper also used Biologically Inspired Features, showing that this method is in fact widespread, and can produce high accuracies on different data sets, and alongside different classifiers. As a result of this paper, the following studies gather more than one demographic feature, to improve accuracy as a whole. One way to improve this accuracy is given by (Han et al, 2014), who again does so using Biologically Inspired Features, which is explored further in (Guo et al, 2009), by extracting these features to aid the hierarchical approach consisting of between-group classification, and within-group regression, to estimate age, race and gender. An interesting part of this study is that it compares results against human observers, and finds that the system is more accurate. Another method of demographic classification for age, race and gender is explored by (Yang and Ai, 2007), who extract Local Binary Pattern Histogram features for texture description, in order to generate a more accurate classifier. This implementation also used AdaBoost, providing a more unique take on the problem.

Sometimes data sets aren’t perfect, which is why (Moghaddam and Yang, 2000) explored gender classification on thumbnail images (21 x 12 pixels), using SVM’s, and tested the performance against other classification algorithms, and human participants, to show its superior accuracy. This paper actually produces incredibly high accuracies, with the SVM model outputting an error rate of 3.4% using low resolution images, whilst humans produce an error rate of 6.7% on high resolution images.

Content based recommendations

Collaborative filtering

# Methodology

## Project Management

### Gantt chart

The aim and objectives have been broken down into a series of tasks, that make it significantly easier to plan a project with the use of a Gantt chart. This gives me more milestones to aim towards and assists me in knowing how on track I am, which is very important given the magnitude of this project. If it turns out I’m not on track, I can act to ensure the project as a whole won’t suffer, such as reducing the scale of certain parts, such as identifying less features. These are the tasks I have laid out:

1. Identify faces in an image

2. Get features of the face

3. Get other features, such as hair

4. Use information to recommend an appropriate advertisement

5. Evaluate performance

6. Infer context of the group (e.g. family, friends, couple etc)

7. Use this information to make more accurate recommendations

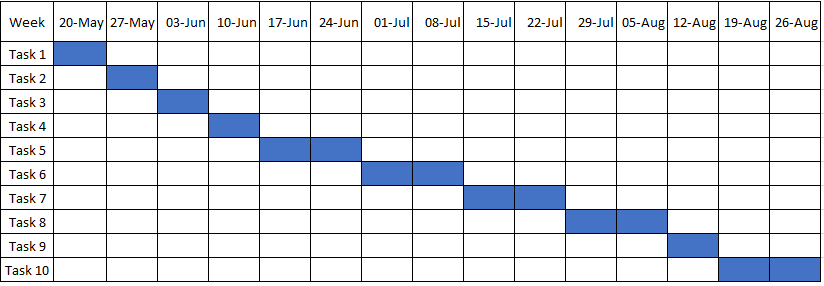
8. Compare the old recommendations to the new

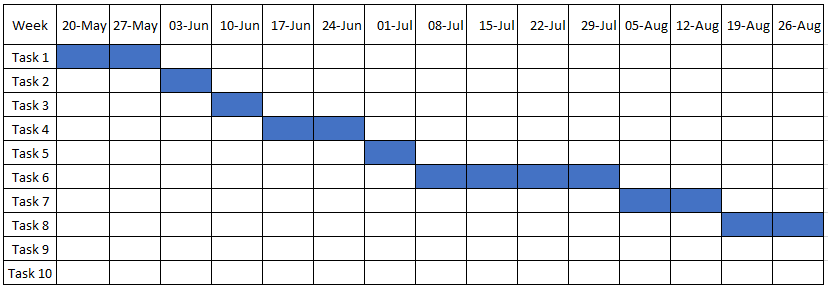
9. Implement the system in a real time, public environment

10. Retrieve information such as time and location

Task 1, 2 and 3 were all about using the Microsoft Cognitive Services API to acquire the demographic information regarding passers-by. These tasks were the most likely to be finished relatively quickly, due to their

Plotting these tasks provided this Gantt chart (figure X), predicting the amount of time the project would take to complete.



However, the project did change slightly throughout development, due to unforeseen circumstances. Therefore, the Gantt chart in figure X shows the actual development time for the project.

The reason for the discrepancy at the start is that it took significantly longer to set up the API and receive a key. However, once this was done, the first three tasks were relatively simple and didn’t take too long, as there is a large amount of documentation available for Azure API services. Task 4, which was to create a system which recommended the advertisement based on the information it received, took an extra week, and in hindsight only timetabling one week may have been an oversight, as this was a large part of the project which required a larger time frame to get right.

Evaluating the performance of the artefact so far by using accuracy metrics took a week less than expected, in part due to how easily the Cognitive API integrated with the labels of the dataset.

Task 6 was to infer the context of a group relationship, such as friends, family or unrelated. This task took a significantly larger time than planned, due to the increased level of complexity, such as the difficulty in figuring out whether people are in a group, or just separate people walking close together. However, instead of simplifying this to keep the project on track, it was seen as more important to potentially not complete the extension tasks, and complete this to a high level, as it was potentially the most important part of the project.

After this, task 7 wasn’t particularly challenging, and took as long as planned, as this had already partially been done in task 4 and was just being updated to take in new information.

The final task completed was task 8, which was to compare the old system to the new in terms of accuracy and the level of depth. Again, this took as long as planned, as the framework for evaluating the system was already prepared, so there were no surprises.

### Risk Matrix

|  |  |  |  |
| --- | --- | --- | --- |
| Risk | Likelihood | Impact | Mitigation |
| API is taken down | Low | High | Research potential replacements for the chosen Microsoft option, such as the Google Vision API, or the Amazon Rekognition API. |
| Estimations in public spaces are low in accuracy | Medium | Low | Use another API or implement a different solution to the problem based on current research. Alternatively, change the scope of the project to only internal environments, such as airports, to reduce variables such as weather obscuring image quality. |
| No suitable dataset exists for faces in the wild (e.g. data is unlabelled, or from wrong angles) | High | Medium | Try to find a dataset which includes the correct type of images, then self-label data to ensure it can still be evaluated for accuracy. |
| Groups are too difficult to distinguish from random clusters of people | Medium | Medium | Focus on having a more fleshed out recommendation system using the identified features such as age and gender. |
| Laws prevent the identification of certain characteristics using computer vision | Low | Medium | Ensure that the system isn’t reliant on any single feature, such as age, but a combination, in case one is considered too private. Don’t use any ethically questionable data, such as weight or race. Don’t store or use data in any way other than for the purpose of this project. |

Fortunately, many of these risks didn’t occur during the process of this project. The API remained live, and was more than sufficient for its purpose, allowing enough calls per minute for the artefact to function exactly as intended. The accuracy in public spaces proved to be better than expected, most likely due to the use of industry best practice as part of the API, as will be explored in the evaluation of the results. TALK ABOUT THE 4TH RISK. Finally, no new laws came into place during the process of completing this project, and it kept well within current legislation regarding privacy and data usage.

However, the risk of not finding a suitable dataset ended up occurring, which could have been a major setback. However, due to this planning, the damage was mitigated by finding an unlabelled dataset (which was taken from a good angle with a relatively high quality of images) and labelling it manually. Although this brought new issues into the project about the accuracy of the manual estimations, it fixed a significantly larger issue, which would’ve been the inability to quantitively evaluate the accuracy of the solution.

## Software Development

(Sommerville, 2011, 32)

Include agile manifesto: Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R. and Kern, J., 2001. Manifesto for agile software development.

And benefits of agile: Cockburn, A. and Highsmith, J., 2001. Agile software development: The people factor. *Computer*, (11), pp.131-133.

For the purposes of this project, the Scrum methodology was used, for a variety of reasons that will be explored. Several different methodologies were considered, such as Waterfall, Spiral or Extreme Programming (XP), to ensure there was a large amount of variety to weigh up.

The decision was made based on three equally important factors: the ability to break the artefact down, the flexibility of the project requirements and the choice between speed and quality/security of the project. In most scenarios, the developer should also consider the skills and location (on site or dispersed) of the team, however as there is only one developer in this project, this does not need considering.

When it comes to being able to break down a project into tasks, the important consideration to make is whether different milestones in the project can function (and potentially be evaluated) alone. For example, this project has been broken down into several bitesize chunks, such as using the Microsoft API to return characteristics of a passer-by, creating an advertisement recommendation algorithm, and inferring the context of a group. Each of these tasks can also be seen as its own self-sufficient unit and can therefore be seen as a project which can be broken down into tasks. However, some projects, such as a video encoding algorithm, can’t be effectively tested until the entire artefact is complete. That kind of project would lend itself well to linear methodologies, such as Waterfall or Spiral, where the project doesn’t need to be seen as anything other than one large task, and seeing that task as one single part to be designed, developed and evaluated as whole is of a great importance, whereas this project fits the agile methodologies such as Scrum and XP, as these tend to split large tasks into smaller chunks, which can be completed in a single iteration or sprint (a predefined length of time, usually between a week and a month).

Whether a project has flexible requirements has a large impact on the approach taken to complete that project. However, every project has some level of flexibility, whether that’s due to temperamental clients, uncertainty in planning or a reliance on third party tools. Therefore, it is sometimes worth accepting that there is a small amount of flexibility, and still choosing a more rigid methodology, but instead trying to mitigate that flexibility, for example trying to lock in an initial set of requirements from a client that can’t be altered, so that the project can be better planned. With that in mind, it is difficult to know where to place this project, as it has a relatively fixed set of requirements (due to an uninvolved client, who in this case is seen as the supervisor of the project) and a well-planned timescale, although it does rely largely on a third-party application. Overall, while this is only one small part of the project, the repercussions could be quite large, and would require a rapid intervention and change in direction, which points towards an Agile methodology.

Finally, a project must choose whether to focus more heavily on speed or quality/security. Quality and security are placed together as both relate to the same idea of meeting the requirements of a project to the highest standard, perhaps going past deadlines in the process. While it would be ideal to have both speed and quality/security, it is in the nature of large projects to have to choose one over the other at some point in the lifecycle. This project is a perfect example of a project that has to focus speed over quality, due to the fact that the deadline is non-negotiable, and therefore won’t allow for any extensions if improving the quality or security of the artefact is prioritised over reaching the deadline. Projects that rely on speed such as this lend themselves well to Agile methodologies, as Agile was largely developed to overcome the issue of projects surpassing deadlines on a regular basis before its inception in the early 2000s (<https://techbeacon.com/app-dev-testing/agility-beyond-history-legacy-agile-development>). Therefore, this project can once again be considered better suited to an Agile methodology than linear.

Choosing an Agile methodology is no easy task either, with the number of alternatives growing ever larger. However, here the choices will be limited to XP or Scrum, as these are reflective of most Agile methodologies, and could be interchanged with most. Scrum was chosen because of two reasons, however these are minor, and XP could’ve also been chosen with little change to the project. The first reason is that Scrum allows for slightly longer Sprints, whereas XP tend to keep them under two weeks. Having Sprints take over two weeks worked well to match the frequency of interactions with the client. The second reason is that Scrum allows the choice of what to work on in what order, whereas XP has a stricter priority order. A more lenient policy worked better for this project, due to the lack of experience in being able to plan priority in advance.

Overall, Scrum worked very well for this specific project, and allowed for significant changes to plans when it started to deviate, ensuring no time was wasted in the way that it would be under a linear methodology such as Waterfall.

## Toolsets and Machine Environments

COMPARE MORE TOOLS (USER REQUIREMENTS)

When it came to the process of developing the artefact, choosing a language was an important, but not necessarily difficult decision. As a Computer Vision project, the choices came down to MATLAB, which is an industry standard in Computer Vision due to it being centred around matrices, which is the simplest representation of an image, Python, which has the largest community in terms of creating up to date libraries that assist with a lot of the work, and C++, which also has a strong following, and allows more low level work to be done in order to develop a more efficient end product.

While C++ has access to libraries such as OpenCV, Python has access to that and more. Python has become the scientific programmers’ tool in part due to its ease of use, but more importantly because the hardest part of most jobs, the Machine Learning algorithms or rendering and outputting an image for example, are done for you. In this project, the API was used through a library called Cognitive Services, the Rest API call was made using a library called Requests, and the JSON data returned was made easier to interact with using the JSON library. During debugging, to check that faces had been found, circles were drawn around faces and outputted in an image using the PIL set of libraries. It is because of the work done by these libraries that this project could have such a large scope within such a short time scale.

To help manage the development of this artefact, a set of tools were used to improve the time efficiency of the project. One such tool was Trello, which feeds into the discussion of choosing Scrum for this project. Trello is a tool which allows the user to create a to-do list, and move items to a ‘doing’ list, and a ‘completed’ list. This is especially helpful when working as a team, however it still provides benefiting to those working in a solo project, as it helps keep track of the progress made so far and can be used to emulate the backlog in Scrum.

GitHub is a version control tool, which allows the user to store and update an entire project, with potentially multiple branches to allow for testing different ideas and approaches. Once again, most of the benefits of the tool are from working as a team, however a massive benefit to a solo project is the ability to go back to a previous version of a project if a change is made which breaks it in some way. This saves a lot of time and prevents the loss of work.

In terms of the machine environment, the artefact was completed on a Windows 7 machine with Python 3.7. However, one of the benefits of Python is the ease of making it cross platform, as only minor changes are needed, such as changing the direction of slashes in file directories.

In practice, this machine would most likely be implemented on a cloud-based system, receiving a stream of images from the site, and sending back an advertisement to display. This reduces the computational demand on site and improves the ability to change the entire system across different sites, all at once. Another benefit of running this on cloud would be that it is easy to expand as more sites are added, making it both scalable and versatile to changes. The cloud system would be easy to install, as it could be housed on a Linux distribution, as long as the required Python version and libraries are installed, which would only need doing during the initial development.

Certain considerations would need to be made to account for the display/camera systems being outdoors, however there have been large improvements in outdoor electronics, with most displays now being made using weather proof, durable glass. The system doesn’t acquire height, and shouldn’t need to operate at night, so the only requirement for the camera is a high enough resolution for the images to be clear.

## Research methods

To evaluate how effectively the artefact answers the research question, multiple steps must be taken. This is because the research question is quite broad, and covers a lot of areas, which must be covered separately. For example, the first section of the research question is “Identifying characteristics of passers-by”, which can be answered by evaluating the accuracy of the outputs of the first stage of the artefact when compared with the labels of the dataset. While this is an effective way to qualitatively test performance, the dataset unfortunately wasn’t labelled, making it difficult to compare to the results of the artefact, requiring the labels to be manually entered first.

While this solved the problem, and allowed the accuracy levels to be collected, this brought the issue of human error, as some of the factors being collected, such as age and makeup, were hard to guess. Although this does bring some doubt into the reliability of these results, the fact that these factors still had high levels of accuracy shows that the labels were likely correct, or at the very least that the artefact was producing human levels of characteristic identification.

The next part of the research question is focused on the advertisements that get recommended to the passers-by. This is significantly harder to evaluate, as the success or failure of this algorithm could be seen as subjective, since there is no one correct output for the algorithm. Therefore, a more lenient test will be devised, in which the algorithm passes or fails for each recommendation, with a fail only being awarded if the advert doesn’t apply to the passer-by. This will then be compared to figures regarding the effectiveness of static advertisements, to figure out the level of improvement.

# Design, Development and Evaluation

## Requirements

This project has six requirements, which are the set of conditions that must be completed in order for the project to be considered fully successful. Although there is a lot of crossover, they differ from the tasks laid out earlier, as some requirements may describe a way of doing something, or a certain aspect which should be avoided, as opposed to tasks which are a set of things which must be done. In a way, these requirements should be seen as the guidelines to follow when producing this project. They are as follows:

1. To perceive the basic characteristics of passers-by, most notably age and gender
2. To be able to operate in a public space
3. To advertise to individuals and groups in a way that mirrors online advertising
4. To infer the context of a group
5. To complete the process of recommending an advertisement with enough time for the passer-by to acknowledge it
6. To comply with data privacy laws

## Design

When it comes to design, this system doesn’t have many user requirements, due to its role as a research tool, not as a commercial product, or a tool that would be used by an end consumer where marketability or ease of use would be issues. Here, the system must meet the aim, which is to identify the features of passers-by, and recommend an advertisement for them. Therefore, features such as a GUI, or alternate forms of inputs, don’t need to be implemented, and so all important information will be outputted using a console based system. This leaves designing the system architecture, which refers to the fundamental structures of a software system. Planning the architecture of a system is associated with improvements in the time and quality of the project (REF), due to the ability to take a more careful, considered approach to the problem.

Zoomed out, the architecture of the system mirrors that of a client-server architecture. Each site would represent a client, capturing the input images of passers-by, and providing the output of the recommended advertisement. However, the algorithms transforming the inputs into the outputs would be performed on the server, to create a centralised system that enables the local sites to require no particularly powerful machines, only needing embedded processors that can send and receive information. In order to do this effectively, cloud computing would be used, due to its ability to scale up and down with the needs of the system, and the fact that it allows for complex processing to be done off site, lending itself well to the client-server architecture.

## Development

When beginning the development of the artefact, the first step was to get a rough idea of the sections that would need to be developed. While these overlap with the list of tasks created earlier, they aren’t identical, as these are specifically the steps that needs to be taken in development, that will roughly be transferred into functions in the final artefact. These steps are as follows:

Step 1: Access the Microsoft Cognitive Services API, and use it to return a set of JSON regarding the image sent as a parameter

Step 2: Parse through the JSON, separating the values returned into usable data (which can be compared with labels)

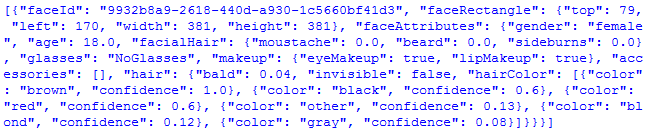
Step 3: Record the gathered information as a user profile

Step 4: Compare the user profile against pre-created advertiser profiles

Step 5: Of all potential adverts, pick one to show the user, and display it

The first step was largely taken up by signing up to use Microsoft’s Cognitive Services API and reading through the documentation to learn how exactly it worked. As Microsoft is a large company, and as such has many different organisations and individuals using their set of APIs, there is an extensive library of documentation and solutions surrounding them. Therefore, this step didn’t take a particularly large time, and most of the code was quite generic. However, work was done at this stage to ensure that best coding practices were upheld during the process of development. For example, setting the foundation for the functions used, and keeping a consistent naming convention with variables, so that the code would be legible and reusable in the future.

By the end of this step, a paragraph of JSON such as the one below in figure X would be outputted, which would need to be parsed in the next stage.

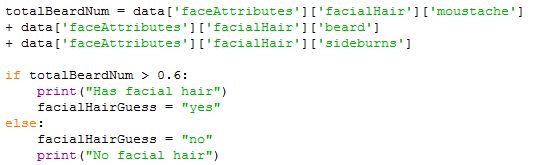


The second step was assisted largely by the ‘Json’ library in Python, which did a lot of the work itself. For example, the following screenshot in figure X shows how the gender would be accessed, which is simply done by finding the correct element in the dictionary.

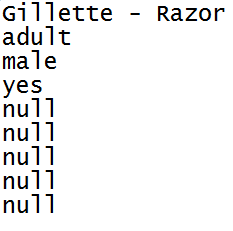


Using other languages, the JSON would’ve needed to be sifted through, taking up more time, and resulting in less legible code. This solution is perfect for the requirements, and simply outputs “Male” or “Female”, exactly as it is labelled in the dataset.

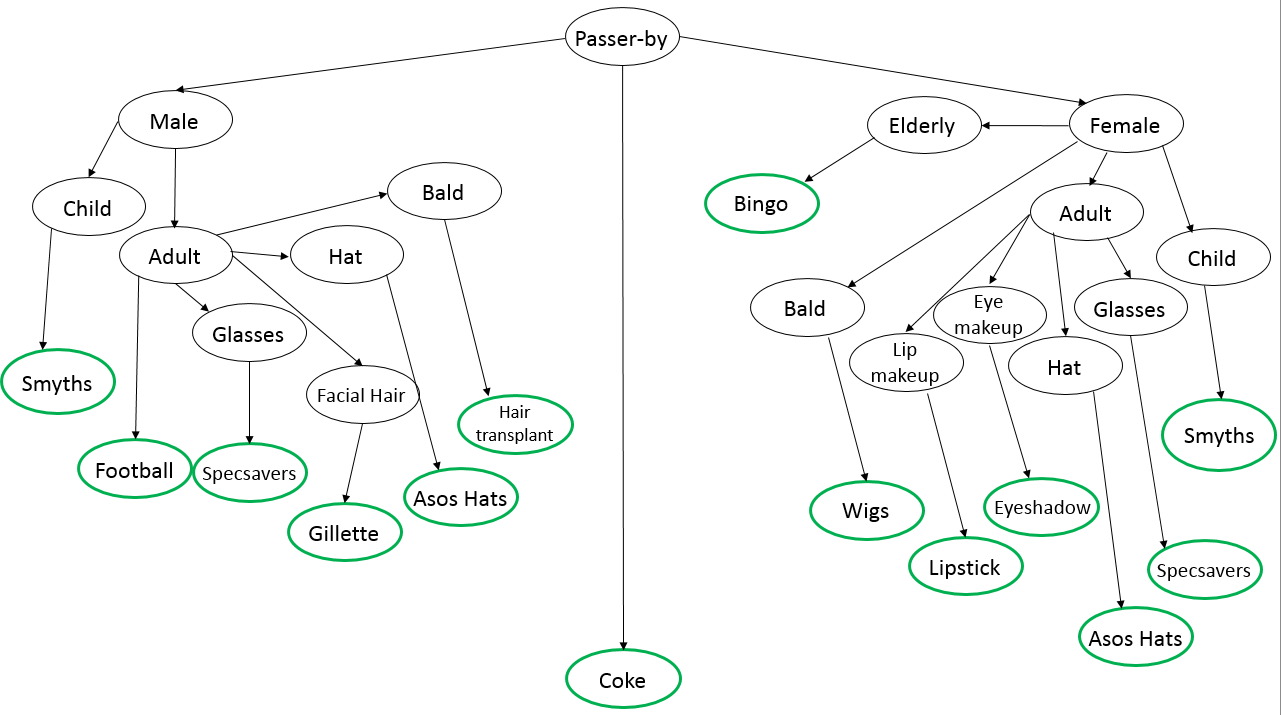
By the end of this step, the data was ready to be converted into a user profile, which would store the age and gender, as well as detecting the presence of facial hair, baldness, glasses, eye and lip makeup and hats. Age would be stored as either ‘Child’, ‘Adult’ or ‘Elderly’, and the other parameters were all Booleans. Therefore, the third step only consisted of converting the fields such as hair and facial hair from numerical data (between 0 and 1) to Boolean data, which was accomplished by finding a cut off point, and stating that anything below it was false, and anything above it was true. This is shown below in figure X:



Having the data stored in this way fed into step four, as the data could be compared quite easily once it was in the same format. Figure X below shows the format of the user profiles being targeted by an advertising unit, which shows several features. The first line is the name of the advertising unit. Next is each of the characteristics, in the order of age, gender, facial hair, glasses, baldness, eye makeup, lip makeup and hats. If the field is given a response (such as the age range or gender), this means the advertiser wants to target based on that factor. If the field is left null, it means the advertiser isn’t interested in that field, and would like to advertise to them either way. As an example, using figure X, Gillette will only advertise to an adult male who is seen to have facial hair. It isn’t concerned with anything else, so will advertise to passers-by regardless of whether they have glasses on, for example. Once an advertising unit is seen as a match to a passer-by, it is added to a list of potential advertisements, so that one can be chosen to display.



To choose an advertisement to add to the list of potential recommendations, the diagram in figure X shows the logic of the system, and how each option is chosen in the case of the mock-up advertisers. In the live system, this would be more complex, with many more options, and potentially multiple choices at the end of each tree. However, for the purposes of illustrating this system, the green circles represent an advertiser, and black represents a characteristic. If the diagram can be followed to an advertiser, it is added as a viable option.



The final step in developing the system was to take this list of potential advertisers and choose one to display to the user. This was done by simply choosing one at random, to ensure that over time, a fair distribution of advertisers is achieved, however is an area where further work could be explored.

## Evaluation

The first part of the project to evaluate is the predicted characteristics of the passers-by. This is done first to ensure that this part of the project doesn’t fail, or it would have large, damaging consequences to the accuracy of the entire project. For example, if a passer-by is thought to have glasses when they don’t, Specsavers would be an incorrect recommendation to make, even though the recommendation system itself may be functioning correctly.

To test this part of the system, the outputted characteristics of the passer-by will be compared with the labelled data, and a percentage of accuracy will be given. As this is a purely quantitative test, an analysis of the figures will be given afterwards, to explain any issues that have been brought up.

The accuracy for the age is measured by checking whether the predicted age is within 10 years of the actual figure. This is quite lenient, however the actual system only checks whether the passer-by is a child, adult or elderly person, and so therefore this provides a good measure of how accurate it can be, by forcing it to pass an even more difficult test than it needs to. The overall accuracy figure for age is 85%, which is a strong figure, especially since outliers are usually only around 2 years off the boundary. Furthermore, some of the images in this dataset are actors and actresses, known for looking younger than the average person their age, and so therefore this is a difficult, and perhaps unrepresentative test for the system.

Gender performs perfectly, achieving a 100% success rate. This is somewhat to be expected, as gender is the easiest of the characteristics to identify most of the time, with only a few rare exceptions. This is because there is a very easy to identify difference in facial features, which the machine learning algorithm can sort effectively.

More surprisingly, facial hair has a success rate of 98%. This figure was expected to be closer to the accuracy of the age, due to the amount of variability in length, colour and shape, and the fact that this system was artificially selecting a number which defined whether or not there was in fact a beard, whereas the Facial Recognition API was returning a value between 0 and 1 for chin, sideburns and moustache. Turning this into a binary value is not what was initially intended for the system, so the fact that this didn’t have a larger impact on overall accuracy is a massive success.

Like gender, glasses performed as expected, with 99% accuracy, and only a single mistake found. This was expected to perform well, again due to the consistency of images with glasses. Machine Learning algorithms perform better when a problem is consistent, as all they are really doing is comparing images (and their labels), so since glasses are always in the exact same place, and follow very similar patterns, it would be hard to misidentify them. The image which wasn’t correctly identified contained glasses with a very small, hard to spot frame, and had the face in a shadow, making it understandable for the system to not pick this up.

The accuracy of the system predicting the presence of hair was 96%, however this could’ve been higher. This is because two of the three errors were a result of men wearing hats. This understandably causes confusion, and should probably have been considered beforehand, as it is a problem which needs a workaround to be created, such as always saying that a passer-by wearing a hat has hair, since this is statistically more likely.

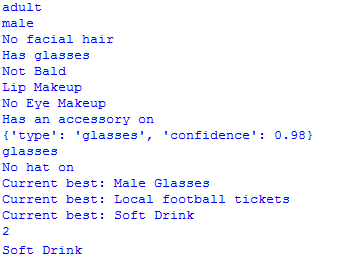
Lip makeup scored a surprisingly high accuracy of 89%, which like age, was expected to be significantly lower. In this case, this is because the natural colour of lips changes on a person to person basis, and therefore it can be hard to say whether they are wearing makeup at all. The same logic applies to eye makeup, which was expected to be similar to lip makeup, but actually dropped to 82%, the lowest of all measured characteristics. This is likely due to the variability of this characteristic, since eye makeup can consist of mascara, eye liner or eye shadow, and these can be in different shades and styles, that look different depending on the who it’s on, especially considering age and skin colour.

Finally, detecting the presence of a hat has an accuracy of 99%. Again, this was as predicted, since hats are usually very easy to distinguish from a face, even from a large distance. In fact, it’s surprising that this isn’t the highest accuracy, however the one instance of the hat being incorrectly located is from an image which recognised the hat as hair. This is likely the result of the image being taken from a strange angle and would be unlikely to appear again.

Overall, the accuracy of the system so far is more than satisfactory, with only a few underperformers. Age underperformed due to reasons that shouldn’t appear in the live system, however makeup could be a point of concern, specifically eye makeup. While 82% would be unlikely to deter advertisers from using a system such as this, moves could be made to increase the effectiveness of the system, such as merging lip and eye makeup to advertise all makeup based on the prevalence of either one of them, based on the assumption that most women that wear one will inevitably wear both.

The next step to take is to measure the effectiveness of the advertisement recommendations. The success of this system is a lot more difficult to measure; is one advertisement better to show than another? What if the passer-by was misidentified? How could the results of the system be shown on a scale of success, or should they at all?

With all these options considered, the choice was made to have a pass or fail system, analysing a set of test cases to ensure that the system was working correctly. It didn’t make sense to have a scale of success, as deciding if one correct advertisement was better than another would be a completely arbitrary task, as their levels of success are subjective. For example, if the system picks up a woman with makeup and glasses, why would a glasses company be a better fit than a makeup company? Either choice would be a suitable fit, and so the system will be evaluated as such.

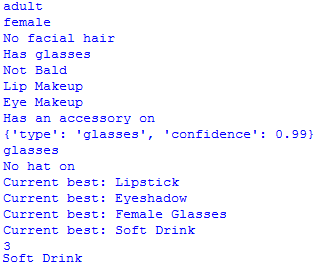
A mix of five faces from the dataset have been chosen to be analysed as part of the system. The images will be provided in appendix 1. They are people of different ages, genders and features, to see if the system picks up on different things. As the system first finds a list of suitable ads, and then outputs the single one it selects at random from that list, all the ‘suitable’ advertisements will be listed, to identify if anything is incorrectly chosen.

Firstly, Image 1 is identified as an adult male, with no facial hair, eye makeup or hat, but is considered to have glasses, hair and (incorrectly) lip makeup, as shown in the complete console output in figure X. The characteristics are now checked against the premade list of advertiser profiles, and the following adverts are collected: Glasses (male), local football tickets and soft drinks.

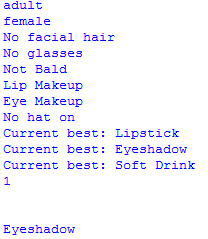
The first thing to do here is to make sure none of the recommended ads are incorrect, by identifying what caused them to be selected. Glasses were identified as a result of his wearing glasses, local football tickets because he is an adult male and a soft drink as this is deemed a universal product, which would want to advertise to everyone, and should appear in each entry. Each of these adverts is justifiable, and therefore has not created any false positives. However, the system could also miss out an advertiser, and therefore the diagram shown earlier in figure X should be followed to try and identify more advertisers. In this case, it does not.

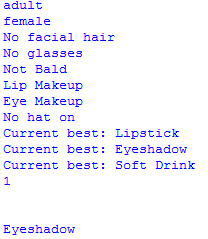
Overall, the system randomly chose to advertise the soft drink.

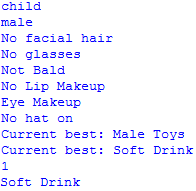
The second image is identified as an adult female with hair, no facial hair or hat and wearing glasses, lip makeup and eye makeup. When compared against the advertiser profiles, the following are outputted: Lipstick, Eyeshadow, glasses (female) and soft drinks. The complete output is shown below, in figure X.

Each of these adverts is based on a correct assumption, such as lipstick, eye shadow and glasses being based on their identification in the image, and the soft drink being an option for everyone.

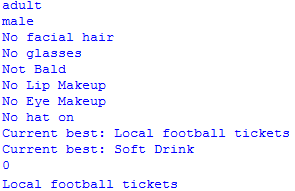
After searching through the diagram, no advertiser has been missed, so the system has correctly identified every potential option, and has passed again. As can be seen in figure X, the system once against randomly chose the soft drink to advertise to the passer-by.

The third image provided a similar set of characteristics as the second image, but without the glasses. This provided the console output shown below in figure X. Predictably, this meant that it selected the exact same advertisements as Image 2, except for the glasses, which are not identified in Image 3. This means that this is also a correct output of potential advertisers.

Similarly, searching through the diagram proves that no advert is missing, meaning that this image is also a success. The advert chosen to be displayed was for eyeshadow.

For the fourth image, a male child was identified, with no facial hair, balding, glasses, lip makeup or hat, however eye makeup was mistakenly identified. This set of characteristics led to two adverts being correctly recommended, Toys (male) and soft drinks. This output is shown in figure X.

To ensure that no advertisement is missed, the diagram can again be followed, which shows that these are indeed the two correct outputs for the system. Overall, the system chose to advertise soft drinks out of the two.



For the final image, a male adult with no facial hair, balding, glasses, makeup or hat was identified. This limited number of features led to two adverts being recommended; local football tickets and soft drinks. Both these adverts are acceptable given the characteristics.

Due to the lack of any features other than age and gender, the diagram clearly illustrates that there are no other potential options that could’ve been explored, making this example also correct. The advert that was chosen to be displayed was local football tickets.

# Findings and Conclusion

Firstly, a deeper insight into the results of the evaluation should be made, to consider the effectiveness of the artefact in meeting the aim and objectives of this project. One such way of doing this is analysing the accuracy of the algorithm which gathers passer-by characteristics. For example, products which advertise only to men or women in static advertisements (i.e. adverts that aren’t targeted) will inevitably be a waste half of the time, as they only relate to 50% of people. However, since gender was accurate 100% of the time, this makes the advertisement twice as cost effective as it was before, as there is less time wasted showing it to people who won’t be interested in it, and in that case advertisers are paying the for space taken by the advert, not the time it is shown.

In the United Kingdom, the percentage of people wearing glasses is around 68% (<https://www.statista.com/statistics/711514/individuals-who-wear-spectacles-in-selected-european-countries/>), so static advertisements are actually quite cost effective as it is in this case, however when it comes to targeting more carefully, such as by gender or age, the effectiveness plummets. For example, 24% of the population in the UK is under 18 (<https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf>), so if around 68% of these people wear glasses, that leaves 16% of the entire population being under 18 and wearing glasses. So, a static advertisement targeting this group would only apply to one in every 6 people that saw it, as opposed to the system outlined in this project, which would essentially be 6 times more cost effective.

While most of the characteristics are similar (with higher levels of targeting providing better results), it is important to analyse the worst as well. Eye makeup only had an accuracy of 82%, meaning around 1 in 5 people shown eye makeup marketing material would be seen as a waste of money. Comparatively, although there are no statistics on the amount of women wearing makeup in public at any one time, it can be assumed from (<https://www.statista.com/statistics/713178/makeup-use-frequency-by-age/>) that this figure is around 50%, which would mean a static advertisement for women’s makeup would target around 25% of people, which translates to being cost ineffective 3 out of 4 times, a much larger figure than the targeted system. This means that even at it’s worst, this system still helps companies save a significant amount of their marketing budget, by reducing waste and helping them show the adverts to those who are more likely to go out and buy the product or service being shown. Therefore, based on this information, the artefact should be considered generally successful at meeting the aim of the project.

However, certain objectives were missed out from the project, which significantly reduced its success. More here.

Another important consideration to make when analysing the success of this project is ethics. There has been a growing concern over the last few years of surveillance, in particular the growing use of facial recognition techniques in public places. While this primarily applies to its use in security, this crosses over to commercial use as well, due to an increased level of care regarding personal privacy from the general public. Seeing how data is being used without consent, such as during the Cambridge Analytica scandal (<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>) and Googles collection of open Wi-Fi data when collecting street view data (<https://www.bbc.co.uk/news/technology-24047235>) has caused a knee jerk response to any technology that can potentially misuse personal information. This means that, if this system were to be implemented, several steps would need to be made in order to maintain a level of trust. The most important step would to be transparent with the public about what is collected, and how it is used, in as simple terms as possible. Everything done in the scope of this project is done as ethically as possible, and so a project of a similar nature should trust this, and be as open as possible, since the most common source of failure is when it is found out that something is being identified which people weren’t made aware of beforehand.

Another consideration to make is what identified information should be used. For example, the Microsoft Facial Recognition API used in this project can also return emotion, and it wouldn’t be too hard to identify race or weight, all of which could be severely misused and should be avoided at all costs.

Finally, the captured data should not be stored, sold or reused in any way. This is a common practice with online companies, with the phrase “If you’re not paying for it, you’re the product” entering the public conversation from its widespread usage. And although it’s lucrative business, it’s hard to get the targets consent for the reuse of data, and is therefore considered unethical, and a breach of trust.

There are also some ethical issues that this project has had to tackle. One of these is the necessary use of stereotyping when creating advertiser profiles. An example of this is recommending football tickets to adult males only. While statistically this is a logical decision, it encourages the idea that going to football matches is only for men and could enforce these ideas that society has been trying to tackle (McDowell and Schaffner, 2011). This goes further when children are considered, who are considered more impressionable to media, including advertisements (Borzekowski and Robinson, 2001). This is explored further in (Pike and Jennings, 2005), which shows that the toys advertised to children can indeed have an impact on what they deem appropriate for their gender. This can be damaging for the individual, and for society in general (Bem, 1981).

Need for better age detection at younger age – diff between 12 and 16

Changing the advert based on characteristics

# Reflective analysis

# References

UNORDERED:

<https://www.theguardian.com/technology/2019/feb/24/are-you-being-scanned-how-facial-recognition-technology-follows-you-even-as-you-shop> – article about face identify in shopping mall

<https://www.theguardian.com/music/2018/dec/13/taylor-swift-facial-recognition-technology-surveillance> – Taylor swift using it at shows

ORDERED:

Aksu, H., Babun, L., Conti, M., Tolomei, G. and Uluagac, A.S., 2018. Advertising in the iot era: Vision and challenges. *IEEE Communications Magazine*, *56*(11), pp.138-144. Available from <https://arxiv.org/pdf/1802.04102.pdf> [accessed 3 April 2019].

Alt, F., Müller, J. and Schmidt, A., 2012. Advertising on public display networks. *Computer*, *45*(5), pp.50-56. Available from <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6193076&tag=1> [accessed 3 April 2019].

Bem, S.L., 1981. Gender schema theory: A cognitive account of sex typing. *Psychological review*, *88*(4), p.354. Available from <https://ahcaf.com/wp-content/uploads/2015/07/gender_schema_theory.pdf> [accessed 3 June 2019].

Borzekowski, D.L. and Robinson, T.N., 2001. The 30-second effect: an experiment revealing the impact of television commercials on food preferences of preschoolers. *Journal of the American Dietetic Association*, *101*(1), pp.42-46. Available from <https://www.researchgate.net/profile/Dina_Borzekowski/publication/246144531_The_30Second_Effect/links/5a4e72690f7e9bbfacfc2da4/The-30Second-Effect.pdf> [accessed 3 June 2019].

Exeler, J., Buzeck, M. and Müller, J., 2009, October. eMir: digital signs that react to audience emotion. In *2nd workshop on pervasive advertising* (pp. 38-44). Available from <https://pdfs.semanticscholar.org/4acc/d415ef99e73fb2ecd4534b7fe5f5793761af.pdf?_ga=2.88568219.1018234308.1552315262-1902740231.1552315262> [accessed 3 April 2019].

Farahat, A. and Bailey, M.C., 2012, April. How effective is targeted advertising?. In *Proceedings of the 21st international conference on World Wide Web* (pp. 111-120). ACM. Available from <https://www2012.universite-lyon.fr/proceedings/proceedings/p111.pdf> [accessed 3 April 2019].

Geng, X., Zhou, Z.H. and Smith-Miles, K., 2007. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *29*(12), pp.2234-2240. Available from <http://dro.deakin.edu.au/eserv/DU%3A30007652/geng-automaticage-2007.pdf> [accessed 3 April 2019].

Guo, G., Dyer, C.R., Fu, Y. and Huang, T.S., 2009, September. Is gender recognition affected by age?. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops* (pp. 2032-2039). IEEE. Available from <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5457531&tag=1> [accessed 3 April 2019].

Guo, G., Mu, G., Fu, Y. and Huang, T.S., 2009, June. Human age estimation using bio-inspired features. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 112-119). IEEE. Available from <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5206681> [accessed 4 April 2019].

Han, H., Otto, C., Liu, X. and Jain, A.K., 2014. Demographic estimation from face images: Human vs. machine performance. *IEEE transactions on pattern analysis and machine intelligence*, *37*(6), pp.1148-1161. Available from <http://vipl.ict.ac.cn/uploadfile/upload/2017020711105957.pdf> [accessed 4 April 2019].

Huerta, I., Fernández, C. and Prati, A., 2014, September. Facial age estimation through the fusion of texture and local appearance descriptors. In *European conference on computer vision* (pp. 667-681). Springer, Cham. Available from <http://hertasecurity.com/sites/default/files/publication/files/ECCVW14_CR_send.pdf> [accessed 4 April 2019].

Kalam, S. and Guttikonda, G., 2014. Gender classification using geometric facial features. *International Journal of Computer Applications*, *85*(7). Available from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.429.3406&rep=rep1&type=pdf> [accessed 7 April 2019].

Li, B., Lian, X.C. and Lu, B.L., 2012. Gender classification by combining clothing, hair and facial component classifiers. *Neurocomputing*, *76*(1), pp.18-27. Available from <https://www.sciencedirect.com/science/article/pii/S0925231211004589> [accessed 7 April 2019].

Lian, H.C. and Lu, B.L., 2006, May. Multi-view gender classification using local binary patterns and support vector machines. In *International Symposium on Neural Networks*(pp. 202-209). Springer, Berlin, Heidelberg. Available from <https://www.researchgate.net/profile/Bao-Liang_Lu2/publication/5663358_Multi-view_Gender_Classification_Using_Local_Binary_Patterns_and_Support_Vector_Machines/links/540289350cf2c48563af8b7b.pdf> [accessed 7 April 2019].

Lyons, D., Pelletier, D. and Knapp, D., 1998, November. Multimodal interactive advertising. In *Proc. Wkshp. Percept. User Interfaces* (pp. 83-86). Available from <https://www.researchgate.net/profile/Damian_Lyons/publication/255564024_Multimodal_Interactive_Advertising/links/53ce6de80cf279d93530a103/Multimodal-Interactive-Advertising.pdf> [accessed 7 April 2019].

McDowell, J. and Schaffner, S., 2011. Football, it’s a man’s game: Insult and gendered discourse in The Gender Bowl. *Discourse & Society*, *22*(5), pp.547-564. Available from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.839.5846&rep=rep1&type=pdf> [accessed 2 June 2019].

Moghaddam, B. and Yang, M.H., 2000, March. Gender classification with support vector machines. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)* (pp. 306-311). IEEE. Available from <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=840651> [accessed 7 April 2019].

Ng, C.B., Tay, Y.H. and Goi, B.M., 2012, September. Recognizing human gender in computer vision: a survey. In *Pacific Rim International Conference on Artificial Intelligence*(pp. 335-346). Springer, Berlin, Heidelberg. Available from <https://www.researchgate.net/profile/Yong_Haur_Tay/publication/262152447_Recognizing_Human_Gender_in_Computer_Vision_A_Survey/links/55fe351a08aec948c4e1feec.pdf> [accessed 7 April 2019].

Ojala, T., Pietikäinen, M. and Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7), pp.971-987. Available from <http://vision.stanford.edu/teaching/cs231b_spring1415/papers/lbp.pdf> [accessed 7 April 2019].

Pike, J.J. and Jennings, N.A., 2005. The effects of commercials on children’s perceptions of gender appropriate toy use. *Sex roles*, *52*(1-2), pp.83-91. Available from <https://www.researchgate.net/profile/Nancy_Jennings/publication/30846400_The_Effects_of_Commercials_on_Children's_Perceptions_of_Gender_Appropriate_Toy_Use/links/53d291fb0cf2a7fbb2e9a76d/The-Effects-of-Commercials-on-Childrens-Perceptions-of-Gender-Appropriate-Toy-Use.pdf> [accessed 3 June 2019].

Shakhnarovich, G., Viola, P.A. and Moghaddam, B., 2002, May. A unified learning framework for real time face detection and classification. In *Proceedings of Fifth IEEE international conference on automatic face gesture recognition* (pp. 16-23). IEEE. Available from <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1004124> [accessed 8 April 2019].

Tian, P., Sanjay, A.V., Chiranjeevi, K. and Malik, S.M., 2012, August. Intelligent advertising framework for digital signage. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1532-1535). ACM. Available from <http://wan.poly.edu/KDD2012/docs/p1532.pdf> [accessed 8 April 2019].

Tucker, C.E., 2014. Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research*, *51*(5), pp.546-562. Available from <https://journals.sagepub.com/doi/abs/10.1509/jmr.10.0355> [accessed 8 April 2019].

Yang, Z. and Ai, H., 2007, August. Demographic classification with local binary patterns. In *International Conference on Biometrics* (pp. 464-473). Springer, Berlin, Heidelberg. Available from <https://link.springer.com/content/pdf/10.1007/978-3-540-74549-5_49.pdf> [accessed 13 April 2019].

Zhang, Y., Liu, L. and Li, C., 2017. Quantifying facial age by posterior of age comparisons. *arXiv preprint arXiv:1708.09687*. Available from <https://arxiv.org/pdf/1708.09687.pdf> [accessed 13 April 2019].

# Appendix