# Identifying characteristics of passer-by's 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

# 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 first 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. The second aim of this project is to implement and evaluate the current state-of-the-art techniques used in Computer Vision to analyse the characteristics and relationships of people in public. To achieve these aims, 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

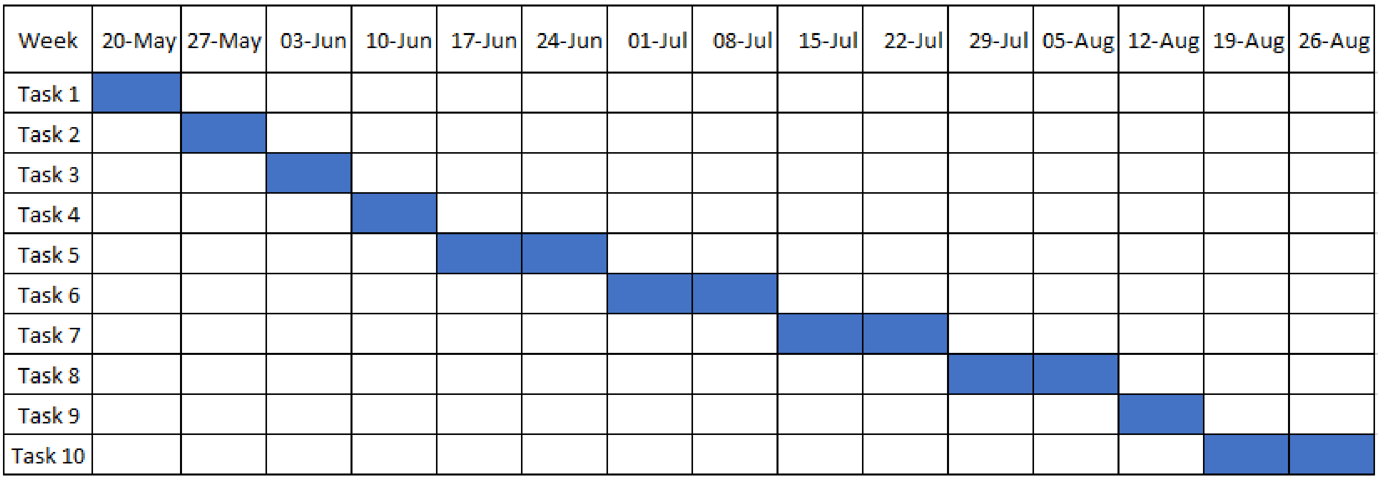
### Gantt chart

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.

| Week | 20-May | 27-May | 03-Jun | 10-Jun | 17-Jun | 24-Jun | 01-Jul | 08-Jul | 15-Jul | 22-Jul | 29-Jul | 05-Aug | 12-Aug | 19-Aug | 26-Aug |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Task 1 | ■ | ■ | | | | | | | | | | | | | |
| Task 2 | | | ■ | | | | | | | | | | | | |
| Task 3 | | | | ■ | | | | | | | | | | | |
| Task 4 | | | | | ■ | ■ | | | | | | | | | |
| Task 5 | | | | | | | ■ | | | | | | | | |
| Task 6 | | | | | | | | ■ | ■ | ■ | ■ | | | | |
| Task 7 | | | | | | | | | | | | ■ | ■ | | |
| Task 8 | | | | | | | | | | | | | | ■ | ■ |
| Task 9 | | | | | | | | | | | | | | | |
| Task 10 | | | | | | | | | | | | | | | |

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 4$^{TH}$ 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

Use API to get info

Plug info in as a sort of user profile (content based recommendation)

Have business create item profile (set of features about item, in this case who to target)

Cross reference against them, i.e. pick an ad that applies to that user

Display it

## Development

## Evaluation

Talk about the need to evaluate qualitatively, to find quality of recommendations, not just accuracy, due to not always being 1 answer.

Not always showing same advert for example.

## Findings and Conclusion

Ethics for company – e.g. toys for boys and girls?

## 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].

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].

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].

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