University of Warwick
Department of Computer Science

GlaxoSmithKline
Immersive Intelligent Manufacturing

Kinect PPE Gowning Conformation

Author
Benjamin J. BIGGS

Industrial Sponsor
Mr Patrick HYETT

Project Supervisor
Dr Abhir BHALERAO

CS344 Final Report
Discrete Mathematics, Final Year
May 2016

Abstract

In order to maintain positive relationships with regulatory authorities, GSK must have confidence that all staff members are properly dressed when located in high-risk facilities (e.g. areas in which employees are exposed to raw product). This project aims to address this need by constructing an automatic verification process to determine whether a lab operator has correctly followed the in-place gowning procedure – i.e. the operator has correctly put on the specified safety clothing/equipment in the appropriate order – before allowing access to the facility. If successful, the developed application will allow GSK, and similar companies, to prove their adherence to EU regulations that ensure both staff and product are kept safe.

Keywords

Computer Vision; Machine Learning; Kinect; Gowning Procedures; MATLAB; C#

Acknowledgements

This project would not have been possible without the help and support of the following people:

- Dr Abhir Bhalerao – Project Supervisor (University of Warwick)
- Mr Patrick Hyett – Industrial Sponsor (GSK)
- Dr Theo Damoulas, Lewis Rendell – Technical advice and discussion around machine learning and statistical methods
- Casey-Elizabeth Hayden – Sketched interpretation of technical diagrams, represented in Figure 5, Figure 10, Figure 11, Figure 13, Figure 14, Table 3, Table 6, Table 7 and Table 8
- Philip Shepherd, Mark Allen and the rest of the IIM team
- GSK data collection volunteers, notably Joana Gomes (key test volunteer) and Eddie Dukes (video camera)
- Liam Steadman, Nicholas Biggs and Anthony ‘Tony’ Potts – General advice and discussion
- Jacqueline Biggs, Emma French, and Stephen Attwood

Contents

List of Figures

List of Tables

1 Introduction and Motivation

1.1 Motivation

1.1.1 Pharmaceutical Regulations

This project is motivated by a genuine business case identified in the pharmaceutical industry. In order to maintain positive relationships with regulatory authorities, GlaxoSmithKline (GSK) must have confidence that all staff members are properly dressed before they enter high-risk facilities (i.e. areas where employees are exposed to raw product). The company's obligations fall into two categories: the health and safety requirements that protect the area operators, and those that govern the manufacture of human-consumable products.

Among many other applications, the personal protective equipment (PPE) used at GSK is purposed to shield the staff from chemical or metal splash, prevent facial impact injuries and entanglement from own clothing. The equipment also serves to prevent cross-contamination – the unintentional transfer of bacteria – that could otherwise result from fallen hair follicles (from scalp or face), skin abrasions or external substances attached to the clothes or skin. Should a pharmaceutical company fail to control these risks, punishments can be severe and even result in a revocation of the company's manufacturing licence.

In the United Kingdom, the European Union laws that govern PPE requirements (SI 1994 / 2326, 1994) are enforced by the Medicines and Healthcare Products Regulatory Agency (MHRA), an executive agency of the Department of Health. Although the international rules are similar, globally-operating companies such as GSK may be answerable to multiple regulatory bodies. GSK choose to resolve the differences between various dress regulations by enforcing their own standard operating procedures (SOP), informally referred to as *gowning procedures*, that set the most rigorous standard over the set of global requirements. The documents associated with these gowning procedures describe each item of PPE and the order in which they should be put on, or *gowned*. It should be noted that in every document of this type the opening paragraphs state that: '*It is the responsibility of: (1) GSK staff to comply with the requirements of this procedure […] and follow the gowning procedure outlined in this document. (2) GSK […] are responsible for ensuring that this SOP is followed.*'

In short, these two bullet points describing the obligations of the company and its staff provide the entire motivation for this project. To aid in the company's quest to satisfy their acknowledged obligations, this project aims to design a process for GSK (and similar companies) to ensure that the gowning procedures used for controlling dress in their internal high-risk facilities are being correctly followed. This project's primary objective is to develop an automated procedure that is able to verify that operators have gowned the specified PPE in the correct order before granting access to their facility. Should this project yield a successful outcome, the application will serve to benefit relationships between its adoptive companies and their regulators, by providing them with a mechanism to prove that their mandatory gowning procedures are being followed.

1.1.2 GSK Gowning Procedure (SOP-PDK-0012)

Since hazards vary across the many high-risk facilities at GSK, gowning requirements are set on a case-by-case basis. To narrow the initial scope, this project focuses on the gowning procedure set for entry into the dispensing, manufacturing, packaging and warehousing areas currently in force at seven high-risk facilities across two major GSK UK sites, in Ware and Harlow.

Apart from the convenience resulting from their close proximity to the research institution, the facilities in these two locations were chosen due to the fact that similar gowning procedures, detailed in standard operating procedure SOP-PDK-0012 are in force. The procedures detailed in this document (GlaxoSmithKline, 2014) are largely identical, as the same item types are set to be gowned in the same order, although some discrepancies occur due to different suppliers being used in the various facilities. As an example, although each of the procedures require the gowning of nitrile gloves at the same stage of the sequence, these may be of different colour or have different markings.

To avoid any potential ambiguity, this project first considered the procedure for entry to GMP Inhaled Powder Manufacturing and Dispensary Suites within Building 5, Ware described in Section 7.2 of the aforementioned document. This project only considers the procedure's 'gowning instructions', i.e. those indicating that a user should put on a PPE item, rather than other instructions such as applying hand gel or collecting coveralls from the store.

Table 1 shows the PPE items and the order in which they should be gowned to comply with SOP-PDK-0012:

1. Mob cap	2. Beard snood *(if required – see next section)*	3. Overshoes
		
4. Coveralls	5. Gloves	6. Goggles
		

Table 1: SOP-PDK-0012 PPE Items

Figure 1 shows a lab operator after gowning these items in the correct order.

Figure 1: Gowned according to SOP-PDK-0012

1.1.3 Beard Snood

Although the majority of the above items are self-explanatory, note that the procedure only mandates a beard snood if the operator has sufficiently dense facial hair. To provide a quantitative measure of whether an operator has a beard, this project aims to develop an automated process for making this determination, in order to work out whether a beard snood must later be put on.

1.1.4 Alternative Use-Cases

It was noted that, should the developed application offer the capability for a user to train the system to recognise new PPE items (and new orderings), there would be great potential for the tool to have applications across many industrial sectors that are not limited to the identified pharmaceutical use-case. Indeed, such an application would be able to enhance the safety conditions of any work process that requires staff to dress accurately in a short time. Among others, the project could be used by firefighters, bomb squads and hospital staff, whose lives may depend whether they are correctly wearing their specified equipment.

Figure 2: Firefighter in PPE (Fire Product Search, 2016)

The application could also have a commercial use as a novel checkout system for high-street clothing shops. Using the application's ability to recognise clothing, a customer could simply wear their chosen garments, present themselves to the application and then leave the store. The price of the identified garments could then be totalled and charged to the purchasing customer's bank account.

Finally, the automatic determination of the presence and density of a user's facial hair could lead to the development of an application to support a marketing strategy for barbers, hair salons or even the Movember charity (Movember Foundation, 2016). Members of the public could be asked to step up to the application and have their facial hair evaluated before and after a complementary makeover for comedic effect.

1.2 Project Aims

This project aims to develop an application that can recognise specific pieces of PPE and the order in which a user puts them on to validate a full gowning procedure. To enable alternative use-cases for other GSK facilities and external organisations, the application should feature a training module that allows an unspecialised user to define new gowning procedures by introducing new pieces of PPE and setting new orderings on them. For gowning procedures that require some users to wear a beard snood, such as the one described in SOP-PDK-0012, the application should contain an integrated beard detection module that determines if the operator has sufficiently dense facial hair to warrant its necessity. Through an industry-provided API, the application should be able to restrict access to the high-risk facility until such time as adherence to the gowning procedure has been completely verified.

Since any developed procedure will be used inside a 'dirty zone' – an area siloed to reduce the risk of staff and external equipment introducing contaminants – no contact can be allowed between human operators and hardware. Consequently, the application must employ alternative human-computer interaction (HCI) methods than the standard mouse and keyboard set-up, to prevent possible cross-contamination. In addition, the application must be suitably designed for use on a large, high-definition and portrait-oriented display that will remain wall-mounted in the Building 5 changing room. The application should also provide the user with continuous feedback that indicates which item is next to be gowned, and display an error if the specified gowning procedure is violated in any way.

Due to the highly-regulated nature of the intended industrial customers, the application must prove to be extremely accurate at verifying if a gowning procedure has been successfully followed, and almost never allow access to an incorrectly-gowned person. In companies such as GSK, computer software is treated in the same way as other laboratory equipment and must pass a rigorous validation procedure before production use is permitted. To ease the natural concerns that validation teams may have over such a novel product, it is essential that the application is robustly tested on a large dataset that adequately represents the general use-case.

Although this project benefits from having an industrial partner, substantial effort should be made to restrict the cost of the project's endeavour but, more importantly, limit the overhead cost of implementing the final application. The developed process should be easy to use and should not substantially increase the time or effort required for an operator to gain access to their high-risk facility.

1.3 Stakeholders

- GlaxoSmithKline – Immersive Intelligent Manufacturing (IIM) Team
    - Due to the genuine business use-case for such an application, GSK's IIM team were targeted as an industrial partner. As well as being able to provide access to a replica laboratory, the company has shown willingness to evaluate the developed application with their industry-standard software validation process.
- Dr Abhir Bhalerao, Project Supervisor
    - With substantial computer vision experience, the supervisor for this project was chosen as being well-suited to perform an advisory and mentoring role during the course of the research and development phases.
- Benjamin Biggs, Author/Researcher/Developer
    - As the research and implementation work carried out for this project has a direct consequence to the author's academic and professional career, he was identified as a key stakeholder. The named individual takes responsibility for conducting all required investigation, producing a testable application and completing all necessary documentation.

2 System Requirements

The project has been developed with respect to the following formal requirements, which were negotiated with the industrial sponsor and the project supervisor. These discussions also gave rise to the listed priority measures that define the impact that would result should a particular requirement be left unsatisfied.

R1. Application should be able to recognise specific PPE items when worn by a human subject. **(Critical)**

- Application must prove to be extremely accurate at accepting/rejecting users' attempts to perform the relevant gowning procedure.
- Application should restrict access to the high-risk facility until the operator has correctly followed the specified gowning procedure.

R2. Application should be able to verify the order in which the PPE items were gowned. **(Critical)**

R3. Application should include a training module that allows a user to teach the system to recognise new gowning procedures. **(High)**

- This will involve providing the user with a mechanism for teaching the application to recognise new PPE items and define new orderings on known PPE items.
- The training module should be suitably designed for an unspecialised user.

R4. Application should include a mechanism for identifying if a user has sufficiently dense facial hair to deem the beard snood necessary. **(High)**

- This will be used when validating procedures that include a beard snood. If a user does not have a beard, the beard snood should not be required.

R5. The recognition part of the application should not require human contact in its general operating mode. **(High)**

R6. Recognition module must be suitably designed for large, high-resolution display in a portrait orientation. **(Medium)**

R7. Recognition module must provide the user with continuous feedback describing the next item to be gowned or reporting an error in the existing sequence. **(Medium)**

R8. Recognition module must be responsive and maintain a 'live' passive feel. **(Medium)**

R9. Recognition module must be extremely simple to use. **(High)**

R10. Recognition module must not significantly increase the time or effort required by an operator attempting to access the lab. **(High)**

R11. The project should remain within a reasonable budget. **(High)**

R12. Application must not impose a significant overhead cost to an adoptive company. **(High)**

3 Potential Solution Designs

The previously-listed requirements can be split into two sets – those that involve the training of new gowning procedures, and those that involve the process of verifying that a specific procedure has been correctly followed. From this point on, these will be referred to as *training* and *recognition* processes respectively and must be fulfilled by any solution strategy.

3.1 RFID Tagging

3.1.1 Explanation

Although primarily used for tracking assets, RFID tags were considered as a method for enabling the identification of PPE. In the training phase, tags could be attached to PPE items that should later be recognised, and encoded such that items of the same type, e.g. the specific purple gloves, exhibit a common identifier that is unique to that type. During recognition, a network of RFID readers could then have been used to continuously scan individual gowning cubicles within the changing room. By designing software that observes the reader network, an event could then have been raised when an encoded tag is introduced to the space, as this indicates the presence of a new PPE item. By analysing the sequence of these introductions, the software could then have inferred which clothing items the user gowned and in which order. Should this sequence conform to the specified gowning procedure, the facility door would then have been made accessible to the correctly-gowned user.

3.1.2 Hardware Options

To implement this system, suitable RFID tags would have been necessary for storing uniquely identifiable values (e.g. GUIDs) for each PPE item type. As there was no requirement for RFID tags to initiate communication with the reader, nor to have an exceptionally large read range, 'active' RFID tags would not have been necessary, so cheaper 'passive' tags would have sufficed. To perform the recognition, a network of RFID readers would have been needed to adequately cover each individual cubicle, reporting which experienced the introduced item.

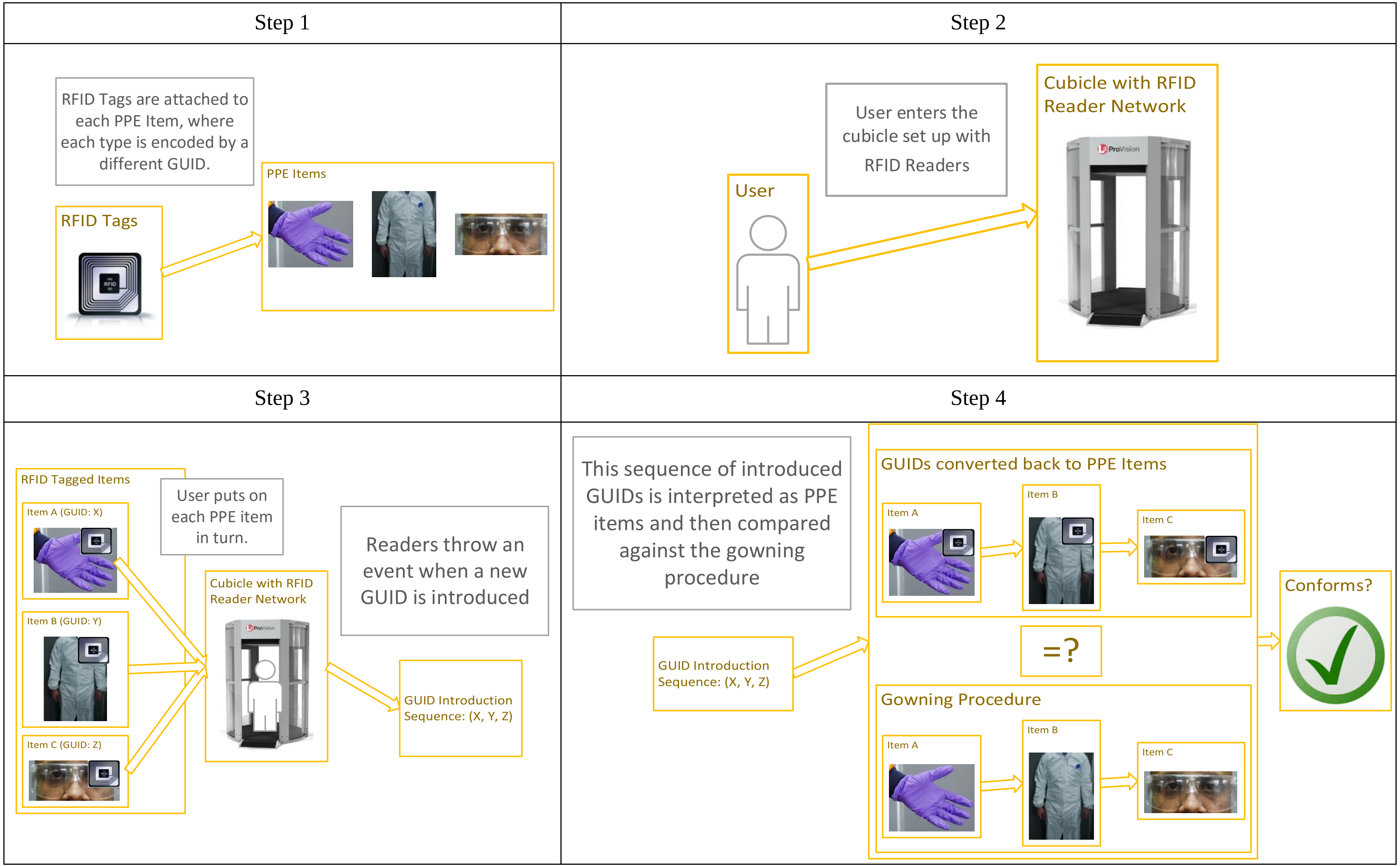


Figure 3: RFID Solution Diagram

3.1.3 Evaluation

This proposal faced a major problem due to the fact that the majority of PPE used across the target industries, including all SOP-PDK-0012 items except the goggles, are disposable. To implement this tagging solution, each adoptive institution would have required their PPE suppliers to attach pre-encoded RFID tags to each item as part of their manufacturing process. Although individual tags can cost as little as $0.07 (RFID Journal, 2016) – approximately 3.7 pence – considering two are required for a pair of nitrile gloves, ordered by GSK at 5.3 pence per pair, the manufacturer could reasonably increase their price by 140% to cover the additional manufacturing costs. Even without purchasing the devices needed to construct an RFID reader network, which incidentally would have consumed the entirety of this project's research budget, the huge overhead cost of customising the existing PPE would be extremely unattractive to potential adopters of this system.

A more trivial fall-back of the proposed RFID system was the absence of any mechanism to perform beard detection. Having users self-acknowledge their own facial hair would have passed a professional validation check, as it is no worse than the current system. However, the new process would have still relied on users making an honest assessment of whether they should wear the beard snood, which may be influenced by the uncomfortable nature of the item.

Aside from this, the RFID system would only have been able to detect PPE items as each was introduced to the cubicle, and would have had no mechanism for determining if items were gowned correctly, or even gowned at all. As a result, such a system could have been easily fooled by a user retrieving PPE items in the correct order without necessarily gowning them. As the reader network would have identified each necessary PPE item in the correct order, the system would release the facility door to allow entry to a user who had not correctly followed the specified gowning procedure. This issue could have been partially resolved by triangulating RFID tag readings to gauge the 3D position of each PPE item and enforcing rules that ensure, for example, that the hat is recorded at a higher position than a glove. However, it is unlikely that such a system would have been sufficiently accurate to determine the difference between a user wearing a glove or simply holding it up with the correct hand.

Despite these issues, the proposed RFID-based system would likely have been highly accurate at identifying tagged PPE items as they were introduced to the scene, as evidenced by their widespread use in asset tracking, personal identification and door-locking systems. The question as to whether a constructed RFID reader network would have occasionally missed the introduction of a correctly-gowned item is difficult to determine without experimentation, but sufficient recognition accuracy is likely to have been obtainable by strategically placing enough readers in each cubicle.

In summary, despite offering higher recognition accuracy on 'taggable' items, employing an RFID strategy would have offered no mechanism for beard detection, been expensive to adopt, and been simple to cheat by a malicious user. Due to these faults, it was seen as highly likely that any application developed according to this method would have failed a formal validation procedure and therefore been unsuitable for this project's intended professional context.

3.2 Computer Vision

Having considered the faults with the RFID-based system, a computer vision approach was conceived and eventually selected as the basis for the produced application.

3.2.1 Explanation

The primary objective for any computer vision solution would be to identify the specified PPE items through visual means, by applying object identification techniques on live video frame data. Any application of this type should exhibit a user-assisted training module that constructs an independent classifier for each PPE item by analysing images of users' body parts in gowned (wearing the item) and ungowned (not wearing the item) states. A well-trained classifier should be able to learn from this training set to allow future classification of similar but unseen images.

For the recognition module, a video camera device should be suitably positioned to observe an entire gowning procedure performed by a user inside a well-lit cubicle with a plain background. This hardware should expose an API to allow custom-built software to process individual frames from this feed. After identifying a user within the cubicle, each PPE item classifier should then be applied in turn to the relevant body region.

A 'beard detector' used to determine whether a user has sufficiently dense facial hair could also be constructed by building a classifier through a similar training process. This specific classifier could then be applied for procedures that involve the detection of a beard snood item, to evaluate whether the item is necessary.

To develop the best possible classifier, the supplied training images and input video frame data should first be cropped to only display the relevant body region for each PPE item. The reasoning for this is that many PPE items are small in comparison to the entire human body, meaning both training and recognition modules may be influenced by unintended differences occurring elsewhere in the frame. These may include differing skin tones and shirt colours.

3.2.2 Hardware Consideration

To ensure the application is suitable for users without specific computer vision expertise, the training module should be able to construct item classifiers from standard RGB images, such as those captured with smartphone cameras. In this case, the pre-processing step that crops images to only contain the relevant body region can be performed manually, either by zooming in before capture or editing the produced images.

3.2.2.1 Standard Video Camera

Perhaps the simplest mechanism for observing a live gowning procedure is to use a standard USB webcam, which have improved dramatically over the last decade. Many feasible options were identified, including some boasting 1080p capability for less than £100 (Microsoft Corporation, 2016g). Should further image quality improvements be required, attention could turn to procuring a Light Field video camera device (Lytro, 2015), which uses a novel technique to combine images captured from several vantage points to produce extremely sharp frames.

By using a standard video camera, there may have been unwanted complications in performing the desirable pre-processing step, i.e. cropping images to only contain the relevant human body region. Although a technique for human joint mapping could have been developed in an open-source image processing package such as OpenCV, this work would likely have been complex and not entirely relevant to this project's primary objectives.

3.2.2.2 Microsoft Kinect

Although primarily used for motion-controlled video gaming, the multi-sensory Microsoft Kinect for Xbox One device provides a mechanism for mapping human body and facial points to perform the aforementioned pre-processing step. Alongside its 1080p video camera and infrared sensor, the Kinect features a high-fidelity depth sensor that provides accurate 25-point skeletal mapping (Pterneas, 2014b), five-point basic facial mapping (Pterneas, 2014a) and over 1000-point facial mapping (Pterneas, 2015). The figures within Table 2 show the position of these skeletal and face points mapped by the Kinect.

By tracking the human skeleton, the device can also interpret a user's body movements, referred to as *gestures*, to provide contactless motion control. As well as providing a mechanism for performing the pre-processing step, this gesture capability can be harnessed to determine if the user is stood in a suitable pose before recognition takes place. By doing this, it would not be necessary to perform a computationally-expensive prediction process on every frame, or even just those containing the user. Instead, the process need only be run on frames in which the relevant body region is clearly visible to the Kinect Sensor.

After several key enhancements from the initial Xbox 360 version, Microsoft have released a Windows-compatible version of the new device alongside the Kinect for Windows SDK (Microsoft Corporation, 2015b) that allows the development of Kinect-enabled software. After purchasing the device and its adapter, a fully-operational set-up can be constructed for less than £200 (Microsoft Corporation, 2016f). Due to the relatively low cost and its rich API collection, the Kinect has lent itself well to a substantial amount of previous academic research, for example by authors pursuing projects that involve human body tracking (Qian, et al., 2014).

<table>
<tr><th>Skeletal Joint Mapping</th><th>Face Point Mapping</th></tr>
<tr><td>

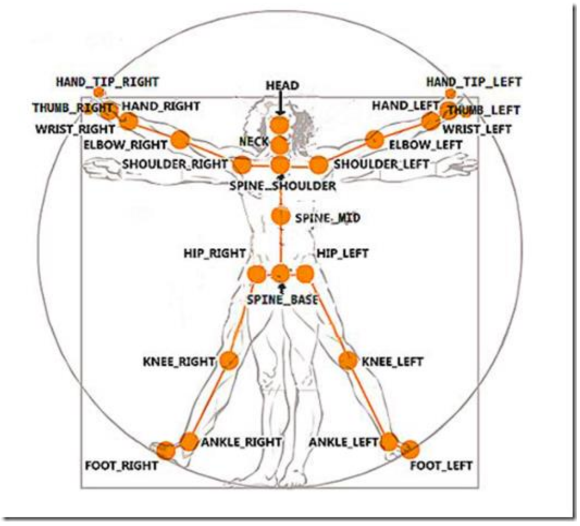


Figure 4: Kinect Skeletal Mapping (Microsoft Corporation, 2016e)

</td><td>

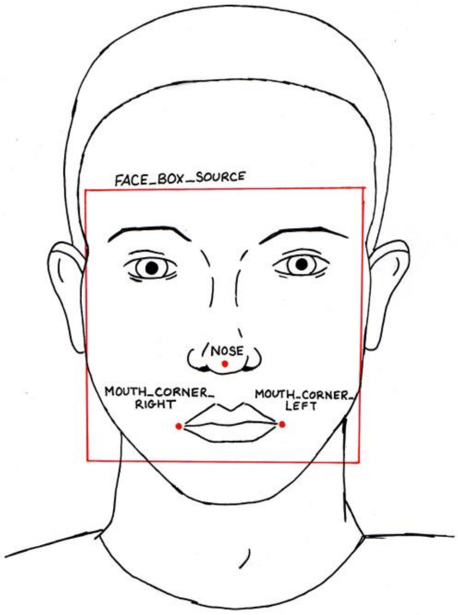


Figure 5: Kinect Face Point Mapping

</td></tr>
<tr><th colspan="2">High Detail Face Point Mapping</th></tr>
<tr><td colspan="2">

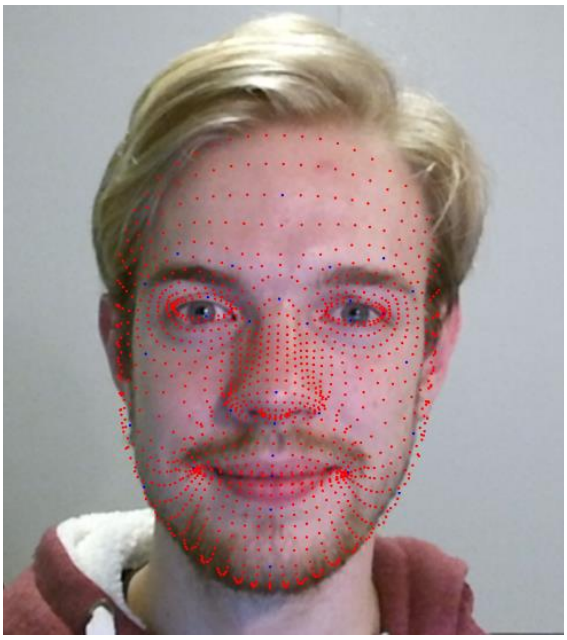


Figure 6: HighDetailFacePoint Mapping

</td></tr>
</table>

Table 2: Kinect Skeletal Joint and Facial Point Mapping

3.2.3 Diagram

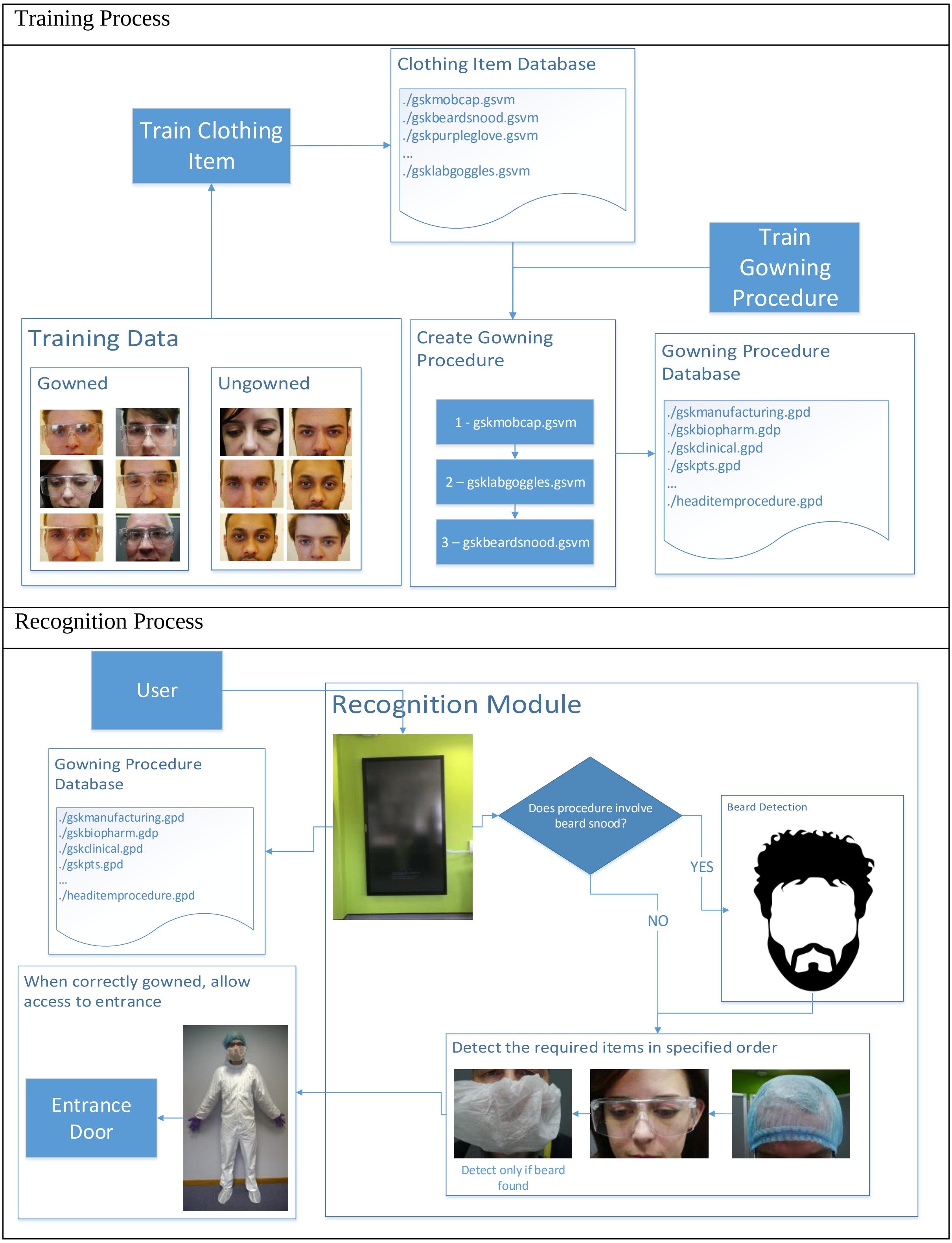


Figure 7: Computer Vision Solution Design

3.2.4 Pros/Cons

It initially appeared that employing a Kinect-based computer vision approach to solving the described gowning problem would fulfil each of the listed requirements. Constructing a vision-based classifier for each detectable item, including the beard, provided a mechanism for validating an entire gowning procedure. Moreover, the Kinect gave a clear direction towards the pre-processing objective that could dramatically improve the reliability and effectiveness of PPE item classifiers. Such a technique would also be difficult to cheat as item classifiers would be trained on users who have correctly gowned each item.

However, a major concern of this strategy was whether such a generic training scheme could produce accurate classifiers for any conceivable PPE item, a problem which appears to be as yet unexplored in the literature.

Firstly, since even the same clothing items can look very different when worn by different people, such a scheme cannot rely on PPE items having a consistent outline shape. Moreover, care should be taken if dependency is placed on colour information, as alongside variation under shadows and different lighting conditions, extra difficulties may arise when classifying transparent items, such as the GSK lab goggles.

Secondly, even by restricting the types of PPE to those described in SOP-PDK-0012, it appears that a number of items exhibit few features that may otherwise aid attempts towards accurate detection. In particular, the lack of logos, barcodes or other detectable markers on the majority of these items may render them challenging to detect in a one-size-fits-all classification scheme.

At least by first intent, there was no need to impose any restrictions on which PPE items would be supported by the system, significantly reducing any overhead cost that otherwise would have resulted from PPE modification. However, if formal testing were to yield systematically unreliable results for particular PPE items, evidence-based recommendations should be made to the industrial customers against their further use.

Finally, despite the many benefits offered by the Kinect, the device’s video camera is weak in comparison to dedicated alternatives on the market. Although it was possible that the modest 1080p resolution could prove to be a drawback when training classifiers on colour frame data, it was known that sharper images could be obtained in other ways. These could include requesting that the user steps closer to the Sensor or, if budget were extended, setting up a multi-device solution that combines the Kinect’s tracking capabilities with a dedicated high-resolution video camera.

4 Solution Architecture and Development Strategy

This project concerns the research and development of an automated computer vision system to address the problem of verifying user-specified gowning procedures enforced in high-risk facilities. The final implementation has been written in C# .NET 4.5.2 and Windows Presentation Foundation (WPF), owing to the industrial sponsor's support requirements.

4.1 Training Module

The training module was designed to allow an unspecialised user to teach the application to recognise new gowning procedures, possibly involving new PPE items. This has been achieved by developing two schemes: the PPE Item Classifier Training Scheme (PIC training scheme) and the PPE Gowning Procedure Training Scheme (PGP training scheme). The PIC training scheme enables a user to introduce new PPE items, whereas the PGP training scheme allows a user to create new gowning procedures by specifying a new ordering on pre-trained items.

4.1.1 PPE Item Classifier Training Scheme

The PIC training scheme has been suitably designed to run on two types of body region images, captured with a standard RGB camera. The first set should be supplied by the user and contain images that depict the relevant body region wearing the new PPE item, referred to as *gowned* images. The second set should be pre-stored and depict images of the relevant body region not wearing the item, referred to as *ungowned* images. The user should also be asked to provide a unique name for the new item and select what kind of item is being trained from a predefined *PPE clothing list*. The options given in this list should be used by the application to determine which body regions the item covers when worn.

An algorithm should then be run over these gowned and ungowned images to construct a *(PPE) item classifier* that accurately determines the presence of the PPE item when shown a previously unseen image of the relevant body region. The algorithm should also be used in the development phase to generate a *beard classifier* to form the basis of the beard detection routine.

The item classifier and its user-supplied metadata should then be combined and serialised to disk in a suitable format for interpretation by the PGP training scheme, which should subsequently be used to incorporate the new item into a gowning procedure. Figure 8 describes this entire process.

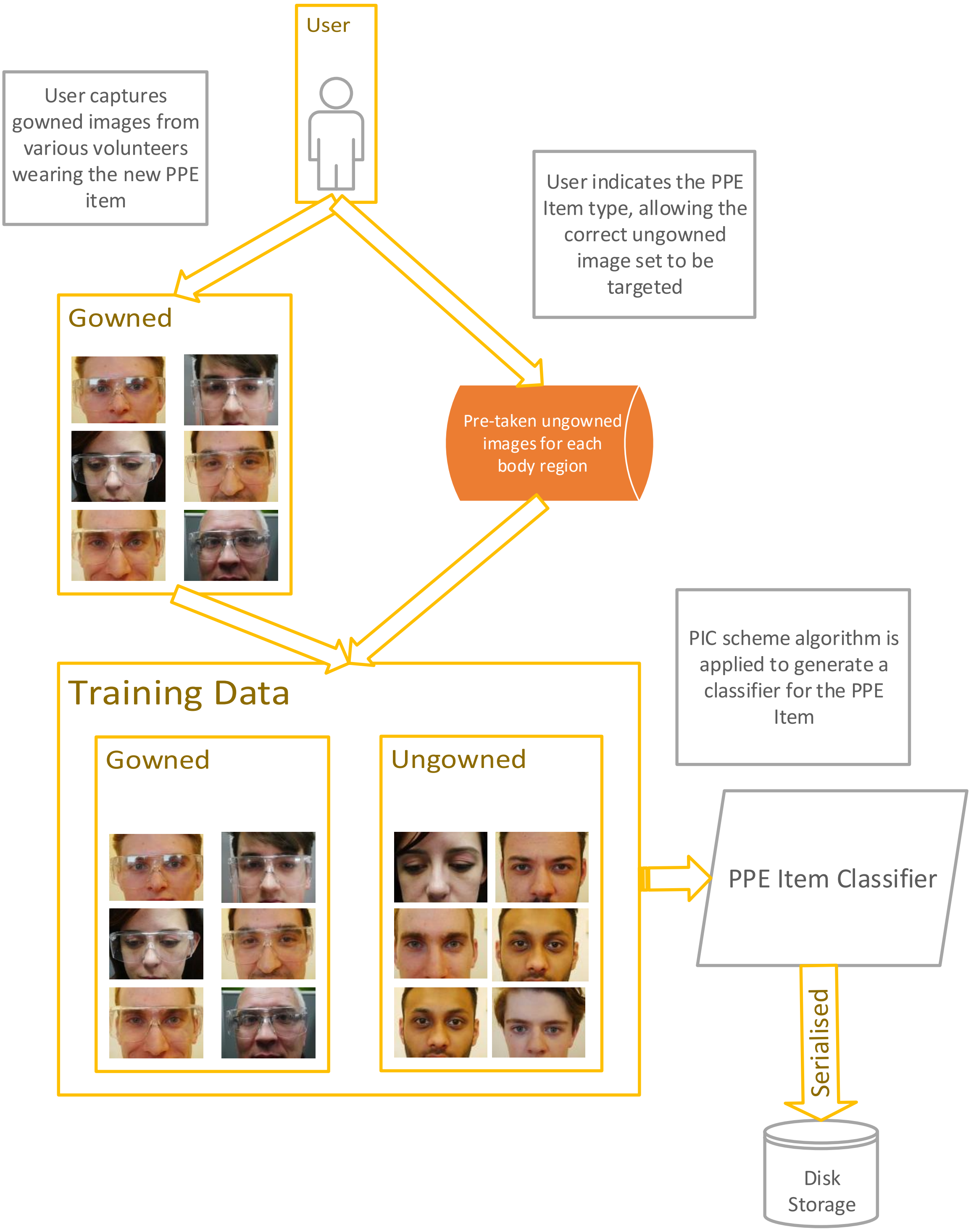


Figure 8: PIC Scheme Diagram

4.1.2 PPE Gowning Procedure Training Scheme

Owing to the fact that some PPE items can be required on multiple body regions – for example, the GSK nitrile gloves are often required on both the left and right hands – it is important to note that gowning procedures actually specify an order on body regions, where each is associated with the item classifier that the recognition module should apply. The PGP training scheme should enable a user to construct a list of PPE items (possibly containing duplicates), although each should be targeted to a unique compatible body region.

The set of ordered item classifiers and relevant metadata, including a unique name for the procedure, should then be combined and serialised to disk in a suitable format for interpretation by the recognition module. Figure 9 describes this entire process.

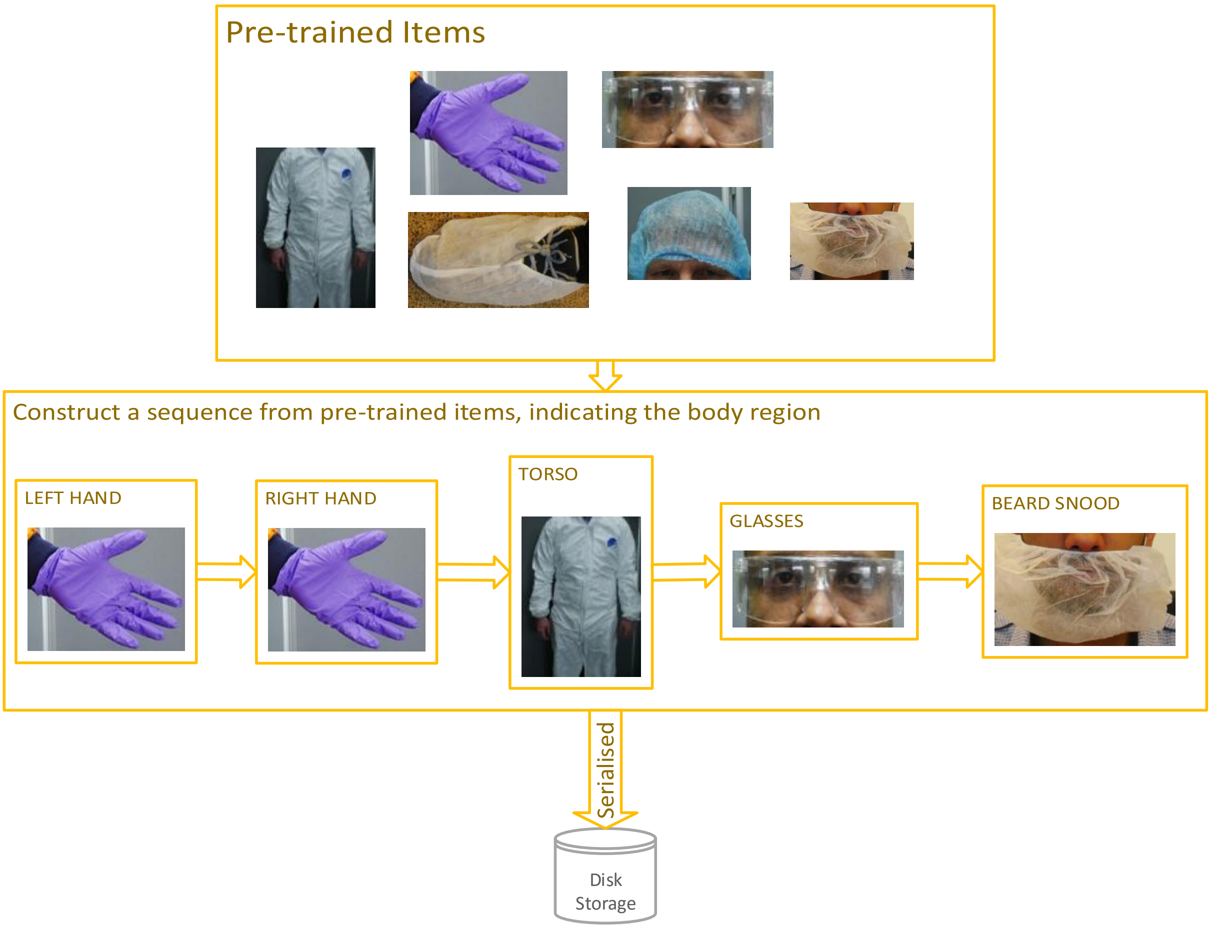


Figure 9: PGP Scheme Diagram

4.2 Recognition Module

The recognition module should instruct and verify gowning processes performed by users who attempt to gain access to their high-risk facility. This can be achieved by deserialising and interpreting the relevant gowning procedure file before running the procedure that checks each required item is correctly gowned by the user in the specified order.

After reading the gowning procedure file, the user interface should be updated to reflect the gowning procedure to be followed by the user. The process should then begin by asking the user to adopt a particular *entrance gesture* to indicate their readiness to start putting on items. If the gowning procedure contains a beard snood item, the application should run its beard detection routine, scanning multiple parts of the user's face to determine whether they have sufficiently dense facial hair to deem the item necessary. If found unnecessary, the beard snood should be removed from this user's gowning procedure.

PPE recognition should then begin by asking the user to adopt a second pose to be verified by the application, the *presentation gesture*, before simultaneously scanning all relevant body regions to ensure that the user begins in a completely ungowned state. As with all PPE detection, Kinect colour frames should be pre-processed to yield tightly-cropped body region images before being passed to the relevant item classifier.

The application should then proceed by checking each body region in the specified order, moving on only when the user has correctly gowned the current item. When this *sequential gowning* process completes, the user should then be stood in a successfully gowned state. This should then be verified by running a second complete body scan, ensuring that no successfully gowned item became ungowned after its initial detection. On successful completion, the application should grant the user access to their high-risk facility, then waiting for another user to adopt the entrance gesture.

4.3 Diagram

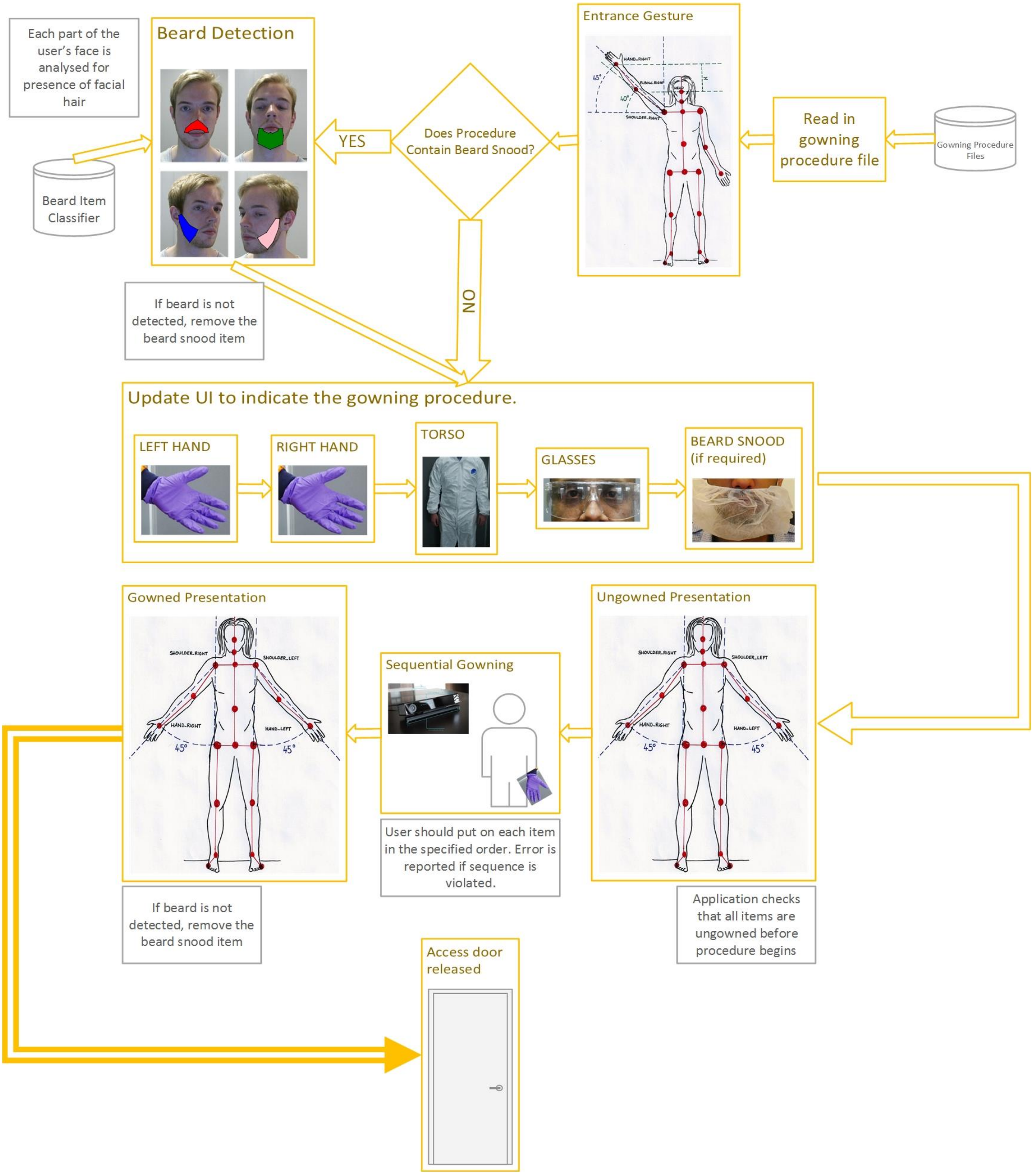


Figure 10: Recognition Module Diagram

5 Evaluation of Chosen Solution, Timeline and Anticipated Issues

5.1 Existing Solutions

Several solutions exist that have been included here, owing to their relevance towards the technical aims of this project.

5.1.1 StileEye

StileEye is an online fashion retailer that provides a computer vision based 'Image Search' function that allows their customers to upload their own fashion images, e.g. dresses or handbags, in order to provide a list of visually similar items from their own product line. Indeed, the company claim to have 'built a self-learning visual engine that understands and recognizes any fashion object in terms of its visual cues (e.g., style, shape, colour and pattern)' (StileEye, 2013).

Although similar feature extraction techniques were conceived for use in this project, StileEye's algorithm cannot detect the presence of known clothing items within uploaded images, and therefore could not directly solve the problem as presented. The company's algorithm also fails to provide a mechanism for facial hair detection and its proprietary nature would have incurred a hefty financial overhead.

5.1.2 Real-Time Clothing Recognition in Surveillance Videos

Some relevant work has been conducted in a recent paper on clothing detection in CCTV feeds (Yang & Yu, 2011). Using a combination of colour histograms and three different texture descriptors, the authors demonstrate promising results on a large dataset.

However, due to their interpretation of low-resolution surveillance videos, the research aims of this paper were restricted to categorising items of clothing into seven pre-defined classes: suit, shirt, T-shirt, jeans, short pant, short skirt and long skirt. Although many of the techniques used in this paper appeared to have strong relevance to this project, the existing research only covers the broad categorisation problem of recognised clothing, rather than a method of one-to-one item matching as required here.

5.1.3 Betaface API

The Betaface API is a web service for face detection and recognition that provides a variety of general classification and measurement data based on a user-supplied image (Betaface, 2015). In particular, the API claims to be able to identify the presence of a beard and approximate the density of a subject's facial hair in an uploaded image. Unfortunately, due to the application's proprietary nature, the company have been unwilling to supply technical details covering the technique they have employed, or to allow access to their underlying dataset.

Through preliminary testing, poor results were obtained when their algorithm was run on images containing small amounts of motion blur, as are commonplace in Kinect frame data. The Betaface API also does not support any kind of clothing recognition, so incorporating this technique would still have necessitated all of the subsequent research carried out in this project.

However, it was seen as desirable to utilise the Betaface API to draw quantitative comparisons against the beard recognition technique that this project later developed. To avoid the implementation complexities that would have resulted from running live Kinect images through a web service model, it would have been necessary to purchase a commercial licence for their SDK. Due to the restricted benefit the proprietary application would have had towards this project, and the significant cost that would be incurred through such a purchase, the API was not further considered in this project.

5.2 Anticipated Challenges

This project's primary challenge has been to design a PIC training scheme that produces provably effective item classifiers when trained on any PPE item. Due to the enormous variety of items that could be introduced to the system, it was seen as important to develop a single and sufficiently generic scheme that performs well, even on items that lack any recognisable features, such as logos or brand markings. However, the scheme should be suitably designed to capitalise on these features if they do exist. The major question for this project was whether the application developed according to the chosen technique would pass a professional validation process required by most of the target industries.

After analysing potential alternatives through online investigation and correspondence with industrial experts, this project appears to be the first specifically addressing gowning recognition. It was therefore likely to be challenging to find online sources or research papers offering strategic advice for the research phase of this project.

Alongside the work required into advanced image analysis techniques, it was also seen as likely that gaining familiarity with WPF and the Kinect SDK would take some time. The decision was initially made to invest in completing relevant tutorials before starting work on the full application. It was also thought necessary to identify suitable open-source image processing packages that would aid in implementing the developed scheme. After evaluating the benefits and drawbacks of a number of alternatives, the chosen package should then be integrated into the WPF Kinect-based solution.

5.2.1 Anticipated Legal, Social, Ethical and Professional Issues

An agreement has been signed between the host institution and the industrial sponsor that covers the allocation of intellectual property rights and ownership of the produced application. As key stakeholders of this project, both organisations made a significant contribution to the requirement sets. The industrial sponsor has also agreed to submit the final application through their validation testing procedure. The relevant documentation for this agreement is included in Appendix C – Confidentiality and Intellectual Property Agreement.

As per the clear regulations regarding the capturing of data from individuals, written approval was obtained from each volunteer taking part in the data gathering phase.

5.3 Research and Implementation Strategy

The project's research phase focused on three natural streams. Preliminary Kinect SDK work was first carried out to develop techniques for gesture recognition and body region extraction as per the described pre-processing step. The main research phase of this project then began by constructing and evaluating multiple PIC training schemes. The project's final challenge was then to implement a full Kinect-enabled C# solution containing a training and recognition module.

Various proof-of-concept applications were to be developed throughout the research phase to demonstrate key programmatic components for use in the eventual solution. These demonstrative applications should show the use of WPF user interface design techniques, interaction between the selected image processing libraries and relevant Kinect SDK features. Several such solutions should feature in this project's demonstrative material, although may not be present in the final submission.

6 Preliminary Kinect SDK Work

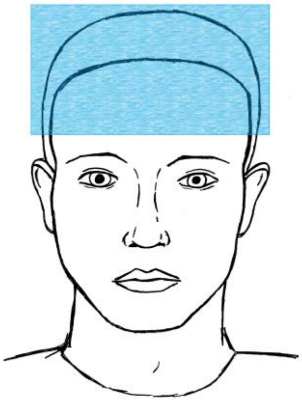
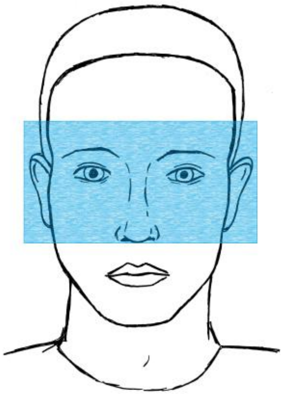
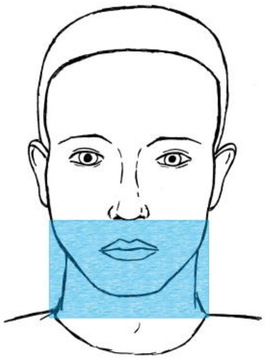
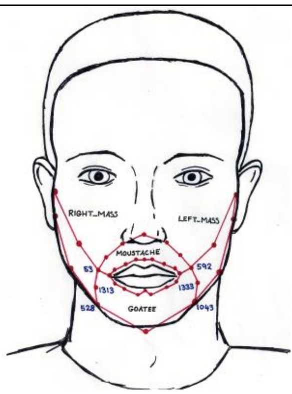
The following sections describe the development of two techniques that have been used to produce the best possible images for evaluation by PPE item classifiers. This has been achieved by careful manipulation of the Gesture, Body Index, Face and HD Face APIs found in the Kinect SDK.

A pre-processing technique has been developed to extract cropped body and facial regions from the live Kinect video feed before the image processing algorithm is applied. Gesture detection has also been developed to reject frames in which the user is either not present or not stood in a readable pose, i.e. those in which the body region(s) are not clearly visible to the Sensor. A similar technique has also been developed to identify the aforementioned entrance gesture, which a user should adopt to indicate their readiness to begin gowning items.

6.1 Extracting Body Components

Before developing the described pre-processing procedure, the relevant body regions were first defined. Owing to the fact that some PPE items cover multiple separate body regions (such as the GSK nitrile glove, required to be worn on both the left and right hands), *clothing types* have been defined that indicate these body region sets. After a new PPE item's clothing type has been specified, a single gowning procedure can include the item multiple times, provided a different body region is selected for each occurrence.

Although enough body regions have been included to support most commonly-used PPE, including those specified in SOP-PDK-0012, the list is not completely exhaustive. For example, there is no defined region whose area would be entirely covered by a full-face visor. Table 3 shows each body region alongside the Kinect frame type used to plot vertices of the desired extraction region.

Body Component Region	Diagram	Covered by
Hat		Face API
Goggles		Face API
Snood		Face API
Beard		HD Face API

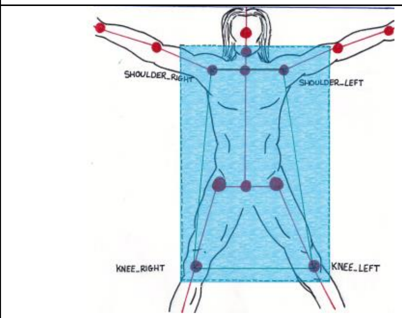
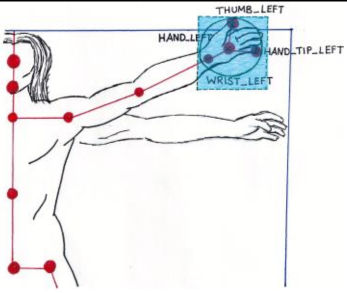
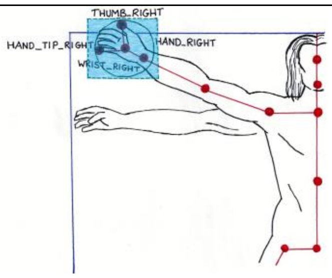
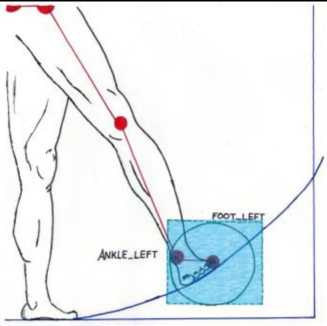
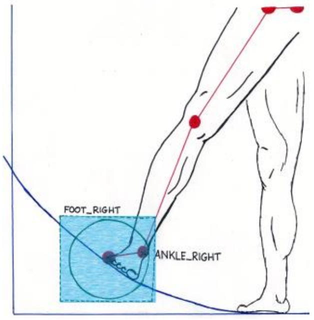
Body Component Region	Diagram	Covered by
Torso		Body Index API
Left Hand		Body Index API
Right Hand		Body Index API
Left Foot		Body Index API
Right Foot		Body Index API

Table 3: Body Component Regions

To cater for PPE items that cover multiple body regions and to ensure an item is not introduced that extends beyond the defined body regions, the clothing types in Table 4 have been defined.

Clothing Type	Allowed on the following regions:
Hat	Hat
Goggles	Goggles
Beard	Beard
Torso	Torso
Glove	Left Hand, Right Hand
Boots	Left Foot, Right Foot

Table 4: Clothing Types

Although the beard region describes an appropriate facial area for the beard snood, it is not suitable for use by the beard detection module. Due to the fact that some people choose to exhibit facial hair in subdivisions of this large region, smaller areas have been constructed to allow beard detection on subjects sporting non-uniform facial hair.

The *beard regions* in Table 5 have been defined for use by the detection module, which will each be analysed independently when seeking the presence of facial hair.

Moustache	Goatee	Left Mass	Right Mass
			

Table 5: Beard Regions

As previously discussed, the latest version of the Kinect SDK supports 25 skeletal joints, 5 face points and over 1000 high detail face points that are dealt with in the BodyIndex (Microsoft Corporation, 2016a), FaceFrame (Microsoft Corporation, 2016c) and HighDefinitionFaceFrames (Microsoft Corporation, 2016d) sections of the Kinect for Windows SDK (Microsoft Corporation, 2015b) respectively. By harnessing these APIs, polygons are mapped to the original scene image by connecting points around each body region's boundary. Since these points are captured using Kinect's depth sensor, their coordinates are first converted to the colour-space frame using the inbuilt CoordinateMapper (Microsoft Corporation, 2016b).

After the vertices of each polygon are obtained, encompassing rectangles are constructed that surround the body region. These rectangles are slightly enlarged about their centre points to account for any small error in Kinect point mapping. The portion of the image contained within the encompassing rectangle is then extracted from the colour frame and written to a new *body component* image.

6.1.1 Extracting Skeletal Components

The BodyIndex API from the Kinect SDK is used to map the 25 skeletal joints as previously shown in Figure 4.

Body component regions are formed by connecting sets of these points and are categorised into two classes, circular or rectangular, depending on which shape the region most resembles. Table 6 shows these regions, the names of the relevant body joints and the encompassing rectangles.

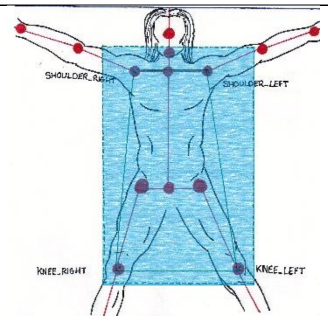
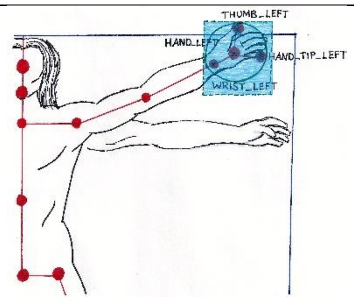
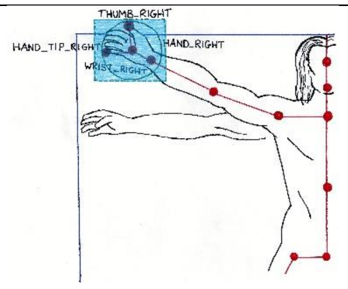
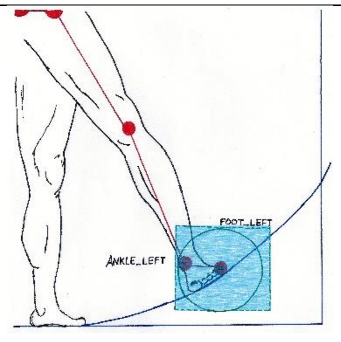
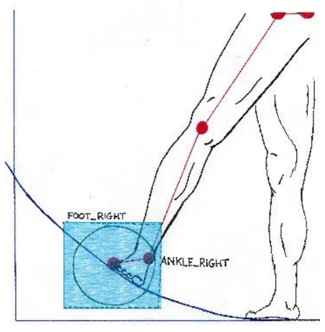
Skeletal Body Regions	Type	Diagram
Torso	Rectangular	
Left Hand	Circular	
Right Hand	Circular	
Left Foot	Circular	
Right Foot	Circular	

Table 6: Skeletal Body Regions

6.1.1.1 Circular Body Components

Circular body components are formed of two joints – the body region's centre point and one that lies on the boundary of the region. To account for region articulation whereby the body component region is rotated away from the camera, multiple extremity points can be defined allowing the algorithm to select the one that is furthest from the centre point. By carefully selecting these points, the application can construct a tightly-fitting polygon around the body region.

Let $p_c = (x_c, y_c)$ represent the centre point of a specific body region and $p_i = (x_i, y_i), i = 1, \dots, n$ represent the region's extremity points. The polygon is then formed by calculating

$$p_{MAX} = (x_{MAX}, y_{MAX}) = \underset{p_i : i \in \{1,\dots,n\}}{\operatorname{argmax}} \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$$

A circle is then constructed with centre point p_c, and radius r given by

$$r = \sqrt{(x_{MAX} - x_c)^2 + (y_{MAX} - y_c)^2}$$

An encompassing rectangle is then formed using an offset ε for safety with top-left point $(x_c - r - \varepsilon, y_c - r - \varepsilon)$, width $2(r + \varepsilon)$ and height $2(r + \varepsilon)$.

6.1.1.2 Rectangular Body Components

Body component rectangles are trivially constructed by connecting four independent vertices to form a simple polygon of maximal area before performing the described enlargement step.

6.1.2 Extracting Facial Components

The FaceFrame API has been used to map a user's face points to allow body component regions to be similarly extracted from the colour frame image. HighDefinitionFaceFrames have not been used in the main sequential gowning process owing to their increased processing time and requirement for the user to stand a short distance from the Kinect. However, HighDefinitionFaceFrames will be used to identify the region subdivisions during the beard detection process, as the increased number of mapped points provides the necessary accuracy in constructed polygons.

The points in Figure 11 are used for the FaceFrame facial mapping.

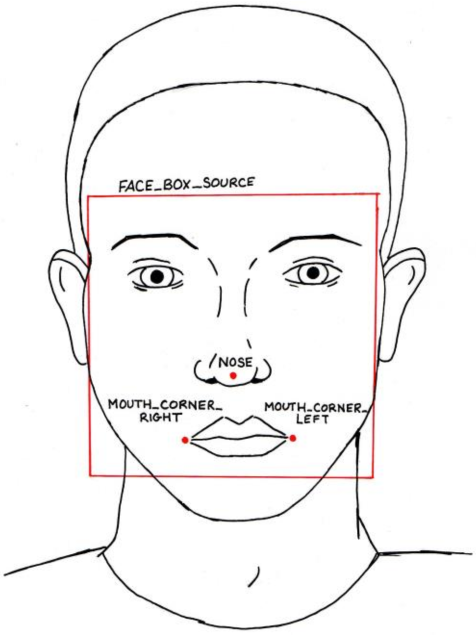


Figure 11: Kinect Basic Face Point Mapping

The hat, glasses and beard snood regions are formed by connecting the annotated face points as shown in Table 7.

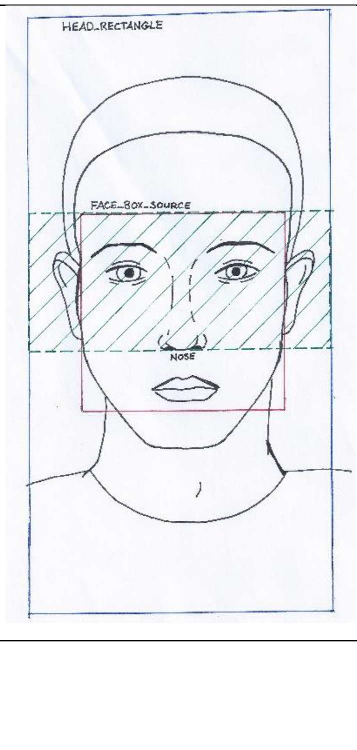
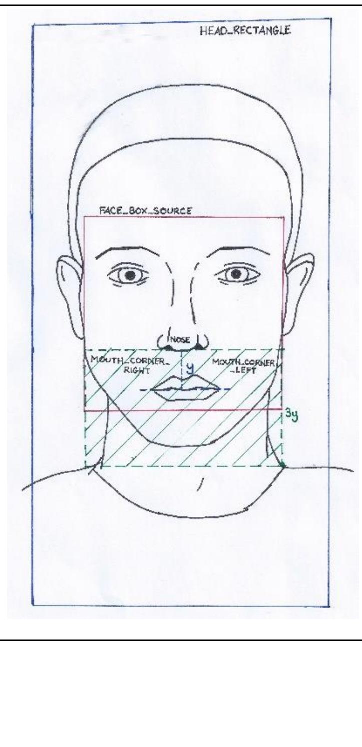
Hat	Glasses	Snood
		

Table 7: Facial Body Regions

6.1.3 Extracting Beard Region Components

Beard regions are constructed by combining HighDetailFacePoints. The regions are then formed by connecting the points as in Table 8.

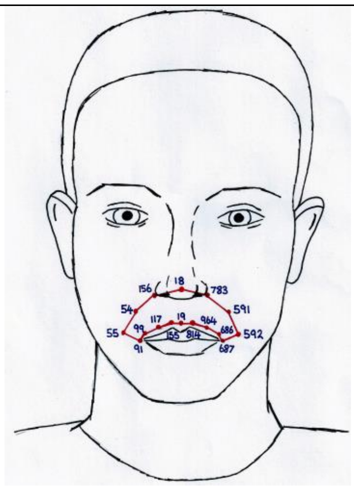
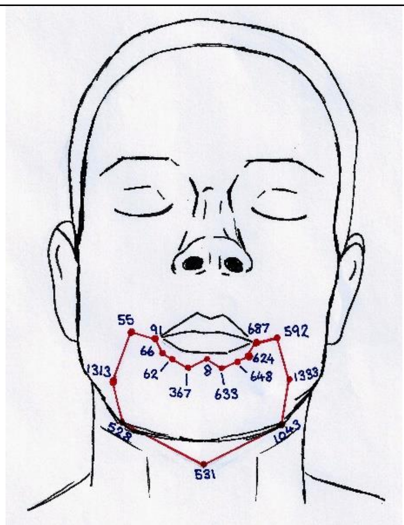
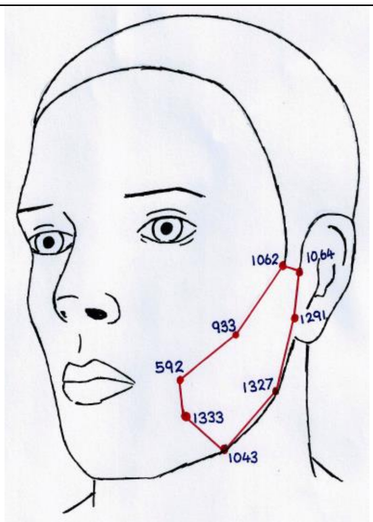
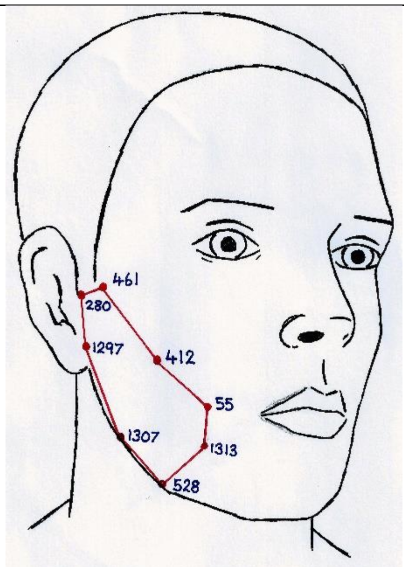
Beard Region	Diagram	Beard Region	Diagram
Moustache		Goatee	
Left Mass		Right Mass	

Table 8: Beard Regions (numbers indicate enumeration index)

Figure 12 shows an application independently extracting each of these beard regions. Note that for beard region images, pixels that are not contained within the original encompassing rectangle are set to black to reduce noise that may otherwise effect the accuracy of produced item classifiers.

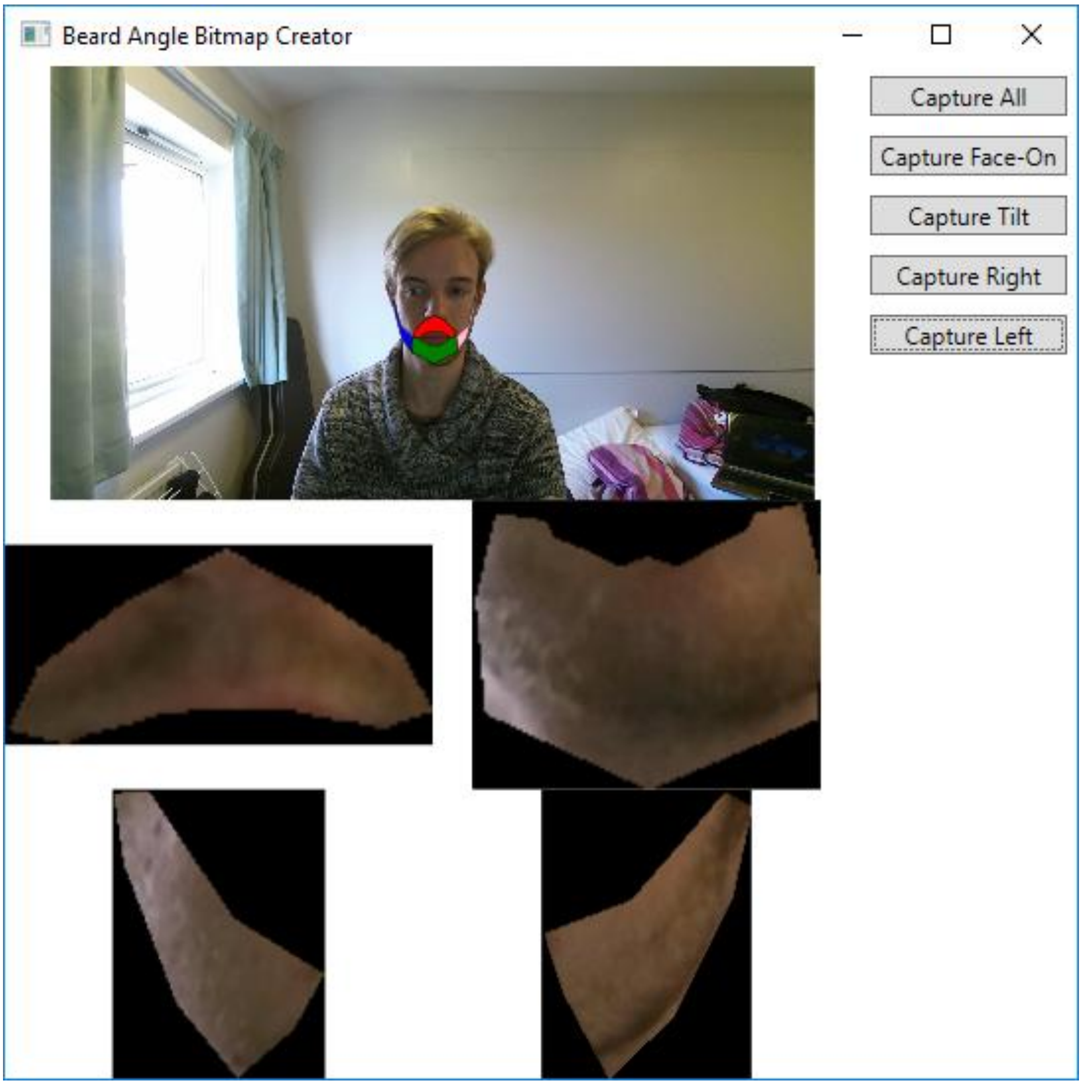


Figure 12: Live Beard Region Extraction

6.2 Gesture Detection

Gesture detection should be implemented in the application to recognise two poses that a user should adopt during the running of the recognition module. Recall that the first pose, referred to as the *entrance gesture*, should be adopted by a user to indicate their readiness to begin the gowning process. The user should be asked to adopt the second position, referred to as the *presentation gesture*, when there is a need to ensure that every item of PPE is clearly visible to the Sensor.

As part of each frame's processing cycle, any human body that is in a tracked state – i.e. the Kinect is able to map their skeleton – will be continuously tested to see if they are stood in either the entrance or presentation gestures.

6.2.1 Entrance Gesture

The entrance gesture involves a user being stood in an upright position with their right arm raised above their head. With reference to Figure 13, the application will 'recognise' this gesture if $\alpha > 40°$, $\beta > 45°$ and $x > 0$.

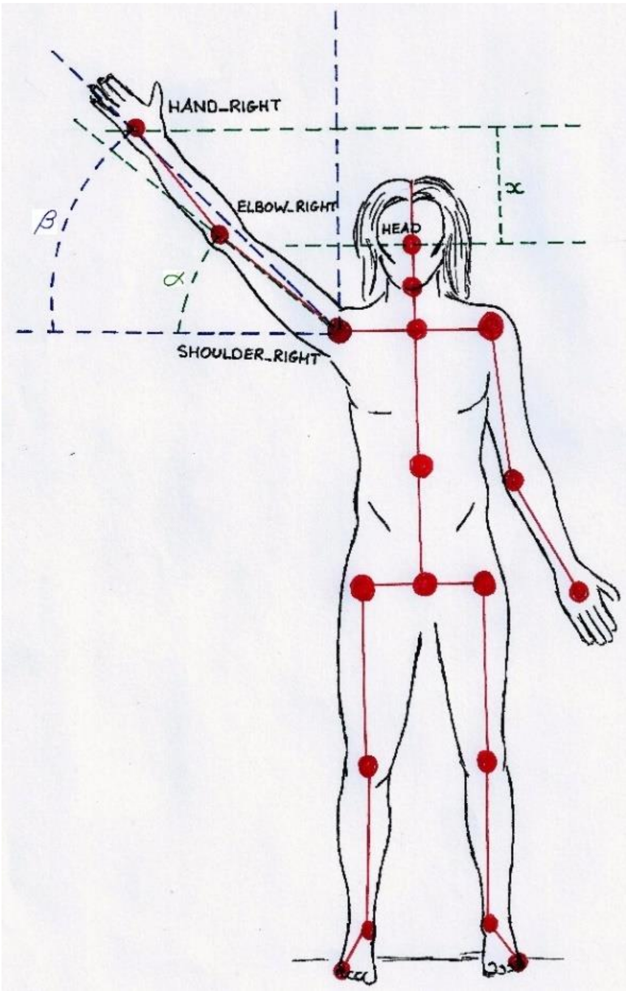


Figure 13: Entrance Gesture

6.2.2 Presentation Gesture

A user can adopt the presentation gesture by standing in the position shown in Figure 14.

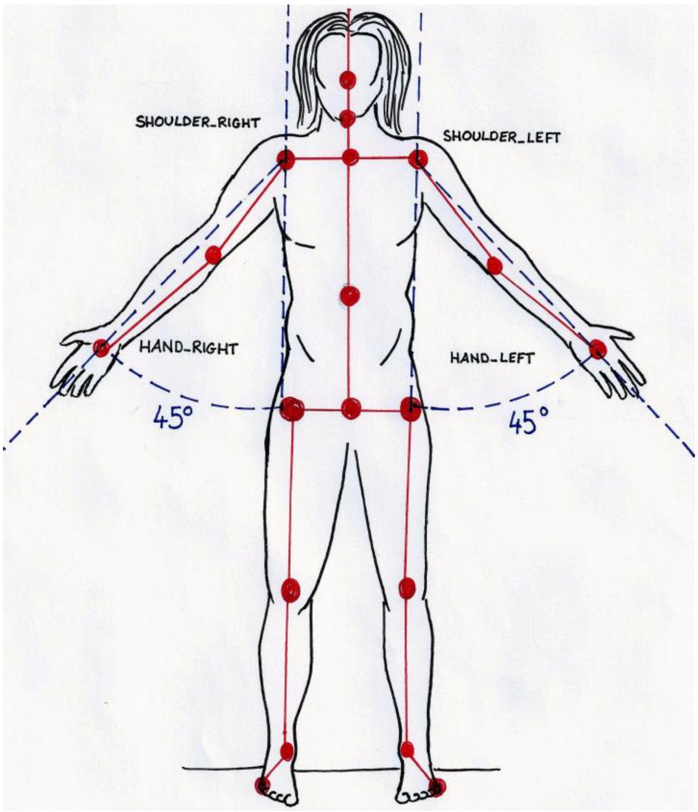


Figure 14: Presentation Gesture

As a more complex gesture, a database file has been generated with the Microsoft Visual Gesture Builder software (Microsoft Corporation, 2015c) that completely describes the required skeletal joint positioning. The application then uses this file to determine whether the user is stood in the presentation gesture.

The Microsoft Visual Gesture Builder software generates the database file by applying a machine learning approach to pre-recorded Kinect skeletal frames from three volunteers. After specifying parameters that specify the type of gesture – for example, which parts of the body it involves, whether is it symmetrical etc. – the software analyses each volunteer's skeletal frames, which are manually labelled to indicate those which show the subject stood in the gesture pose. The presentation gesture was trained on a training set partition of this data before using the testing set partition to confirm the accuracy and tolerance.

7 Developing a Gowned/Ungowned Image Classifier

7.1 Strategy and Preliminary Data Collection

In order to gain familiarity with complex image processing techniques, the decision was made to use MATLAB with its Image Processing Toolbox (MathWorks, 2015a) as the primary research tool for developing a successful PIC training scheme. By developing and testing a variety of techniques, a viable algorithm for constructing item classifiers was created, which could then be re-implemented as a core part of the C# training module. Alongside the substantial online documentation and availability of sample code, MATLAB's self-contained nature and computer vision toolbox provided sufficient support to conduct this research phase without continually needing to seek extra libraries.

7.1.1 Preliminary Data Collection

To evaluate any developed techniques, it was first necessary to capture a dataset containing gowned and ungowned images for a number of PPE items. The dataset contains body region photographs from a number of volunteers who were asked to wear each PPE item in turn and then to provide an image of each region without any item being worn. In this initial research stage, data was collected from each volunteer for the GSK lab goggles, purple nitrile gloves and white beard snood. In addition, images were captured that show the beard region from clean-shaven and bearded volunteers sporting a range of facial hair styles.

Although this initial dataset only covers a subset of possible PPE items – indeed, it does not even cover those set in the GSK Building 5 requirements – they were chosen for the initial dataset as they exhibit the challenges listed in Table 9.

Item	Challenge	Dataset Size
GSK Goggles	Transparent, reflective lenses result in colour varying due to skin tone. Thin, barely visible lenses.	20
GSK Gloves	Cause no change in shape to the hands when worn.	16
GSK Beard Snood	Hugely inconsistent 'shape' that depends on how the wearer put it on.	14
Beard	Since not an 'object' it is not clear how best to recognise the presence of facial hair.	13

Table 9: Preliminary Dataset PPE Items

7.1.2 Evaluating Techniques

In order to assess the effectiveness of each technique, the gowned and ungowned images for each PPE item were partitioned into a training and testing set. Classifiers were then trained using only training set data with their effectiveness being evaluated by their performance on the testing set. Results were then demonstrated as a confusion matrix, which allows accuracy, precision (positive predictive value), false negative rate and false positive rate to be calculated.

For binary classification problems, such as the gowned-ungowned problem presented in this project, the formulae displayed in Figure 15.

		Classifier Prediction	
		Gowned	Ungowned
Actual Label	Gowned	True Positives (TP)	False Negative (FN)
	Ungowned	False Positive (FP)	True Negative (TN)

Figure 15: Confusion Matrix

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

7.2 HSV Colour Thresholding

7.2.1 Explanation

The first attempt at building a PPE Item Classifier Training Scheme solely relied on colour information to discriminate between gowned and ungowned images. As many PPE items are vibrantly coloured, it seemed likely that this would play a strong role in the final solution.

The HSV Colour Thresholding solution relied on developing a series of functions, referred to as *filter functions*, where each directly corresponds to a PPE item. A filter function accepts a pixel as a parameter and returns a binary value that indicates whether the pixel should be 'filtered in or out' based on its three-channel colour information. Precisely, filter functions should be of the following form:

$$f : S \to \{0, 1\}, S \subseteq \mathbb{R}^3$$

The item classifier for a particular PPE item will use its associated filter function to map each pixel from the input image to a new binary image, one in which each value is either an *on-pixel* or an *off-pixel*. The number of on-pixels present in the new filtered image is then calculated as a percentage of the image's size and the classification result for the input image – either gowned or ungowned – is given by testing to see if the *on-pixel percentage* is above or below a set threshold value.

The item classifier will therefore make predictions using the following algorithm:

Constructing HSV Colour Thresholding Item Classifier

1. Input: Input Image, Filter Function, Threshold Value
2. Output: Classification result – either gowned or ungowned
3. Use the associated filter function to map each pixel of the input image to a new filtered binary image.
4. Calculate the on-pixel percentage:
    a. Enumerate the on-pixels
    b. Divide by the image's size in pixels and multiply by 100
5. Test if the on-pixel percentage is above the set threshold.
    a. If so, return 'Gowned'
    b. Otherwise, return 'Ungowned'

7.2.1.1 Filter Function

The filter functions that correspond to PPE items are designed to map pixels of similar colour to the PPE object to one and map all other pixel colours to zero. Applying such a function to a gowned image therefore yields a new binary image where the only on-pixels are located in the region that corresponds to the PPE item's location. An entirely zero-valued image should be returned if the input image contains no pixels of similar colour to the PPE item, ideally when an ungowned image is provided. Figure 16 demonstrates the result of applying a suitably designed filter function to a gowned and ungowned image from the GSKPURPLE (GSK nitrile gloves) dataset.

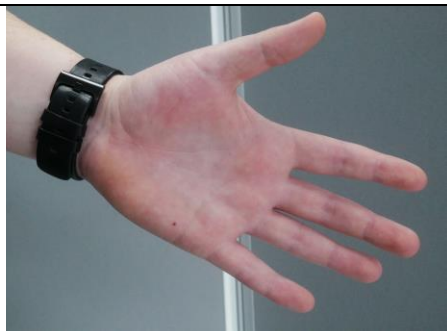
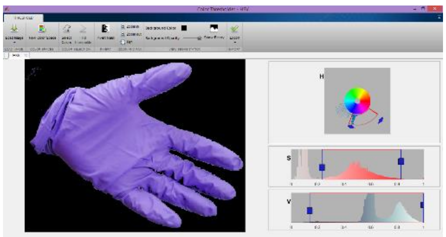
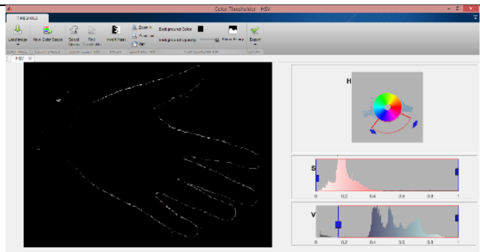
	Gowned	Ungowned
RGB		
Reconstruction of filtered binary image		

Figure 16: Filter Functions

To account for instances of similarly-coloured pixels appearing in ungowned images, each PPE item will have a set threshold that defines the number of mapped on-pixels required (as a percentage of the input image size) to return a 'gowned' classification result. Suitable filter functions can be designed by defining a range of colours that map to on-pixels, and otherwise map to off-pixels.

7.2.1.2 HSV Colour Space

Owing to the adverse effect that inconsistent lighting can have on the RGB colour space, the decision has been made to begin the classification procedure by converting each pixel to the HSV colour space.

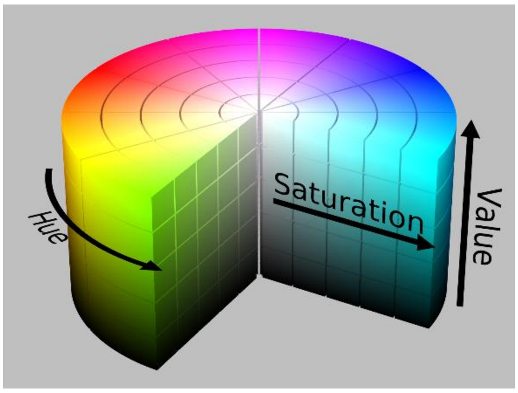


Figure 17: HSV Colour Space (Horvath, 2010)

The Hue, Saturation, Value (HSV) colour space is a commonly-used model use to describe a pixel in terms of its hue (or tint), shade (or saturation) and brightness (or value) – see Figure 17. By representing pixels in these three channels, it is made significantly easier to segment detrimental effects that shadow or inconsistent lighting may cause to the image. Due to the fact that red, green and blue components are all correlated with the amount of light that hits the object, it is made far more challenging to deal with these inconsistencies using the RGB colour space.

After applying the transformation from RGB, a range will be calculated for each HSV channel that best represents the colour variation caused by the relevant PPE item. Each filter function will then proceed by determining if the H, S and V values of the given input pixel, mapping to an on-pixel if all fall inside the range, or an off-pixel otherwise.

7.2.2 Solution Design

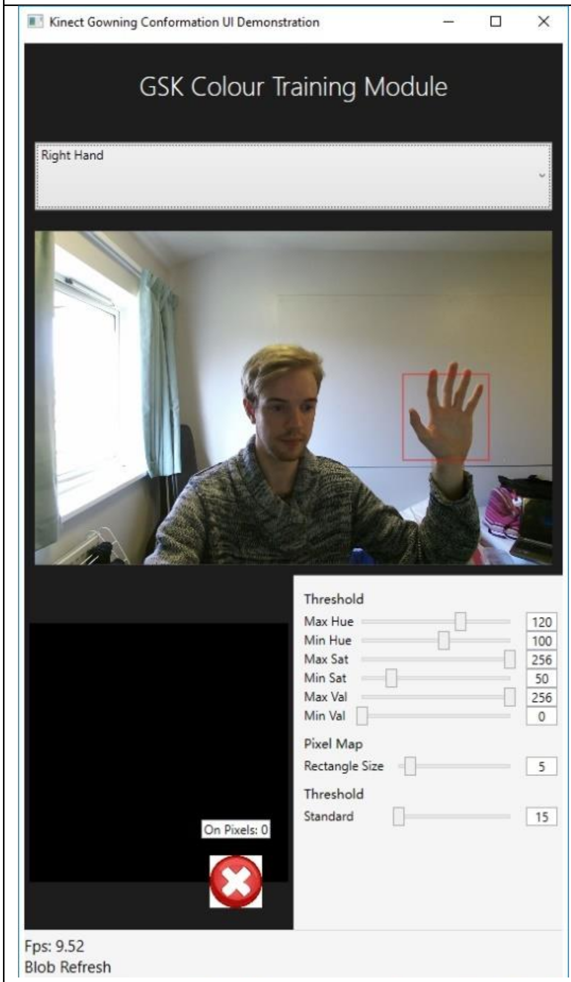
An application has been identified to approximate the HSV ranges for each PPE item's filter function. Minimum and maximum values are calculated using MATLAB's Colour Thresholder (MathWorks, 2014) by manually adjusting sliders to best capture the colour variation within each PPE item. These six values are set with the intention of yielding a higher on-pixel percentage when the filter function is applied to gowned images than when applied to ungowned. After these ranges have been calculated, the on-pixel percentages are calculated for each training set image. A threshold value is then computed for each PPE item by taking the mid-point between the medians of the on-pixel percentages over its gowned and ungowned images; that is,

$$Threshold = \frac{median(OPP_G) + median(OPP_U)}{2}$$

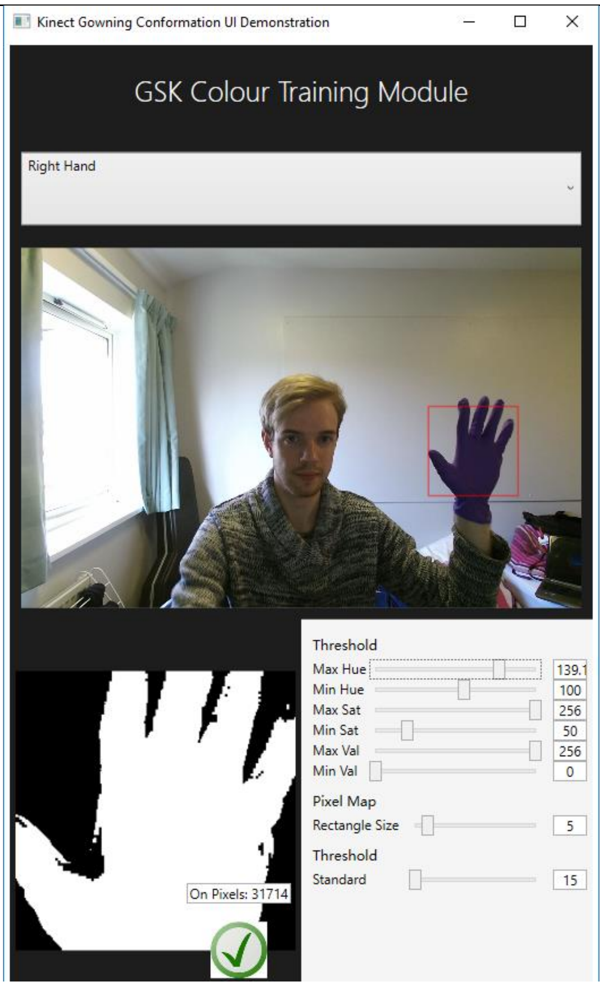
where OPP_G and OPP_U are, for the sets of gowned and ungowned images respectively, the vectors of on-pixel percentages.

To ensure that range and threshold values are effective when filter functions are applied to live Kinect data, a C# WPF 4.5 application (Nathan, 2013) has been constructed that performs real-time HSV thresholding on pre-processed frames. Figure 18 shows the application performing live HSV thresholding to produce on-pixel percentage values and a gowned or ungowned classification result on a number of PPE items.

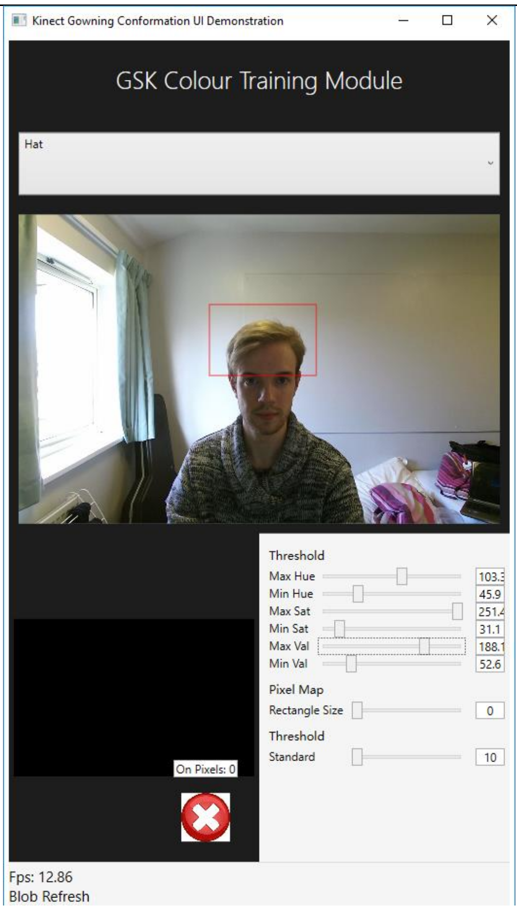
Ungowned Hand Region



Gowned Hand Region



Ungowned Hat	Gowned Hat



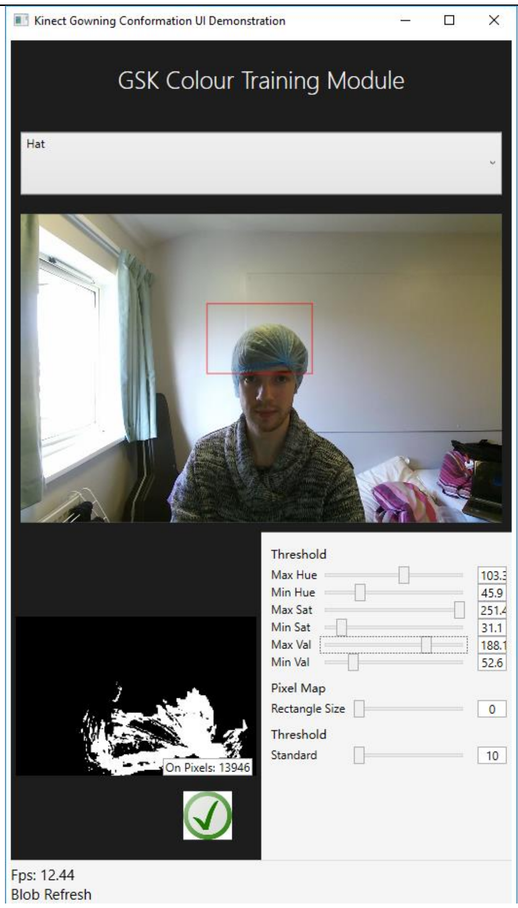


Figure 18: Live HSV Colour Thresholding

7.2.3 Implementation

The scheme has been evaluated by manually setting HSV ranges for each PPE item using the MATLAB Colour Thresholder and verifying their effectiveness with the custom-built Kinect application. A third application has been constructed to produce on-pixel percentages for each image by applying each filter function to its respective training set. Threshold values are then calculated for each item by applying the previous formula.

The third application which originally applied the relevant filter to an input image set has been extended to report each image's classification result according to the set threshold. By identifying the number of correct and incorrect predictions when classifiers are run on the testing set, informative confusion matrices and all other measures can then be constructed.

7.2.4 Preliminary Testing

7.2.4.1 HSV Ranges

	Min Hue	Max Hue	Min Sat	Max Sat	Min Val	Max Val	Threshold
beard	0.025	0.2	0.358	1	0.209	0.372	1.382
gskgoggles	0.129	0.978	0	0.282	0.596	0.793	2.877
gskpurple	0.686	0.932	0.248	0.805	0.169	1	11.846
gsksnood	0.086	0.618	0	0.375	0.462	0.742	8.318

Table 10: Preliminary HSV Ranges

7.2.4.2 Results

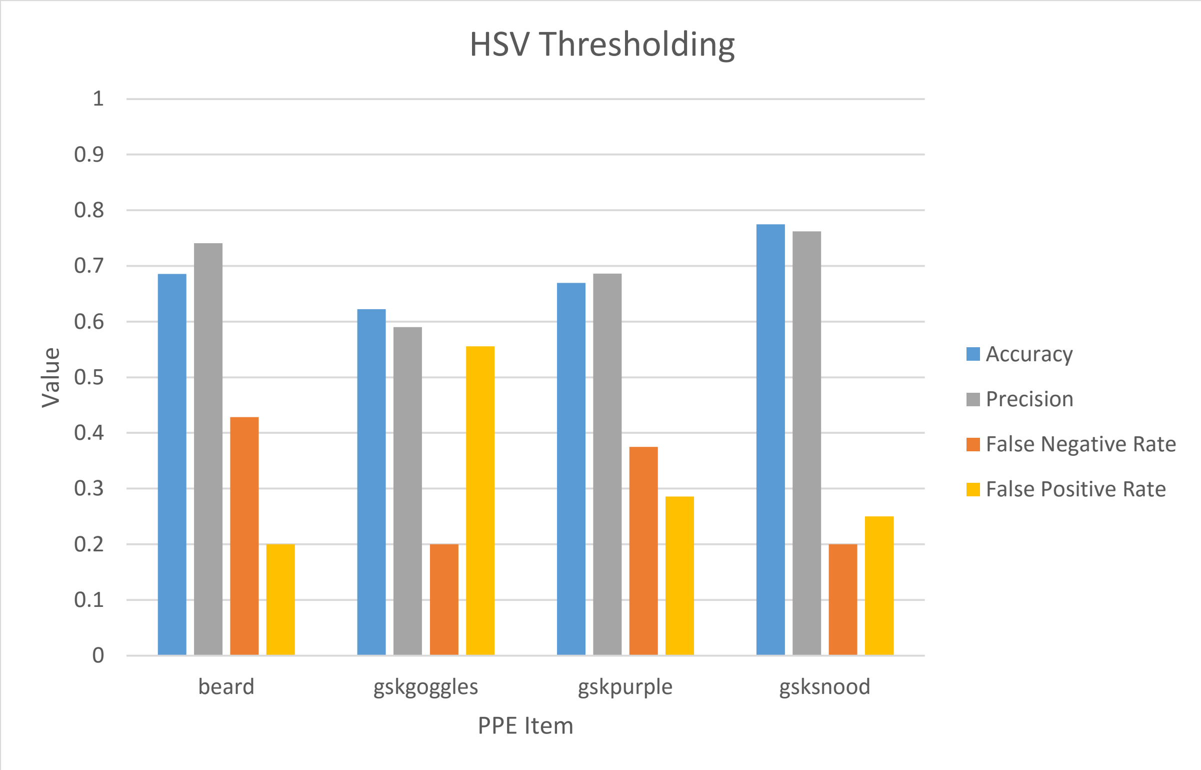


Figure 19: HSV Thresholding Test Results

7.2.5 Evaluation

The data indicates that the HSV Colour Thresholding scheme is a moderately reliable classification method. Although there was little variation between the items' accuracy levels, higher values did occur for the gloves, beard and beard snood, which was likely caused by their vibrant colour. Unsurprisingly, lower values were obtained for GSK lab goggles, which is likely to have been caused by their almost-transparent nature, which provides little distinguishing colour.

Substantial difficulty was encountered when calculating HSV ranges across many of these classes, particularly those with images captured under varying light levels or background colours. Although this issue was minimised by using the HSV colour space, the inaccuracies that occurred when selecting ranges were likely to have detrimentally affected the classification accuracy for all PPE items, including those that had pronounced colour definition.

In reality, the proposed HSV Thresholding solution seemed impractical without first resolving these inconsistencies. This could have been achieved during data capture by ensuring a constant light source or installing a ‘green screen’-style background, although it would still have been necessary to remove the manual MATLAB filtering method, which would be too complex for a user unspecialised in computer vision.

7.3 SURF Bag of Visual Words with SVM Classifier (SBS)

7.3.1 Explanation

To resolve the problems of overreliance on colour information, an alternative technique was developed that instead uses shape information to classify input images. By running a Speeded-up Robust Features detector on the training set, gowned and ungowned images were encoded in terms of their local interest points, which generally correspond to edges and corners. In order to use this information to classify unseen images, a Bag of Visual Words framework was used to standardise the interest point information across both classes. By associating each training image’s standardised interest point data with its correct classification label, a commonly-used classification model known as Support Vector Machine (SVM) was then trained on the entire set.

7.3.1.1 Speeded-up Robust Features (SURF)

Speeded-up Robust Features (SURF) is a technique used to extract local interest points from a given greyscale input image. A SURF detector produces a set of scale and rotation-invariant key points that can be used by matching algorithms to detect pre-trained objects (see Figure 20). Each SURF point is represented by a 64-dimensional vector that encodes the point’s location and local neighbourhood information.

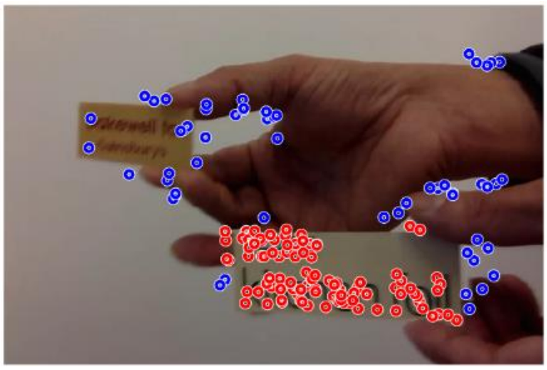


Figure 20: SURF for Object Matching (Bhalerao, 2015)

Even prior to designing and testing the algorithm, this technique was expected to face the converse problem to that found using the HSV Colour Thresholding scheme. Owing to SURF’s restriction to single-channel (or greyscale) images, the technique has a sole reliance on shape information and was therefore likely to systematically misclassify items that cause little deformation to the body region when worn. For example, owing to the few differences between greyscale versions of gowned and ungowned images from the GSK Nitrile glove dataset (shown in Figure 21), it seemed unlikely that a SURF-based classifier would prove effective.

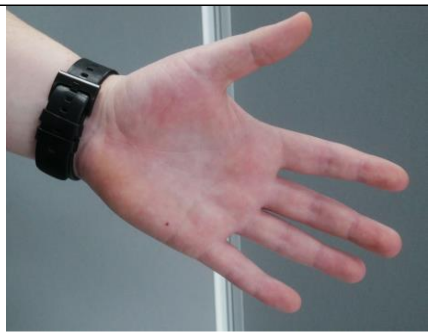
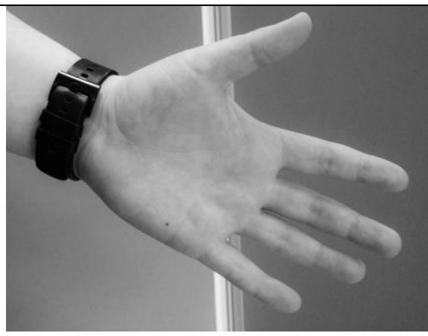
	Gowned Image	Ungowned Image
RGB		
Greyscale used by SURF		

Figure 21: SURF Invariance to Colour

By design of the algorithm, it is unlikely that running the SURF detector over two separate images – even those taken from the same class – would yield the same number of key points. To prevent this disparity resulting in some images being over-represented during the training process, it was necessary to standardise the SURF information over each dataset image. This could have been trivially achieved by retaining only the ‘strongest’ k points for each image, where k is given as the fewest number of key points returned over the entire training set. However, owing to its proficiency for representing relationships between pre-categorised images, a bag of visual words approach has been used to standardise the SURF information.

7.3.1.2 Bag of Visual Words (BoW)

Bag of Visual Words is an approach used to represent individual images in terms of the entire input set (Csurka, et al., 2004). The process begins by constructing a vocabulary of k-many features, which are chosen to be the most significant across all input images.

Although the features are generally more abstract, the following example may aid with explanation. Consider the process of forming a vocabulary over an input set that contains multiple images of humans wearing goggles and multiple images of humans not wearing goggles. By analysing the SURF points over all of these images – irrespective of gowned or ungowned class – significant features may be returned that, for example, represent a logo marking, human nose or plastic frame.

By constructing this vocabulary of k-many significant features, *feature vectors* are then constructed by encoding each training image as a histogram of its significant-feature occurrences. In other words, each training image is assigned a k-dimensional vector where the i^{th} element describes the number of occurrences of the i^{th} significant feature in the image.

With this technique, is was hoped that feature vectors across the gowned and ungowned classes differ in some recognisable way. To refer back to the example, it should be that the class of images that show humans wearing goggles contain more logo and frame features than the class of images that do not contain goggles. Since both classes depict the face region, they are equally likely represent the nose feature. By virtue of this technique, it is considerably more likely to obtain highly informative feature vectors.

7.3.1.3 Support Vector Machine (SVM)

Once the feature vectors for each class have been calculated, they are then provided to a supervised classification model, which uses each input set's known class labels to develop a prediction function that is capable of categorising similar but previously-unseen images. Owing to its frequent and successful use in image classification, this project will use a support vector machine as the prediction model.

A support vector machine is a commonly-used supervised model that calculates a classification boundary that best partitions a training set's feature vectors in a high-dimensional *feature space*. For linear SVMs that rely on linear kernels, this boundary is a straight line, so only its gradient and intercept need be determined. However, it is possible to define more complex SVMs that allow their classification boundaries to have extra degrees of freedom by increasing the feature space's dimensionality, using a technique known as the 'kernel trick'.

An SVM must be trained on a labelled set of vector features, where each feature's label indicates its correct class (in this case, either gowned or ungowned). The SVM's internal 'compute' function then executes an algorithm that fits the best classification boundary line according to the mapped data points.

To classify an unseen feature vector, the SVM first represents the vector in feature space coordinates and returns the class label that corresponds to the side of the classification boundary it inhabits. As in the previous proposal, confusion matrices can be constructed by running the trained SVM on the testing set partition.

7.3.2 Implementation

To demonstrate the effectiveness of this technique, a MATLAB application has been developed to construct the relevant SVM classifier for each PPE item represented in the training set. The application generates a suitable PPE item classifier by extracting SURF information from the training set images, encoding feature vectors using a constructed bag of visual words and training a SVM classifier. The application is also able to produce detailed accuracy statistics by running the classifier on the testing set partition.

As before, a secondary application has also been written to demonstrate the scheme's suitability to live Kinect data. A C# application has therefore been constructed that analyses a PPE item's gowned and ungowned dataset to generate a SURF-based SVM classifier and allows a user to test the accuracy by performing real-time prediction on pre-processed Kinect frames.

7.3.3 Diagram

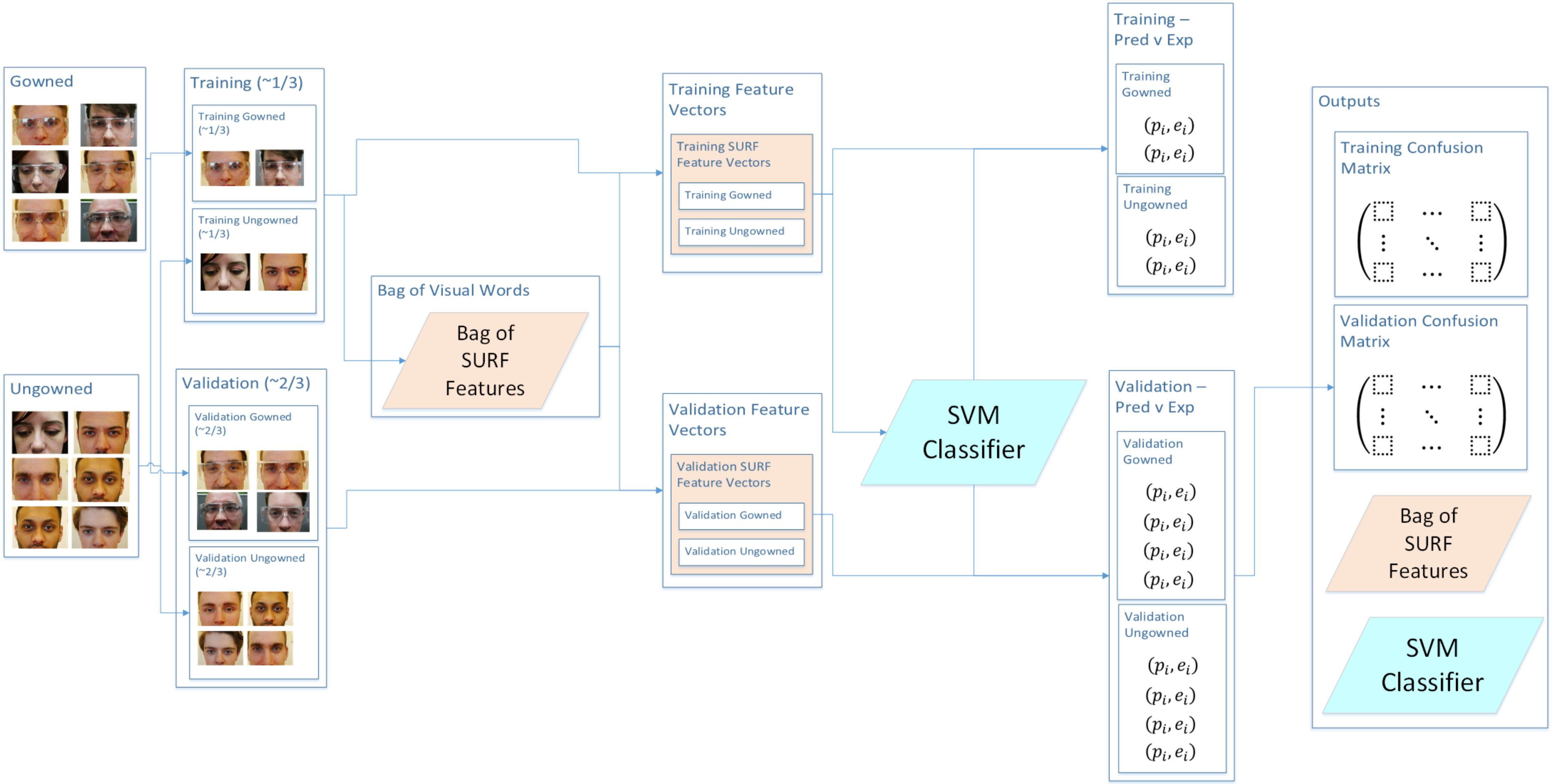


Figure 22: SBS Solution Diagram

7.3.4 Preliminary Testing

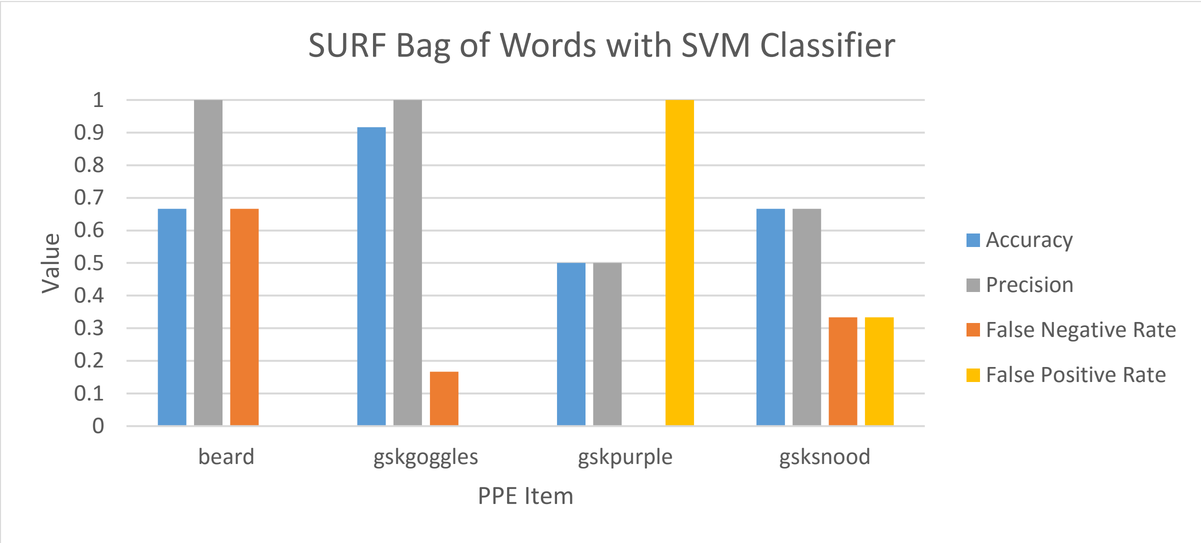


Figure 23: SBS Preliminary Test Results

7.3.5 Evaluation

Although this data suggests a small overall improvement when compared to the HSV Colour Thresholding scheme, the results confirm the expected accuracy drop in the GSK purple gloves classifier, which causes little shape change when worn. Although colour information has previously been shown to be important when generating a GSK beard snood classifier, the data suggests that shape information could be equally useful.

Importantly, however, this technique demonstrates the benefit of using SURF information when classifying items that only exhibit distinctive shape, rather than colour, indicated by a significant improvement in the GSK goggles classifier accuracy. Although some accuracy improvement was observed when detecting beards, the false negative rate remains too high to consider industrial use.

To make any significant improvement to the existing proposals and enable the construction of classifiers for items that do not exhibit strong colour (e.g. GSK lab goggles) or shape (e.g. GSK Nitrile gloves), it appeared that a scheme would be required to combine both features.

Moreover, since neither technique adequately constructed a beard detector, it was thought that a third feature was needed. Although shape and colour were not sufficient to determine the presence of facial hair, it was hoped that integrating a measure of each class’s texture information would yield stronger results.

7.4 Multi-Feature Bag of Visual Words with SVM Classifier (MFBS)

7.4.1 Explanation

The MFBS approach is an enhanced version of the previous scheme that constructs PPE item classifiers using multiple image features, which in this case represent shape, colour and texture information. This scheme required investigation into additional feature extraction techniques to allow an image's colour and texture information to be suitably represented, and also into methods for combining these features to construct an improved SVM classifier.

7.4.2 Solution Design

7.4.2.1 Colour Feature Extractor

A local colour descriptor was constructed to produce a set of vectors that indicate an image's average colour across a set of extracted sub-blocks of a fixed size. By virtue of the pre-processing step, the target body region in both gowned and ungowned images appear in approximately the same place, meaning that the location of each colour point is likely to be useful. To capitalise on this, each five-dimensional feature vector encodes the sub-block's centroid alongside its average colour.

Due to its ability to provide a quantifiable measure of the visual differences between colours, images are first converted to the three-channel LAB colour space, where L represents 'lightness' and a and b represent the colour dimensions. Although LAB provides the most accurate colour representation (MathWorks, 2016c), it is often difficult to implement, given the lack of simple conversion formulas from RGB or HSV. This is due to the fact that LAB is device-independent, unlike RGB and HSV, which therefore must first be transformed to an absolute colour space before conversion to LAB.

The following algorithm, adapted from an existing MATLAB implementation (MathWorks, 2016a) and demonstrated in Figure 24, describes how an image's local colour information is represented.

Local Colour Feature Extractor
1. Convert RGB image to LAB colour space using library. 2. Split image into x, y independent cells of size 16×16. 3. Construct a colour feature vector (L, a, b, X, Y) for each cell where (L, a, b) is the average LAB value over the cell and (X, Y) are the cell's centre-point coordinates, normalised to a range $[-0.5, 0.5]$, allowing feature vectors to be compared to images of varying dimensions. 4. Return the set of colour feature vectors.

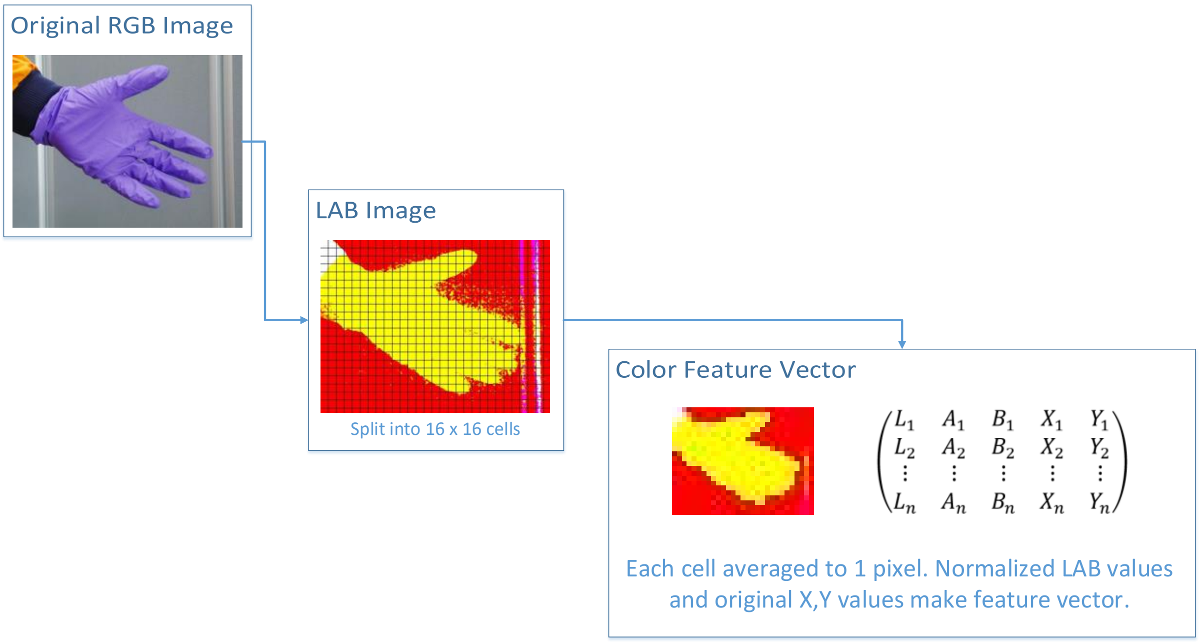


Figure 24: Local Colour Extractor Diagram

This feature extraction technique has been initially evaluated using a custom-built application that trains an SVM using images from three categories: boats, flowers and horses. As each category expresses a distinctive colour (boat images are predominantly blue, flowers yellow and horses brown) it is made easier to judge the method's effectiveness over the original dataset. Extracted features from the three classes were passed to the SVM trainer after being encoded as a bag of visual words.

Boat	Flower	Horse
		

Table 11: Local Colour Feature Extractor Alternative Dataset

	Boat	Flower	Horse
Boat	1	0	0
Flower	0	1	0
Horse	0	0.5	0.5
Accuracy	0.83		

Figure 25: Local Colour Feature Extractor Alternative Dataset Test Results

The above test shows a relatively high classification accuracy, demonstrating the local colour feature extractor's effectiveness. The technique has therefore been used to represent colour information in the MBS combined classifier.

7.4.2.2 Texture Feature Extractor (HOG)

The histogram of orientated gradients (HOG) is a well-known technique used to describe an image's local texture information. The method first splits the input image into cells of a predefined size and produces a feature vector that indicates the gradient information in each cell. The method has been used extensively for detecting pedestrians in CCTV images (Ellis, et al., 2009) and for problems in optical character recognition (Ebrahimzadeh & Jampour, 2014), but has here been used to classify an object's material.

A suitable balance between information loss and feature length has been identified by optimising the descriptor's cell-size parameter. Figure 26 shows HOG features for a sample glove image using three different cell sizes. It can be observed that cells of 64×64 size appear to lose information, whereas 16×16 cells produce features of extremely high dimensionality. To best satisfy these two concerns, the HOG descriptor appears to be set best to work on cells of size 32×32, although 64×64 had to be used to prevent the occurrence of memory overload exceptions. The formal testing section evidences the impact this had on SVMs built using HOG features.

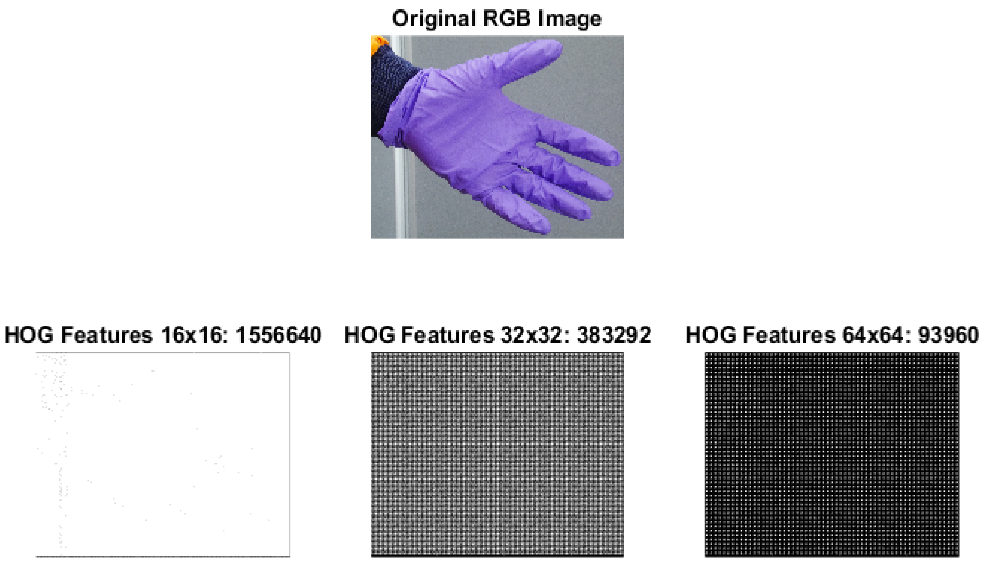


Figure 26: Selecting HOG Cell-Size

An application has been written to evaluate the effectiveness of HOG features for texture classification by training an SVM on three material types (bark, wood and brick) taken from a textured surface dataset (Lazebnik, et al., 2005). Extracted HOG features were encoded with a bag of visual words before being passed to the SVM training algorithm.

Bark	Wood	Brick
		

Table 12: Local Texture Extractor Alternative Dataset

	Bark	Wood	Brick
Bark	0.75	0.16	0.09
Wood	0.07	0.91	0.02
Brick	0.02	0.04	0.95
Accuracy	0.87		

Figure 27: Local Texture Extractor Alternative Dataset Test Results

The high accuracy results obtained in the above tests demonstrate the effectiveness of using HOG information to classify different material types. HOG features have therefore been used to represent each class's texture information in the MBS combined classifier scheme. Formal testing has been included in a later section of this report, evaluating the extractor's suitability to facial hair classification as well as the other PPE items used in this project.

7.4.2.3 Constructing a Multi-Feature Model

Research was conducted to identify a suitable method for combining extracted SURF, Colour and HOG features in a single, improved classifier. The three feature extractors produce the following sets:

- s SURF feature vectors of length 64
- c COLOR feature vectors of length 5
- $t = 1$ HOG feature vectors of length 93960

Even by employing a technique to ensure each extraction method produces an equal number of feature vectors, an image's three descriptor types cannot simply be concatenated due to the varying dimensionalities. Moreover, as each feature vector type encodes the interest point's location, these would be in direct conflict if combined into a single vector.

This issue has been resolved by employing a similar bag of visual words scheme, although this time, an individual bag is required for each feature type. Each training set image is encoded into three normalised vectors of identical length – a SURF feature vector, a colour feature vector and a texture feature vector. As each image's three feature vectors are now of standard length and only contain general (rather than local) feature information, they can be concatenated to form a *multi-feature vector*. The matrix formed by appending each multi-feature vector is then row-normalised to ensure that each constituent feature is equally represented (i.e. carries an equal weight) before it is passed as training data to construct a *multi-feature* SVM classifier.

7.4.2.3.1 Post-SVM Combination

Although not implemented, an alternative scheme was also considered that uses ensemble learning techniques. Rather than training a single SVM classifier on multi-feature vectors, three classification models could be trained where each is given one of SURF, colour or texture feature vectors. By employing a suitable probabilistic classification model, cross-validation or linear programming could be used to find the ideal weighting over the predictions made by these classifiers, i.e. that which yields the highest confidence value.

For example, let (s_i, c_i, t_i) indicate three confidence values when a produced SURF, colour and texture classification model is run on a training image i. Assume these values are in the range $[0,1]$ and higher values indicate a higher likelihood that the input image shows a *gowned* body region.

Three weights w_s, w_c, w_t would be calculated that provides the strongest confidence value for the ensemble classifier f, which also takes values within the range $[0,1]$ with higher values reflecting a higher gowned probability:

$$f_i = w_s s_i + w_c c_i + w_t t_i$$

$$w_s + w_c + w_t = 1$$

High confidence measures are obtained when high values are returned for gowned images and low values are returned for ungowned images. To account for this, the following definition provides a single confidence measure that accounts for both input classes:

$$g_i = \begin{cases} f_i & \text{if image is gowned} \\ 1 - f_i & \text{if image is ungowned} \end{cases}$$

Again, $g_i \in [0,1]$ with higher values indicating higher confidence. The confidence returned by using these weights over an entire image set I can be expressed by taking the L_2 (Euclidean) norm over each input image:

$$g(I) = \left(\sum_{i \in I} g_i^2 \right)^{\frac{1}{2}}$$

These values w_s, w_c, w_t could be calculated using leave-one-out cross-validation or by implementing a SIMPLEX algorithm to solve a suitably-constructed linear program. Due to the increased time required by running three independent probabilistic classifiers and the additional implementation challenges, investigating this technique will be left for future work.

7.4.2.3.2 Algorithm

To generate a multi-feature bag of visual words model to produce an SVM classifier, the following algorithm will be employed.

Multi-Feature Bag of Words with SVM PIC Scheme
1. Partition into training and testing sets. 2. Extract SURF, colour and texture features using the described algorithms. 3. Construct three independent bags of visual words – SURF, colour and texture bag. 4. Encode training set images with each bag, generating a SURF feature vector, colour feature vector and texture feature vector of a set length for each. 5. Row-normalise these vectors with respect to each other, to ensure each carries equal weight. 6. Concatenate these normalised vectors to form a single multi-feature vector for each training image. 7. Train a SVM classifier using these labelled multi-feature vectors, where the labels indicate the correct class labels.

7.4.3 Diagram

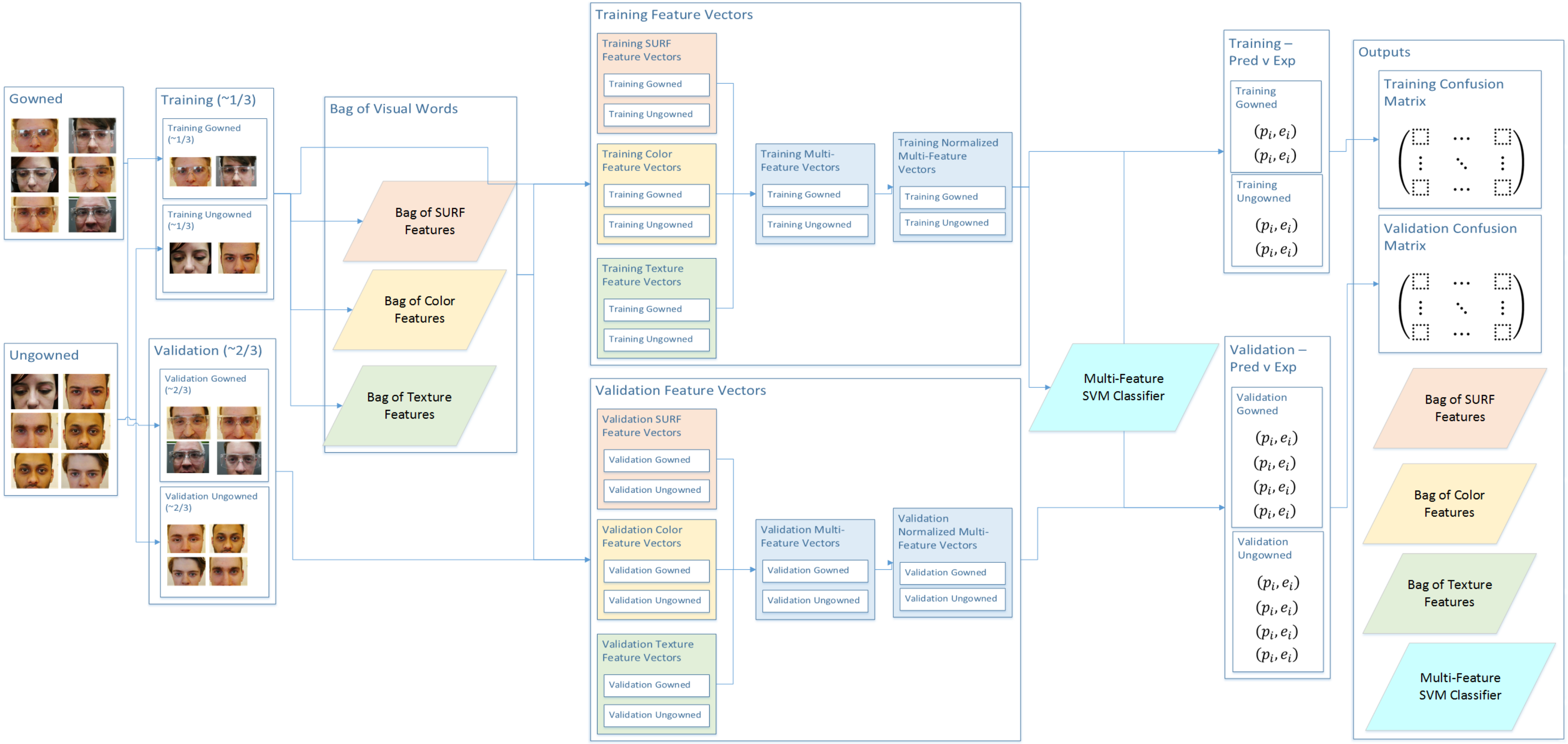


Figure 28: MBFS Solution Diagram

7.4.3.1 Developing a Test Framework

To demonstrate and evaluate the effectiveness of this technique, the algorithm has been implemented in MATLAB. In order to reduce running time, the MATLAB code has been compiled into a C program using the MATLAB Compiler with results reported to disk. The application also serialises the constructed bag of words and feature vectors, allowing them to be loaded at a later time.

In keeping with the previous two investigations, this final scheme has also been implemented in C# to prove its effectiveness on live Kinect data. As this code forms the basis of the eventual Kinect Gowning Application, it has been omitted from this section in favour of a fuller explanation later.

7.4.4 Preliminary Testing and Evaluation

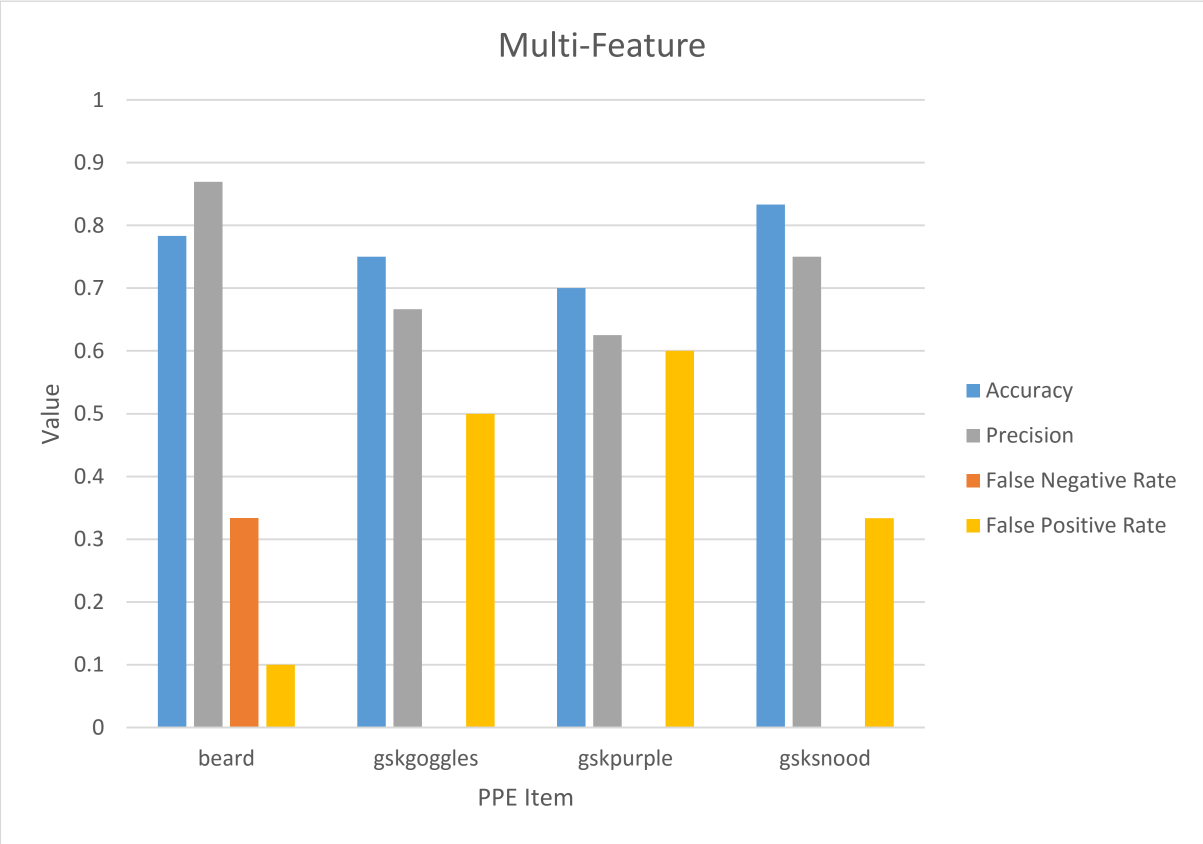


Figure 29: MBFS Preliminary Test Results

Unlike the previous attempts, the MBS PIC scheme has been able to generate strong classifiers for both goggles and gloves, showing its ability to cater for items that have no distinguishable colour or cause little shape change when worn. Although false negative rates appear to have decreased, the new scheme has not reduced false positive rates, which are the most critical from a regulatory standpoint.

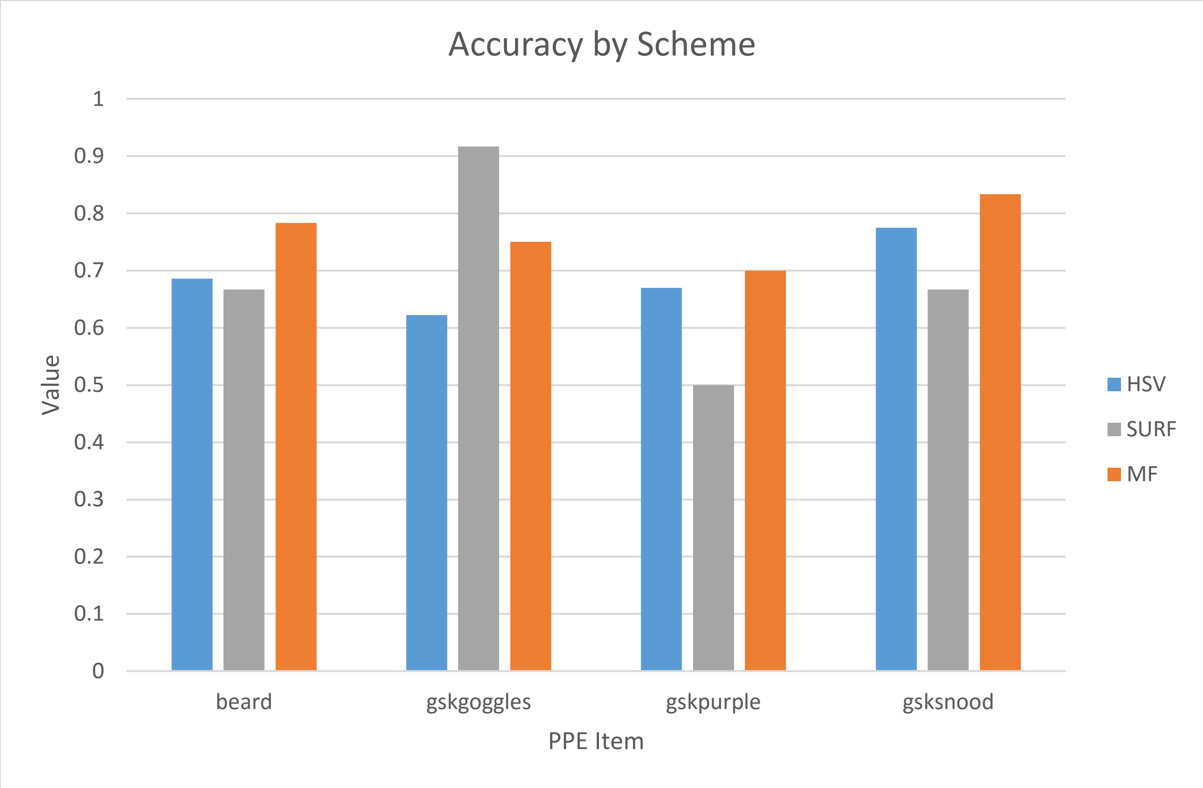


Figure 30: Preliminary Comparison of Schemes

In comparison to the other techniques, the multi-feature classification approach indicates a significant accuracy improvement on three of the four PPE items. However, the data does indicate a fault with the combined scheme, in that the accuracy level can in fact decrease in cases that the feature extraction techniques are in direct contradiction.

7.5 Max-Rule Bag of Visual Words with SVM Classifier

In order to ensure that the combined classifier can be no worse than those built on its constituent feature extractors, a final technique was developed that generates a SURF, colour, texture and Multi-Feature SVM and selects whichever indicates the highest accuracy rating when run on the training set data. The Multi-Feature and SURF SVMs are generated according to the described methods and the others are instead constructed by manipulating the SBS-scheme to apply the local colour or HOG feature extractor. Figure 31 shows the accuracy levels obtained by running the MBS scheme and SVMs built from the scheme's constituent feature extractors.

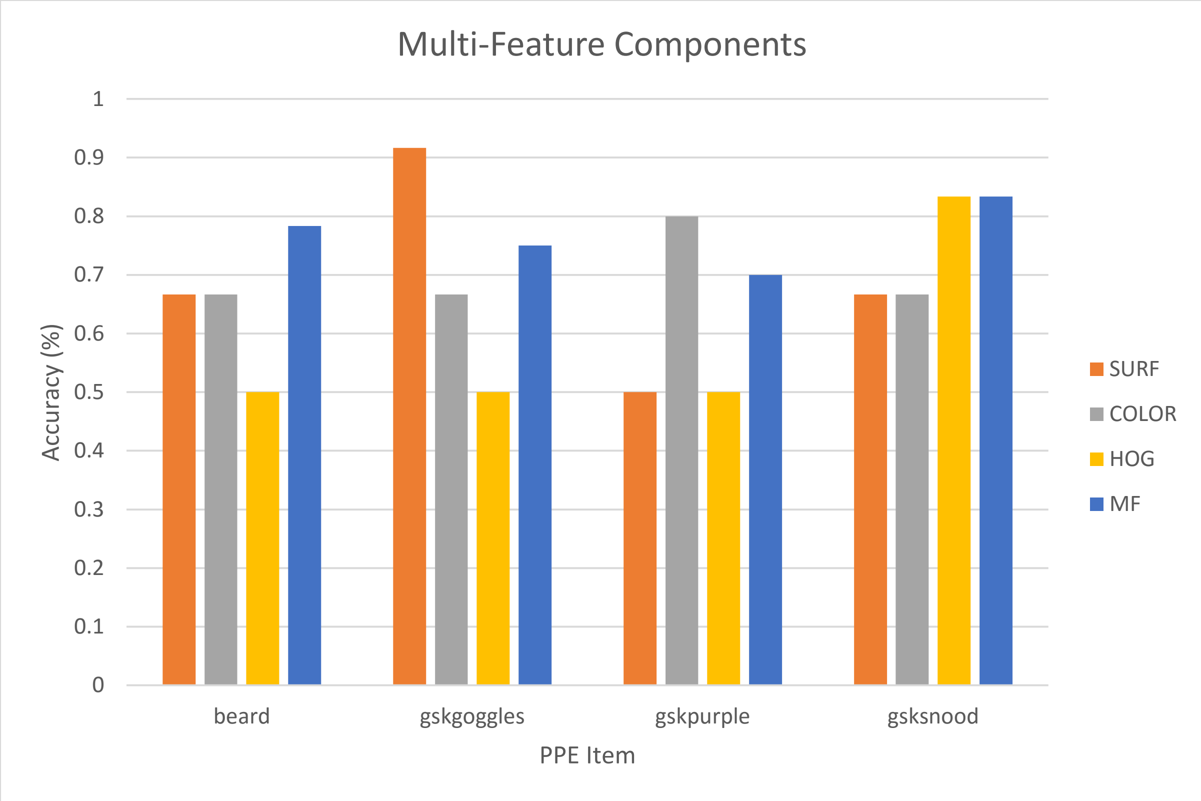


Figure 31: Multi-feature SVM Evaluation against Components

Figure 32 shows the result of employing the Max-Rule classification scheme, where the best-performing classifier was identified by running training set predictions and tested on the testing test.

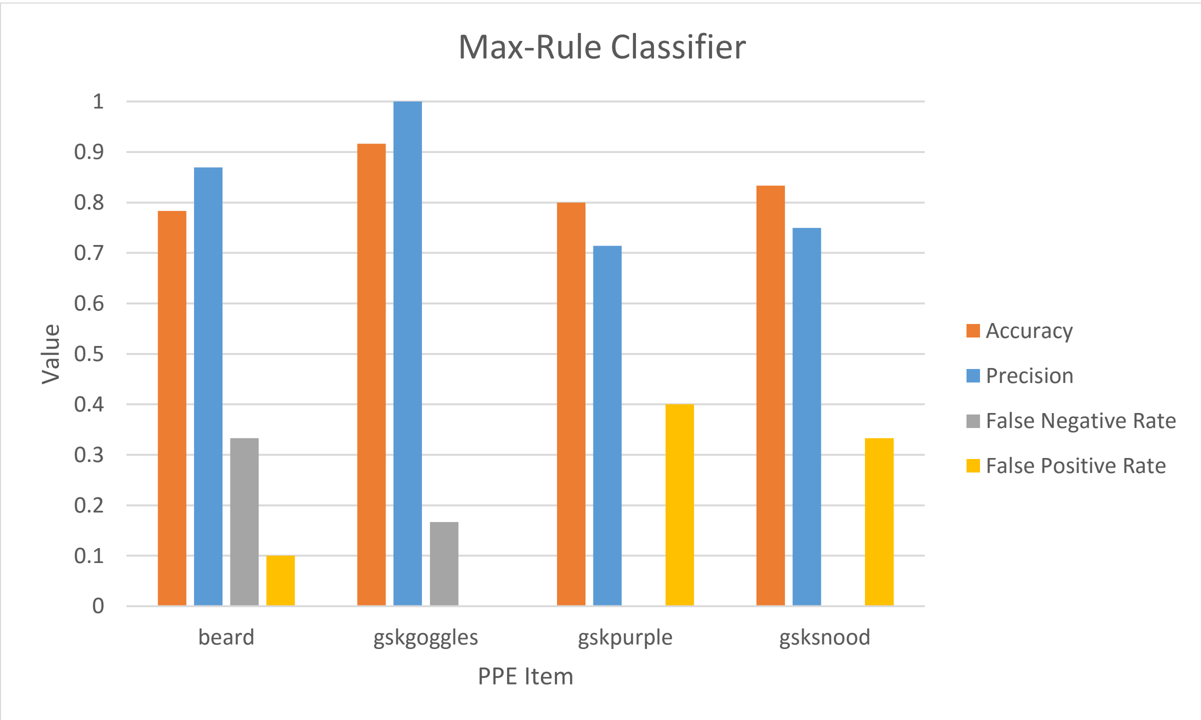


Figure 32: Max-Rule Classifier Test Results

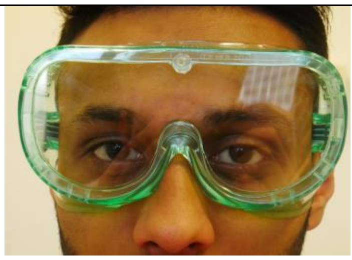
By virtue of the sufficiently-representative training set, the Max-Rule scheme has selected the classifier for each PPE item that happens to yield the best testing set accuracy. Due to the need to calculate three independent bag of words models and four independent classifiers, the Max-Rule training process is far more time-consuming. However, since the scheme simply selects a single classifier to use for each PPE item, the prediction process should take no more time. Since this training process need only occur once and could simply be left to run overnight, this overhead has been considered acceptable.

Although caution was held at this early stage due to the relatively small and non-exhaustive dataset, the preliminary investigation has demonstrated a number of viable PPE item strategies for formal evaluation.

8 Formal Testing and Comparison

8.1 Testing Strategy

In order to formally compare these techniques, a much larger dataset was collected that includes images of volunteers wearing all GSK Building 5 PPE items as well as some additional items. These new PPE items have each been used to demonstrate a particular challenge that otherwise would not have been tested. Additional ungowned images have also been captured for all necessary body regions. Table 13 shows these new items and the reason for their inclusion in the test procedure.

Item	Body Region	Picture	Comments
Gardening Gloves	Left/Right Hand		Gloves with texture information. Expecting difficulties distinguishing from background.
Yellow Gloves	Left/Right Hand		Prove the colour information distinguishes purple and yellow gloves
Dark Goggles	Glasses		Goggles with key colour information.
Green Goggles	Glasses		Highly reflective goggles may affect images.
Birthday Hat	Hat		Multi-coloured item should demonstrate benefit of local colour extractor over HSV.

Turkey Hat	Hat		Hat changes shape depending on which way around it is worn.
Slipper	Left/Right Foot		May be challenging to distinguish from work footwear.
Arsenal Shirt	Torso		T-shirt with pronounced colour and SURF information.

Table 13: Additional PPE Items and Justification

A MATLAB application has been developed that will be used to perform all testing, allowing the best scheme to be selected for a C# implementation to form the final application. Although some variation is likely to occur between the MATLAB test framework and a potential C# implementation, particularly with processing time, the differences should be in proportion and thus allow techniques to be compared.

The application produces evaluation statistics for each of the six PIC schemes, including the three additional schemes, which re-implement the SURF Bag of Visual Words with SVM classifier by instead using the colour and texture features and give rise to the Max-Rule scheme. By doing this, the accuracy of extractors can be compared that allow conclusions to be made about which features are most important for which PPE items. All schemes are run on each PPE item, trained using a random 30% training set partition and tested using the remaining 70% testing set partition.

Code	Scheme Name	Notes
HSV	HSV Thresholding	HSV threshold ranges were manually calculated/recalculated using the training set for all PPE.
SURF	SURF Bag of Visual words with SVM	Runs SURF extractor, creates bag of visual words and trains SVM.
COLOR	Colour Bag of Visual Words with SVM	Similar to SBS scheme, but uses local colour extractor rather than SURF.
HOG	HOG Bag of Visual Words with SVM	Similar to SBS scheme, but uses HOG extractor with calculated cell size rather than SURF.
MF	Multi-feature Bag of Visual Words with SVM	Concatenated normalised feature vectors from each of SURF, COLOR, HOG to create a combined classifier.
MAX	Max-Rule Classifier	Selects the best classifier from SURF, COLOR, HOG, or MF depending on training set accuracy results.

Table 14: PIC Schemes for Formal Testing

Barring a small amount of cross-over designed to aid efficiency, such as reusing SURF, colour and HOG bags and feature vectors when construct the multi-feature and thus Max-Rule SVMs, the training of each scheme requires a great deal of independent calculation. Considering this in combination with the heavily enlarged dataset, the application understandably has an extremely long execution time.

To speed up this process, the MATLAB Compiler (MathWorks, 2016b) has been used to compile the application into a C executable and is run on a high-specification machine initialised with the MATLAB runtime, hosted on Microsoft Azure (Microsoft Corporation, 2016h). The dataset has been indexed and attached to the machine using virtual disk technology and the uploaded test application has been set to run from a remote command. Using this cloud-hosted machine, tests run faster and are able to continue uninterrupted until they complete.

The results from each test will be used to indicate which PIC scheme produced the best overall performance over the PPE items. A cost-benefit analysis will also be considered to establish whether a scheme showing a non-negligible classification accuracy improvement comes at an unacceptable prediction time increase. False positive instances for each scheme have also been evaluated to indicate the frequency with which a subsequent application would have erroneously reported a gowned item. In contrast to false negative instances that occur when a gowned item is not recognised in a single frame, false positive instances are extremely dangerous as a combination could allow facility access to a person who has not successfully completed the gowning procedure.

Conclusions are also made to identify the most difficult PPE items to detect, to allow evidence-based recommendations to be made to companies that choose to use the system produced by this project. As previously highlighted, SURF, colour and texture performances have been evaluated on individual items to justify the original motivation towards adopting a multi-feature classifier.

8.2 Data Capture

A group of 16 volunteers was identified to take part in this project. Images were captured of the body regions in Table 15, in states indicated by their class name.

Body region	Class	Dataset Size
Beard	BEARD	34
	UNGOWNED	55
Glove	GARDENING	55
	GSKPURPLE	56
	YELLOW	54
	UNGOWNED	48
Goggles	DARK	62
	GREEN	68
	GSKGOGGLES	62
	UNGOWNED	34
Hat	BIRTHDAY	51
	GSKMOB	67
	TURKEY	74
	UNGOWNED	50
Shoes	GSKOVERSHOE	47
	SLIPPER	47
	UNGOWNED	47
Snood	GSKSNOOD	63
	UNGOWNED	34
Torso	ARSENAL	62
	GSKCOVERALLS	54
	UNGOWNED	47

Table 15: Enlarged Dataset for Formal Testing

In all cases, PPE item classifiers will be trained to recognise the difference between two classes: a single gowned class (such as GSKPURPLE) and the relevant ungowned class of images depicting the relevant body region wearing no PPE item. Since only the relevant PPE items for each gowning procedure should ever be accessible to the user, it is not necessary to include alternative items of the same type in the ungowned category. For example, when training a GSKPURPLE item classifier, the ungowned class need not contain images taken from another type of glove.

8.2.1 HSV Ranges

The HSV ranges presented in Table 16 were manually calculated for the HSV Thresholding scheme using the MATLAB Colour Thresholder.

	Min Hue	Max Hue	Min Sat	Max Sat	Min Val	Max Val	Threshold
arsenal	0.92	0.054	0.606	1	0.455	1	12.759
beard	0.025	0.2	0.358	1	0.209	0.372	1.382
birthday	0.04	0	0.74	1	0.192	0.898	3.87
dark	0.021	0.804	0.029	0.036	0.09	0.533	0.017
gardening	0.482	0.611	0	0.28	0.494	0.861	15.106
green	0.126	0.634	0.204	1	0.146	0.859	26.895
gskcoveralls	0.068	0.706	0	0.275	0.453	0.893	15.374
gskgoggles	0.129	0.978	0	0.282	0.596	0.793	2.877
gskmob	0.135	0.642	0.129	0.56	0.293	0.902	21.394
gskovershoe	0.987	0.667	0	0.151	0.54	1	2.281
gskpurple	0.686	0.932	0.248	0.805	0.169	1	11.846
gsksnood	0.086	0.618	0	0.375	0.462	0.742	8.318
slipper	0.386	0.724	0.134	0.664	0.251	0.64	5.589
turkey	0.073	0.17	0.382	1	0.472	0.847	12.278
yellow	0.103	0.198	0.507	1	0.428	0.903	5.7

Table 16: Manually-Calculated HSV Ranges

8.3 Comparison of Techniques

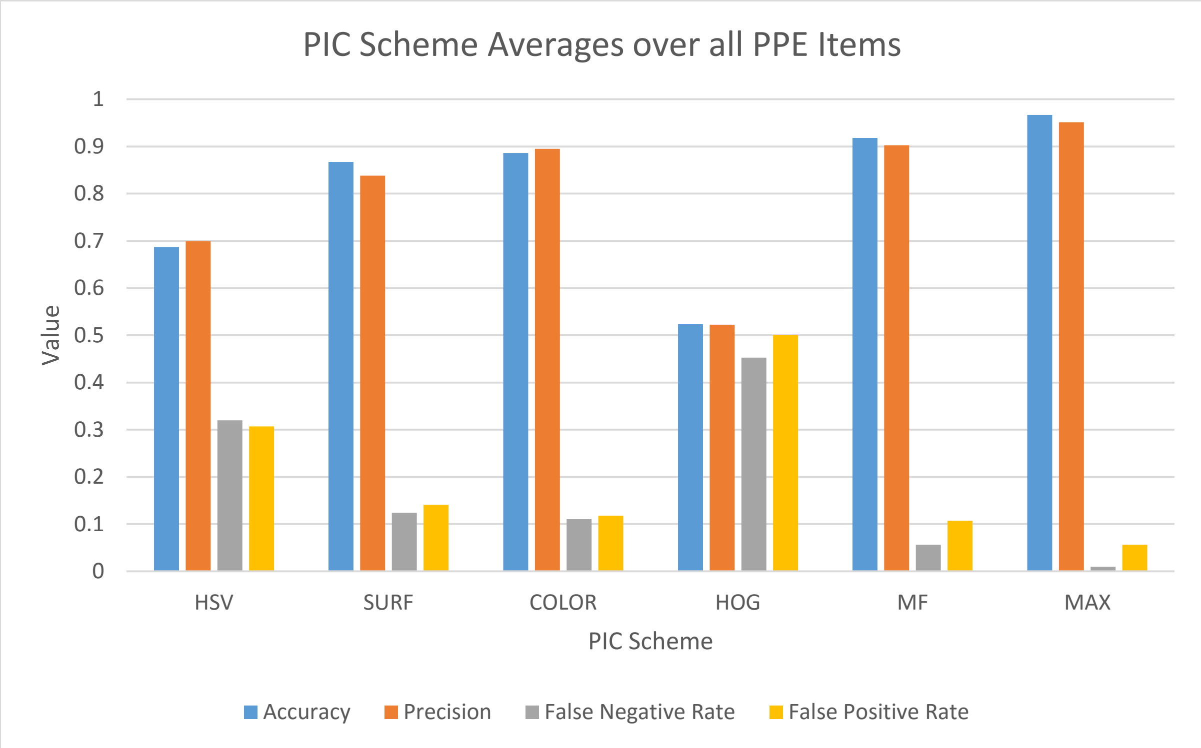


Figure 33: PIC Scheme Average Performance

Figure 33 shows the average performance statistics obtained by training and testing each PIC Scheme on each PPE item. Indicated by the final series (labelled 'MAX'), the data indicates that a remarkably high classification accuracy was obtained when the Max-Rule classification scheme was employed. The scheme yielded a 96.7% accuracy and 95.1% precision rating over all PPE items and an 0.95% false negative rate (occurring when gowned images are misclassified). Although the 5.7% false positive rate is the lowest among each tested scheme, it could still be considered too high for industrial implementation. To combat this, average classification values should be taken over multiple live frames before reporting the overall result. The more frames that the application considers, the higher the accuracy reading and, critically, the lower the false positive rate.

Shown by the superior values for the COLOR scheme over the HSV scheme, the data suggests that the automated training process for constructing local colour feature vectors outperformed the manual method for calculating HSV ranges. Although partially due to the difficulties encountered when filter functions were constructed for transparent or multi-coloured items, the HSV scheme's inferiority is also likely to have been caused by its inability to define where particular colours are located.

The HOG scheme classification accuracy results are shown to be the weakest among all considered schemes. It is likely that these poor results were caused by the necessity to increase the extractor's cell-size parameter in order to prevent the application causing a memory overload. In preliminary testing HOG values were shown to be useful, but their benefit could not be adequately demonstrated on such large datasets with limited equipment. Without testing a form of the multi-feature classifier that does not include HOG features, it cannot be claimed that the texture component decreases the overall classification accuracy. Despite its poor individual performance, it could be that the texture features contribute vital information on images that are systematically misclassified by the other schemes. In the absence of any indication that HOG features play a detrimental effect, they have remained included in the multi-feature and max-rule schemes, albeit with a smaller cell-size parameter setting.

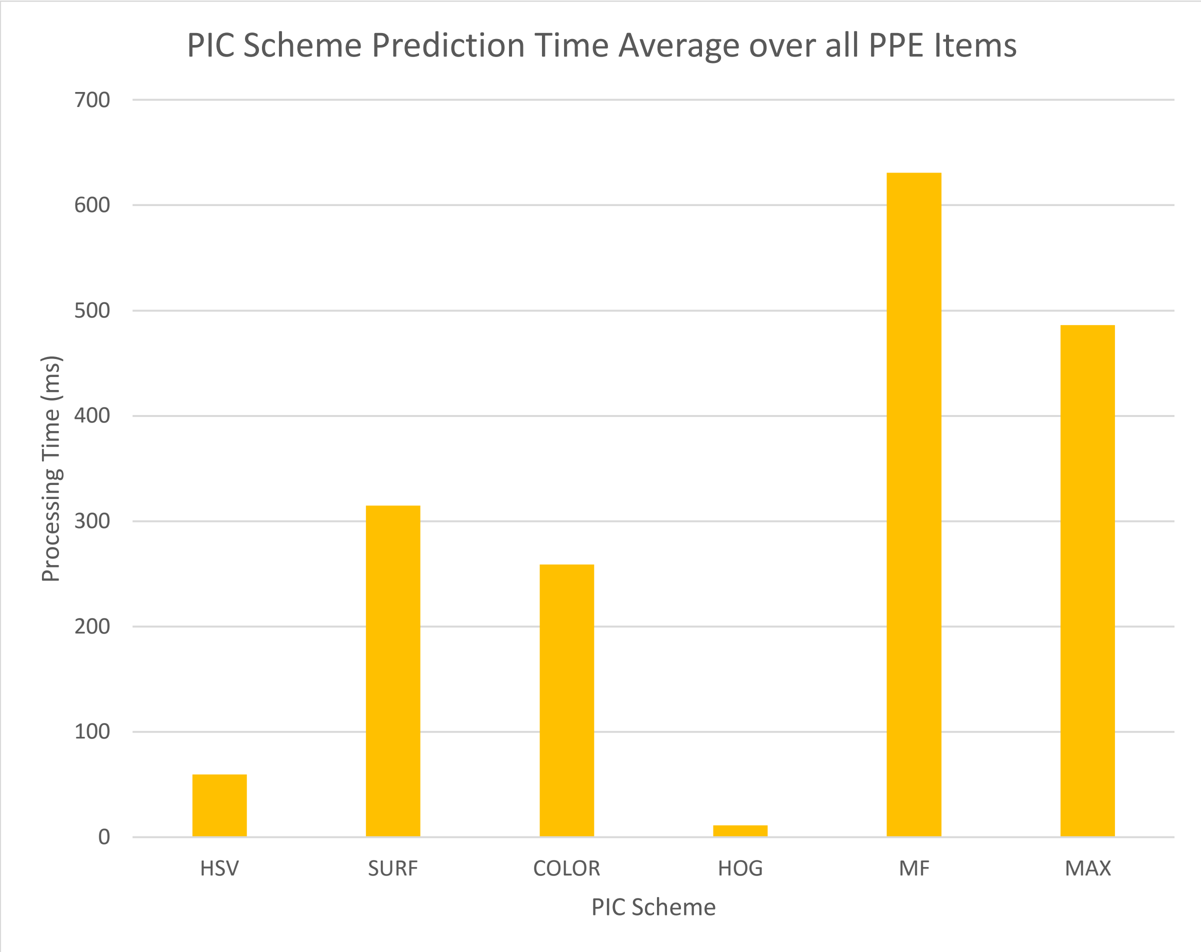


Figure 34: PIC Scheme Average Prediction Time

Since the Max-Rule scheme can only be as computationally expensive as the most time-consuming of the other schemes, the average prediction time across all items was unsurprisingly lower than that of the multi-feature scheme. However, since the multi-feature scheme was the most frequently used by the Max-Rule scheme, the processing time increased accordingly.

The prediction time data also supports the claim that the HOG extractor was run with an overly-elevated cell-size parameter. The rapid speed at which HOG-based predictions appear to have run over input images suggest that insufficient points were being returned for the construction of an accurate independent classifier.

8.4 Comparison of PPE Items

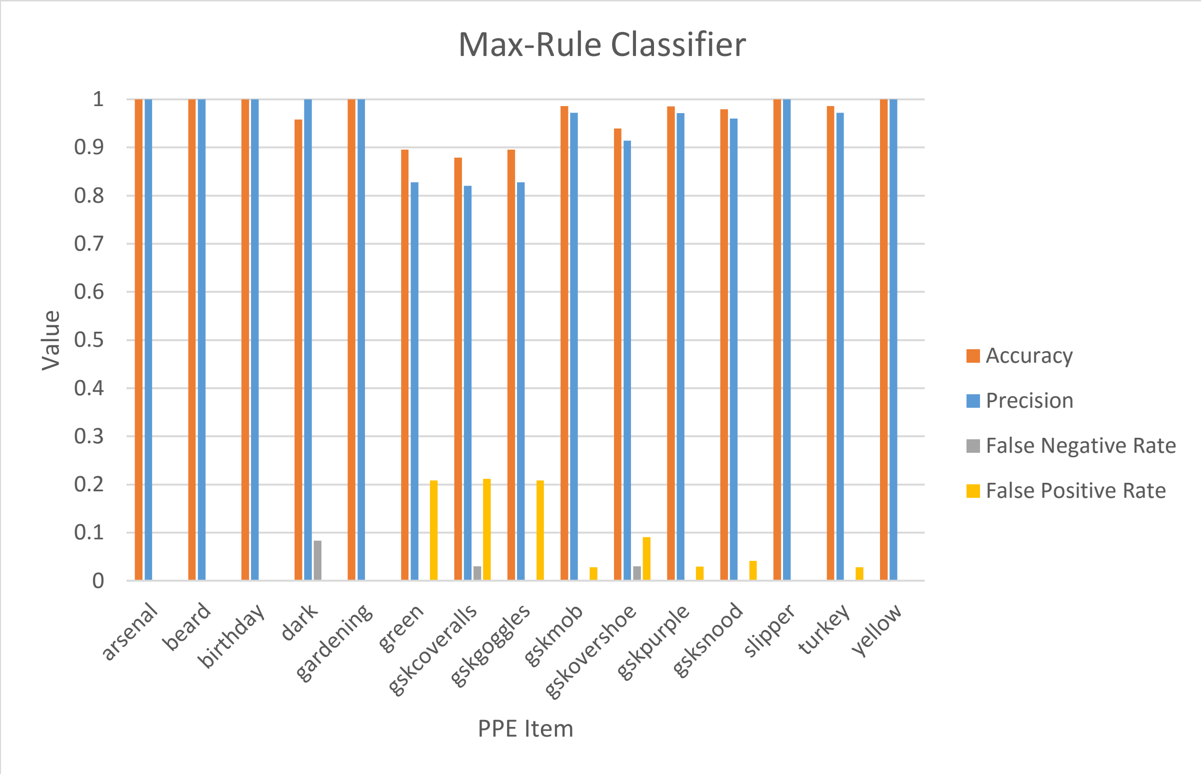


Figure 35: Comparison of PPE Items under Max-Rule Classifier

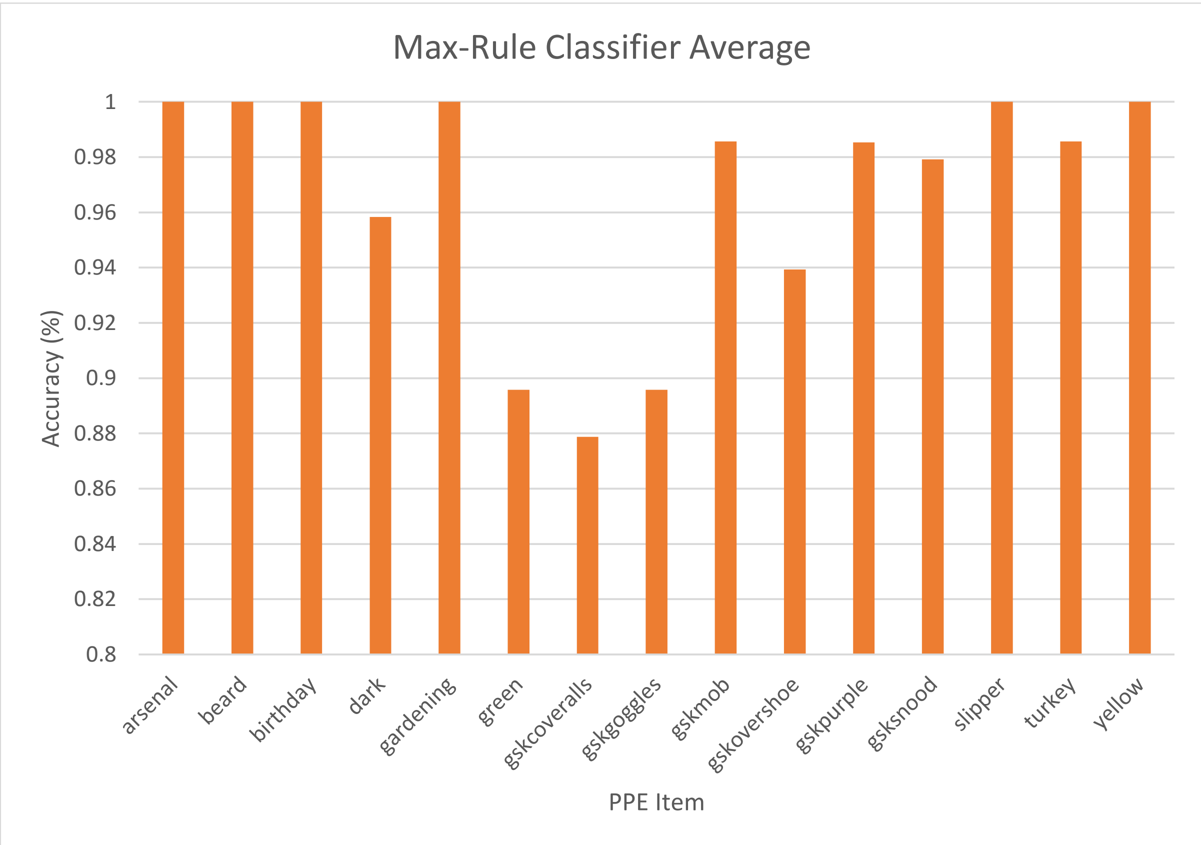


Figure 36: Accuracy Comparison of PPE Items under Max-Rule Classifier

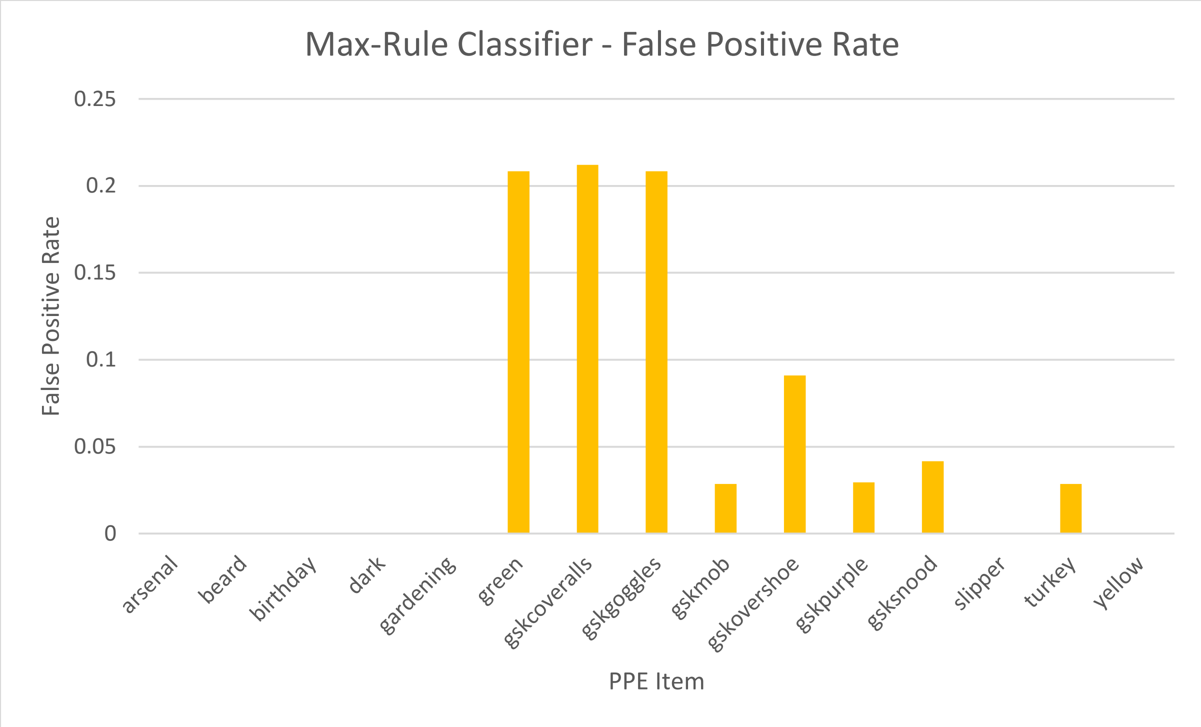
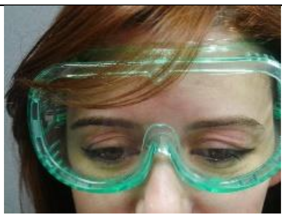
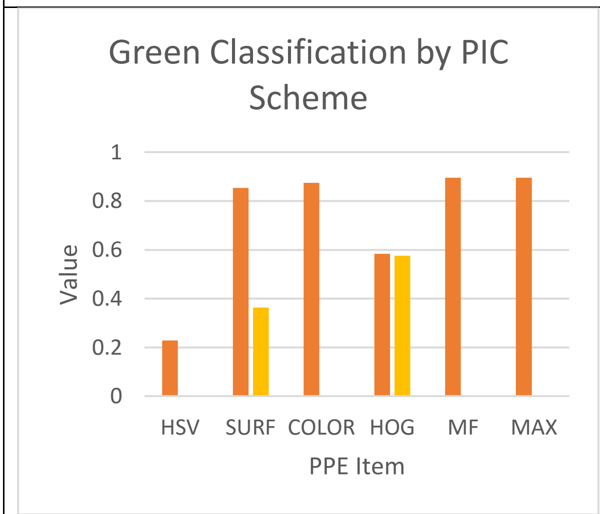
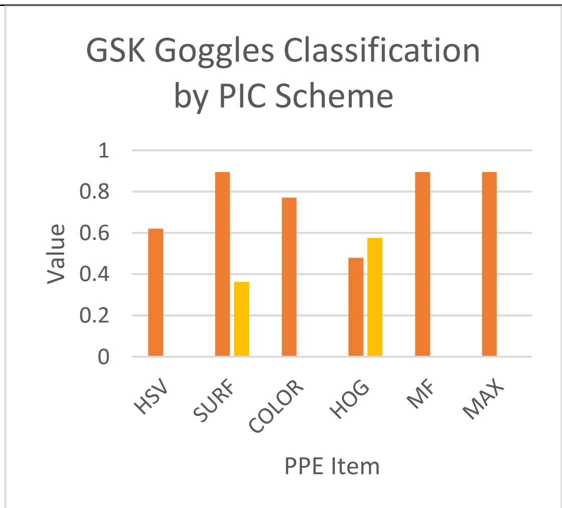


Figure 37: False-Positive Rate Comparison of PPE Items under Max-Rule Classifier

The data indicates that while an accurate (>95%) PPE item classifier was built for the vast majority of items, four items appeared to be the most challenging and also caused the highest false positive rates – at under 1%, the false negative rate has been considered negligible. These items are discussed in Figure 38.

<table>
<tr><th>GREEN (89.6%)</th><th>GSKGOGGLES (89.6%)</th></tr>
<tr><td></td><td></td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

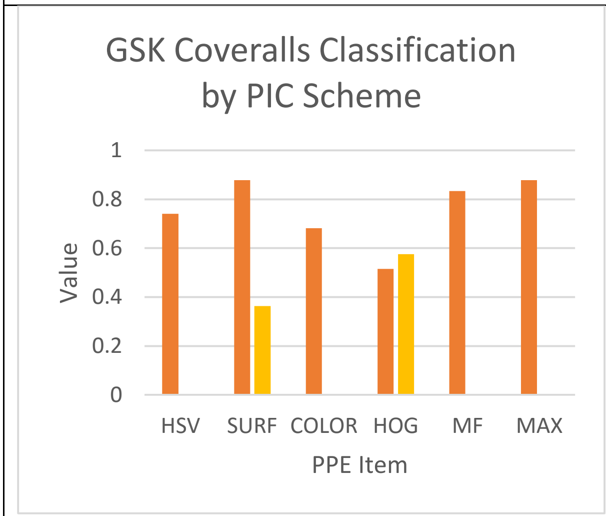
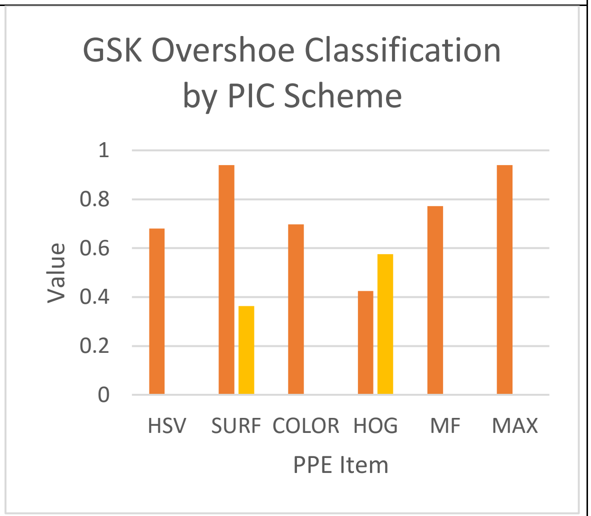
<table>
<tr><th>GSKCOVERALLS (87.9%)</th><th>GSKOVERSHOE (93.9%)</th></tr>
<tr><td></td><td></td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Figure 38: Most Challenging PPE Items Test Results

Although the multi-feature and Max-Rule schemes have been designed to compensate for items that show weak SURF, COLOR or HOG information, these items have each been shown to exhibit insufficient information across all three types.

For the green and GSK goggles, it is likely that the presence of head hair over the frames, reflective nature of the lenses and variation under skin-tone caused difficulty for both SURF and COLOR.

The shape of both GSK coveralls and GSK overshoes change depending on the user (note that the overshoes are put on over the wearer's original shoes) and, in addition to their lack of distinctive markings, have not provided strong SURF information. Moreover, since both items are almost uniformly white, variation in light and shadow locations between images are more significant. HOG features have also shown to be poor indicators for each of these items, most likely for the reasons already given.

To prevent these concerns causing classification errors during production use, this project's customers should be particularly advised against future procurement of plain, white items. To avoid transparency issues, such as those encountered with the green and GSK lab goggles, companies should also be advised to use tinted versions, such as safety versions of the dark sunglasses, or introduce distinctive frames to promote detection. In general, opting for vibrant, multi-coloured items with distinctive logos or other markings appears to produce the highest classification accuracy.

Although HOG features were originally chosen to support beard detection, Figure 39 instead shows the suitability of SURF features, which generated a 100% accuracy rate on an 89-image dataset.

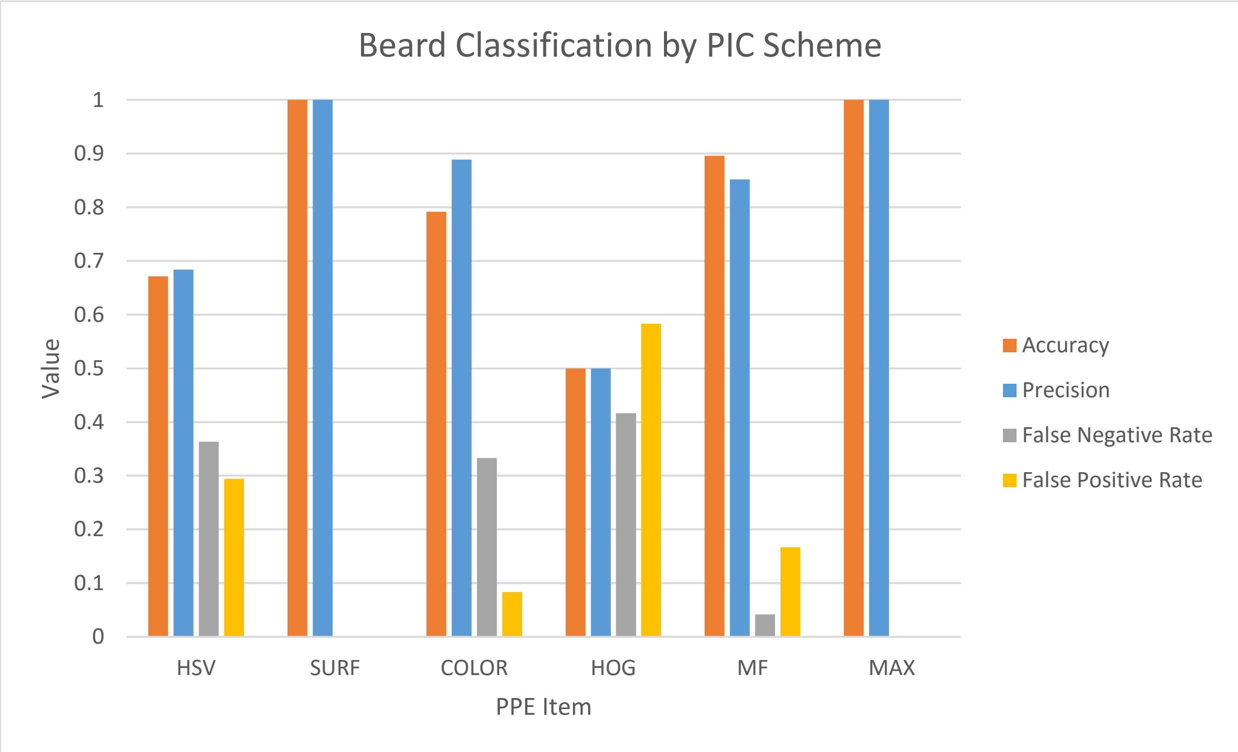


Figure 39: Beard Classifier Test Results

9 Kinect Gowning Application

9.1 Introduction

The final stage of this project involved building a complete Kinect-based implementation suitable for validation testing by the industrial sponsor. As per the conclusions made in this report's research section, the Kinect Gowning application uses the Max-Rule PPE classification scheme. The solution has been implemented in C# .NET and uses WPF as the design framework, as per the recommendation for Kinect-enabled software, and packages both training and recognition modules.

The following sections cover the design and functionality of each module and conclude with notable features, including the use of external image processing packages to train and use PPE item classifiers, .NET multi-threading techniques that ensure a responsive UI, and the serialisation framework used to define a standard layout for exported files. Pre-processing and gesture recognition techniques have been omitted from this section due to their prior explanation given in the 'Preliminary Kinect SDK Work' section of this report.

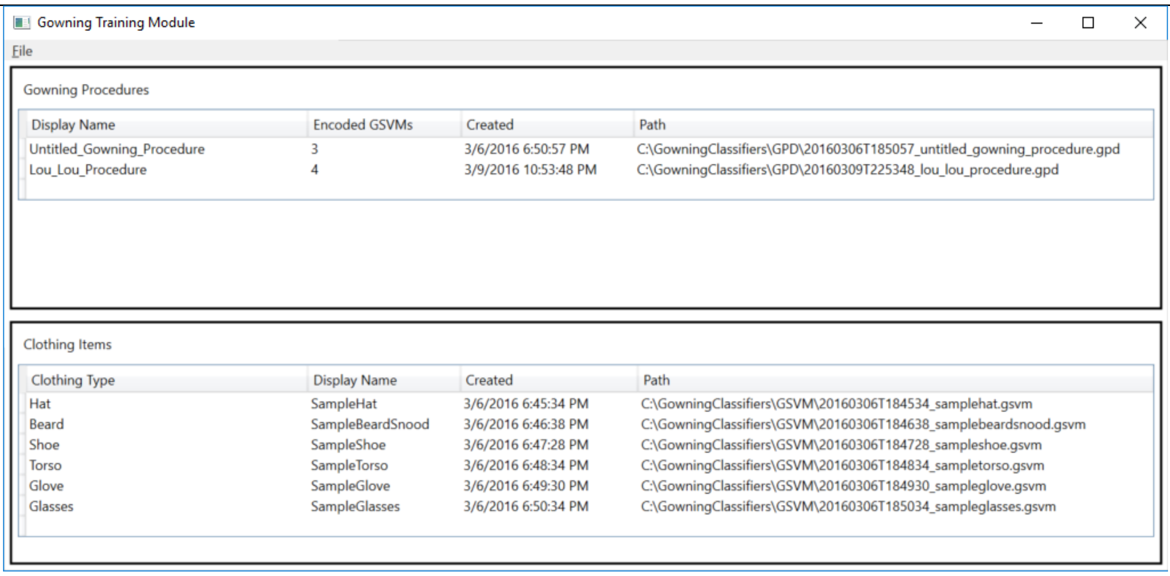
9.2 Training Module

Although the training module is not Kinect-enabled, it has similarly been implemented using WPF to maintain consistency with the recognition module. The training module has two main functions that are performed by the 'Train New Item' and 'Train New Gowning Procedure' windows, which are both described in Table 17.

Window	Implements	Functionality	Output
Train New Item	PIC training scheme	Allows a user to train a new PPE item according to the PIC training scheme.	Exports a GSVM file to disk that contains: • Multi-feature SVM (the item classifier) • Clothing Type • PPE item display name • Creation Time • File path • SURF BoW • Colour BoW • Texture BoW The SURF, Colour and Texture bag of words are included to allow files to be re-processed by the training module.
Construct Gowning Procedure	PGP training scheme	Allows a user to specify a new gowning procedure according to the PGP training scheme.	Exports a GPD file to disk that contains: • Ordered list of re-serialised GSVM files • Gowning procedure display name • Number of PPE items used • File path • Created time The GPD file is formatted suitably for interpretation by the recognition module.

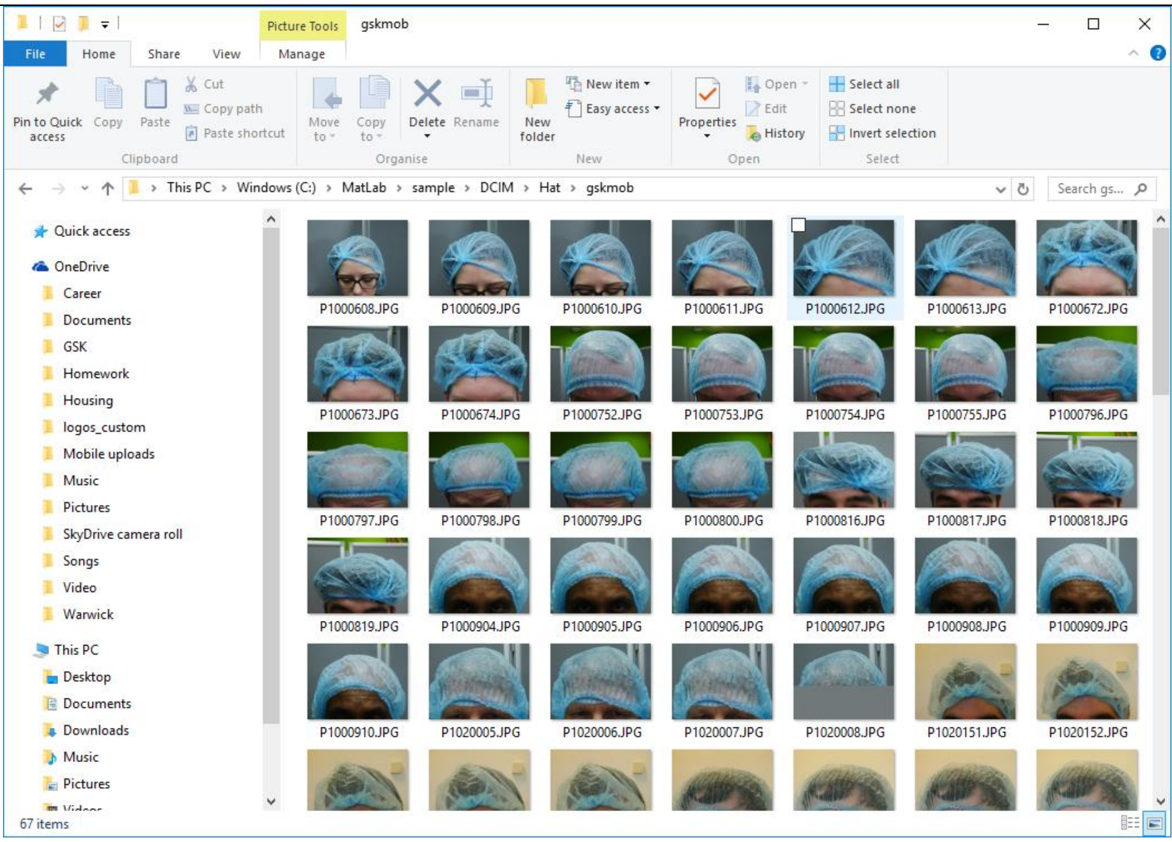
Table 17: Training Module Functions

9.2.1 PIC Training Scheme

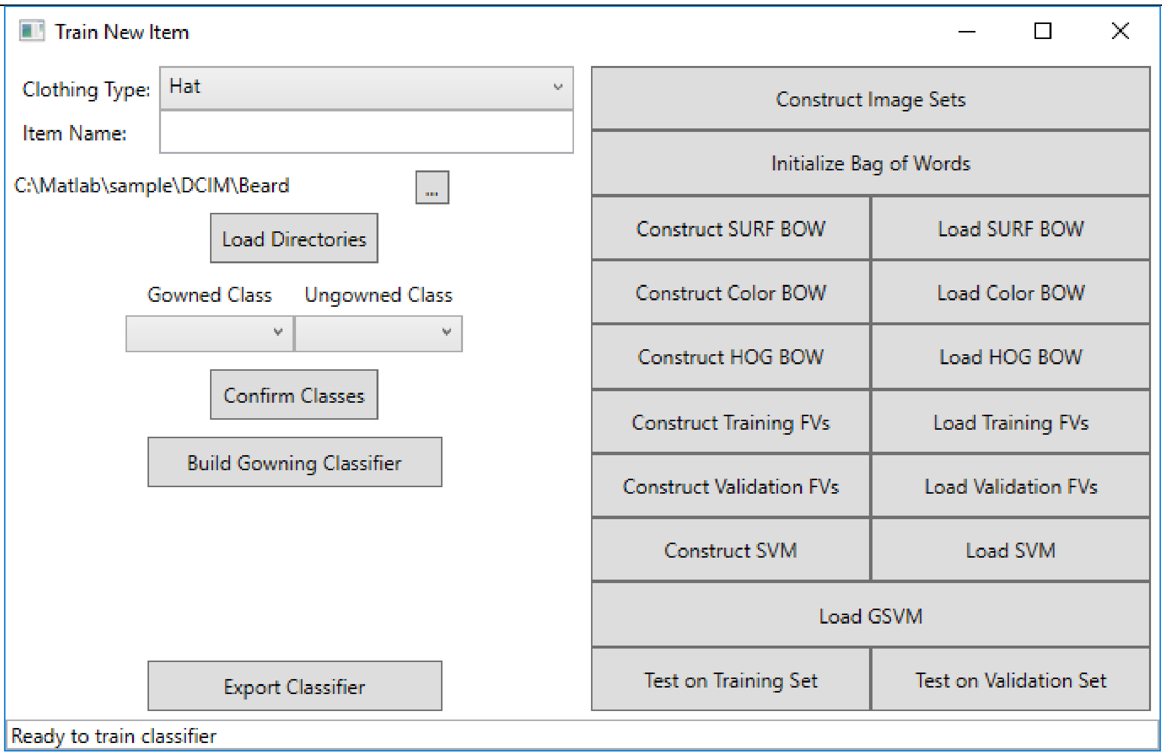


1. The overview page displays a table of pre-trained PPE items and gowning procedures by deserialising GSVM and GPD files found in the set output directory.

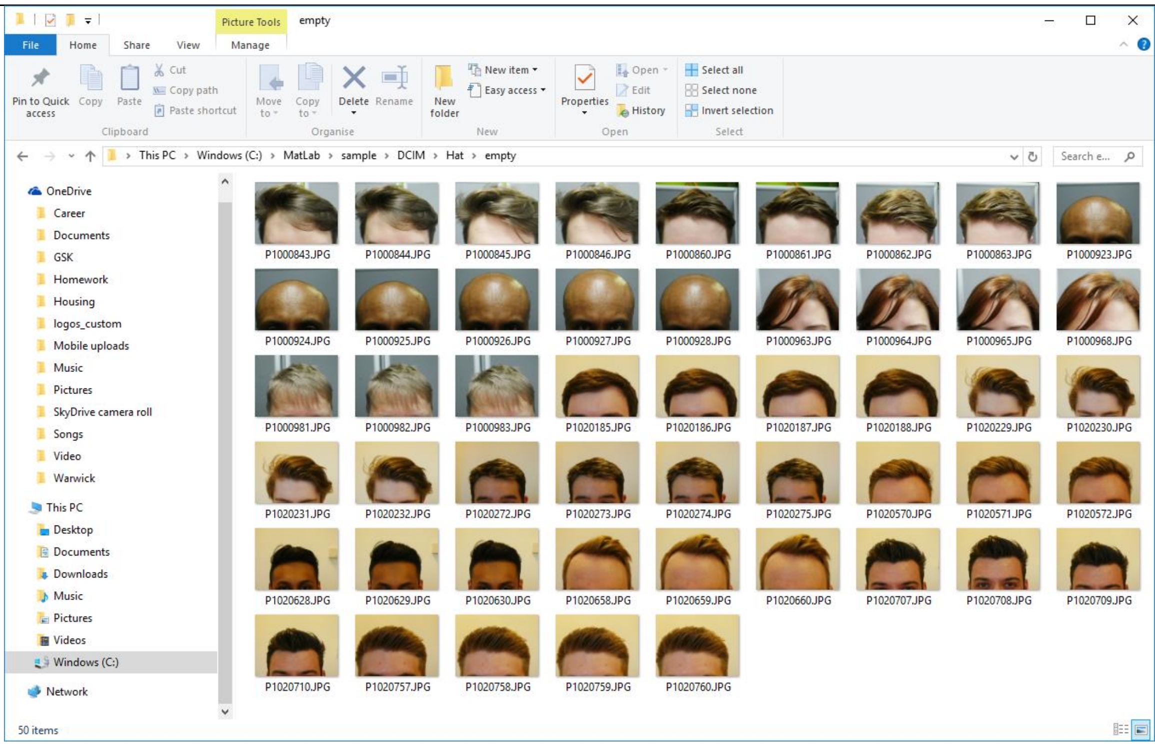
2. The 'Train New Item' window, accessible via File -> Train New PPE Item, can be used to generate a GSVM file for a new PPE that contains a generated item classifier and necessary metadata.



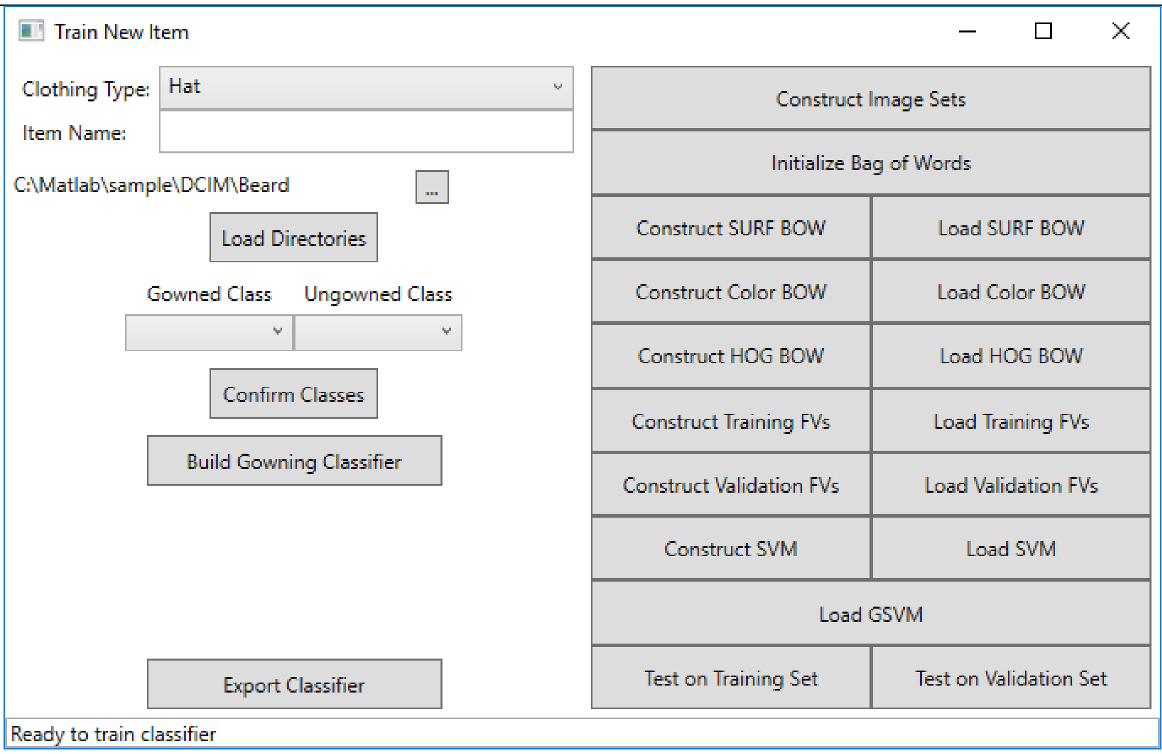
3. A directory must first be constructed that contains multiple gowned images from a number of volunteers.



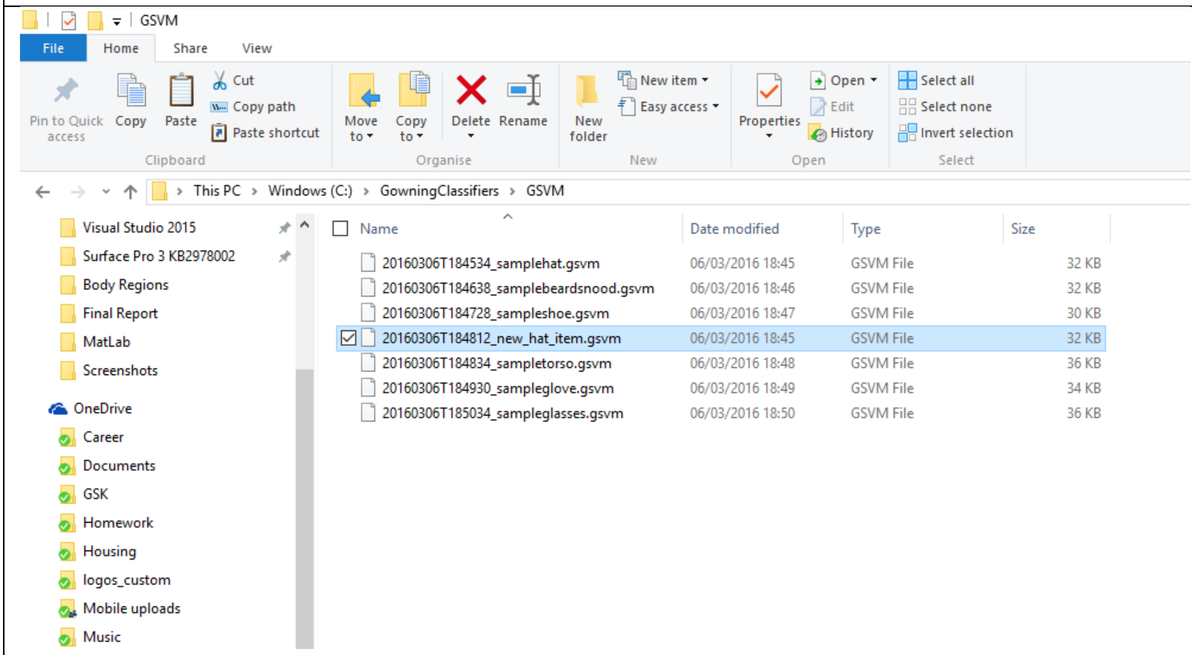
4. The user must then specify which type of PPE item is being trained, provide a unique name for the new item and set the gowned directory they constructed in the previous step.



5. This application then sets the ungowned directory for the appropriate body region. These sets were captured by the developer and are supplied with the training module.



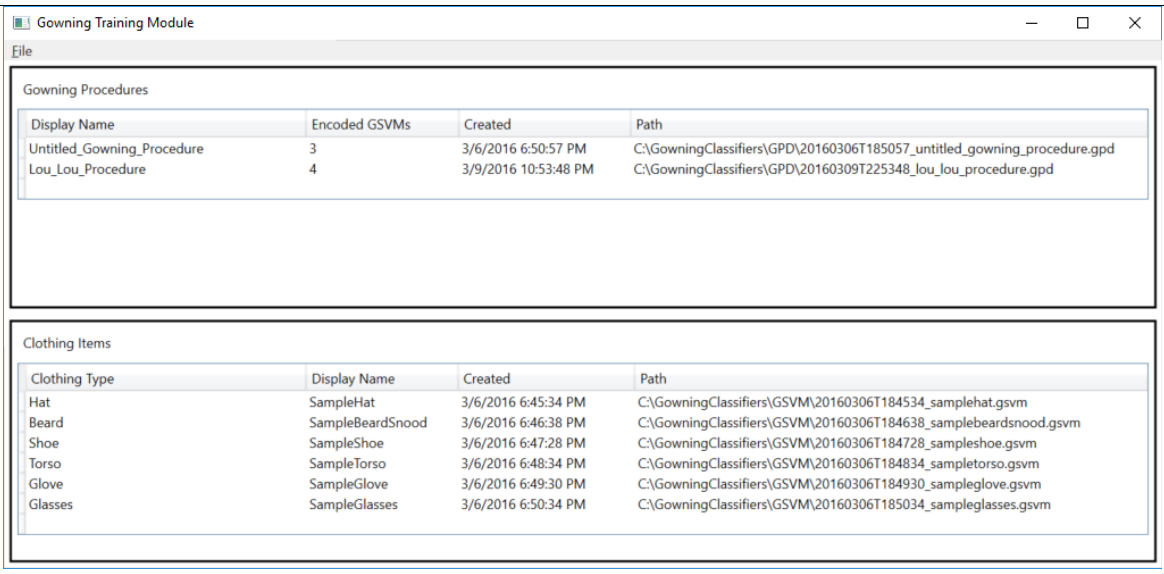
6. The GSVM file can either be constructed by selecting a sequence of the buttons in the right-hand column or simply by clicking 'Export Classifier'. The right-hand column allows an advanced user to 'save' or 'load' their progress midway through training.



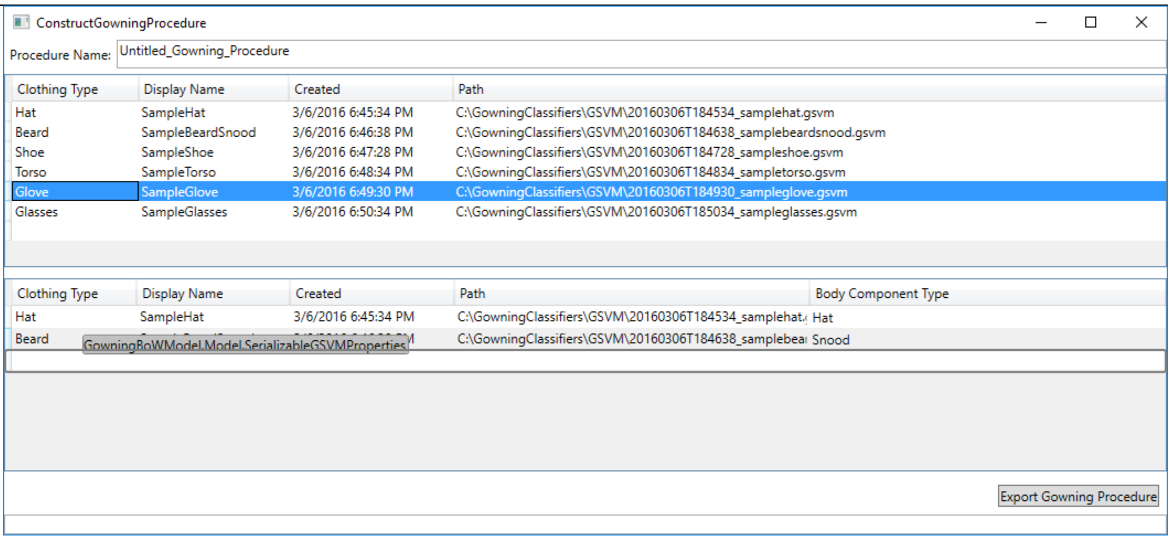
7. The file is then serialised and saved to the nominated GSVM output directory.

Table 18: PIC Training Scheme Implementation

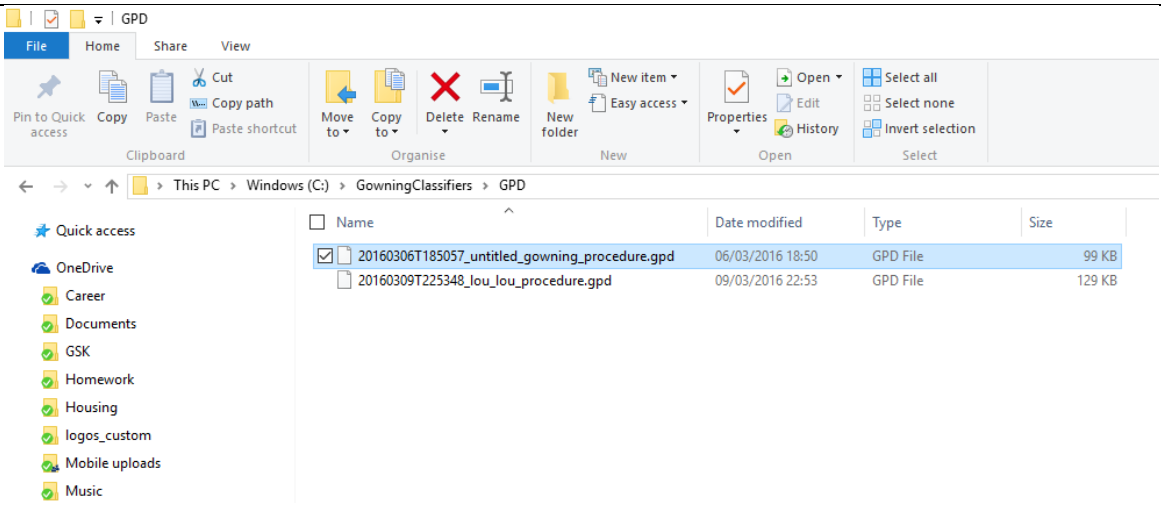
9.2.2 PGP Training Scheme



1. The ‘Train New Gowning Procedure’ window, accessible via File -> Train New Gowning Procedure, can be used to export a new sequence of recognition steps that the recognition module can use to perform live verification.



2. The pre-trained items listed in the top box can be dragged to the bottom box to form a new procedure. Since PPE items can appear many times during the same procedure, each can be dragged more than once. To indicate which body region should be inspected when detecting each PPE item, the user must specify a Body Component Type from each dropdown list.



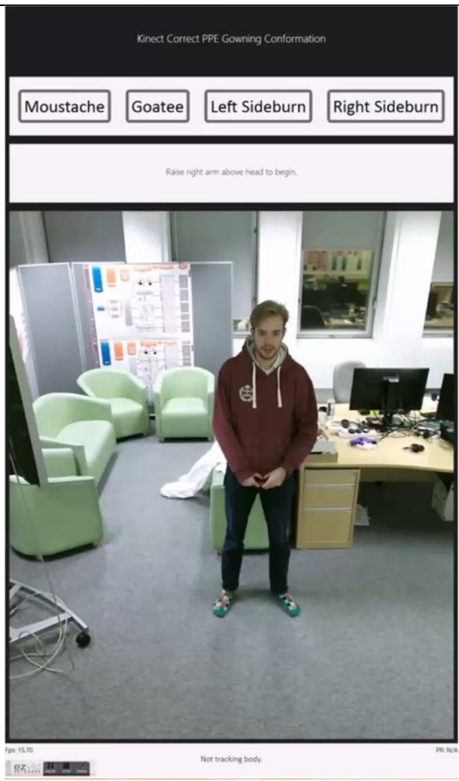
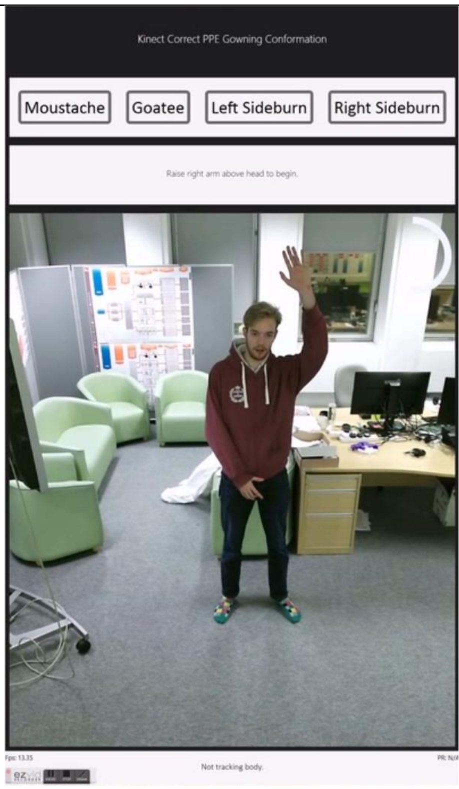
3. The Export Gowning Procedure button first runs a validation routine to ensure duplicate body regions do not appear, alongside standard exception handling such as checking the procedure is non-empty.

4. The GSVM file for each selected PPE item is then re-serialised in the specified order alongside various necessary information, such as the procedure's name and each PPE item's selected body component type.

Table 19: PGP Training Scheme Implementation

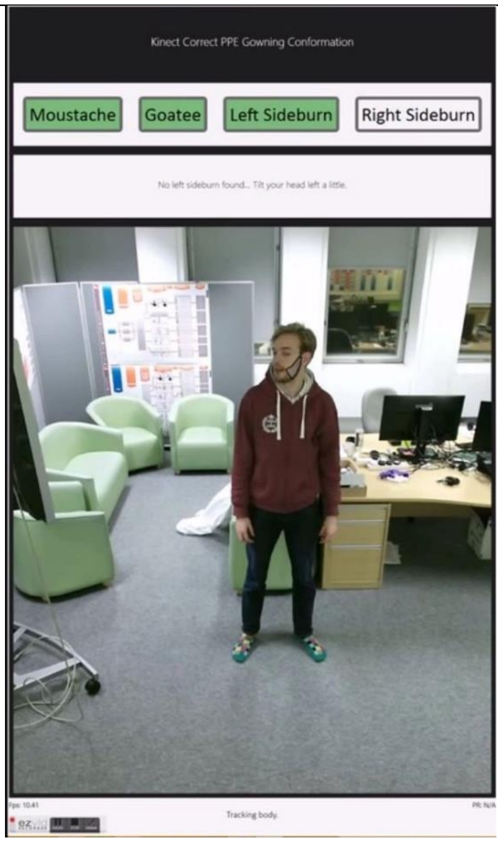
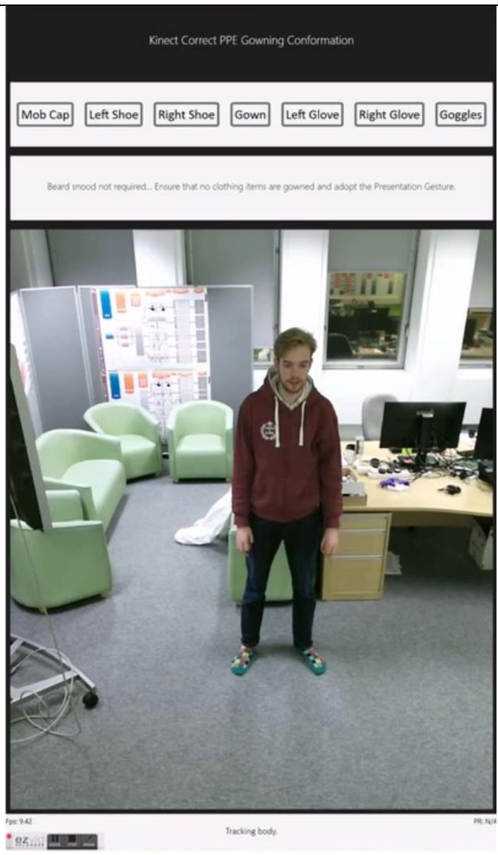
9.2.3 Recognition Module

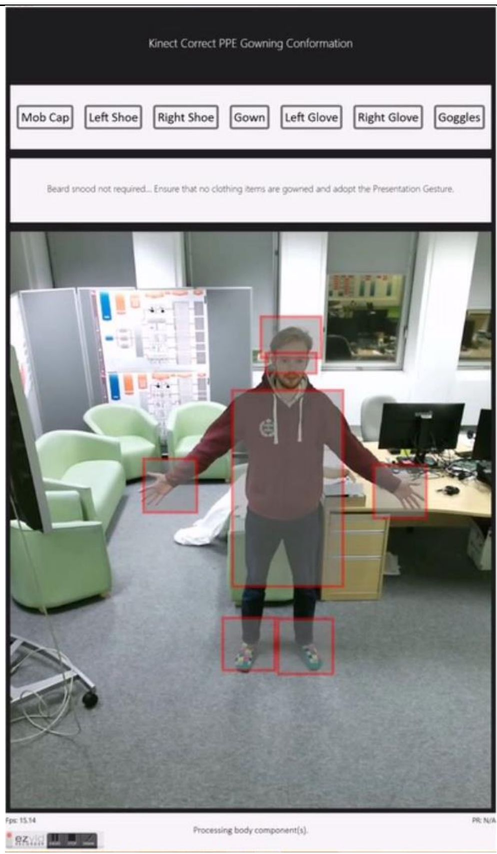
Table 20 shows the steps involved in the recognition module.

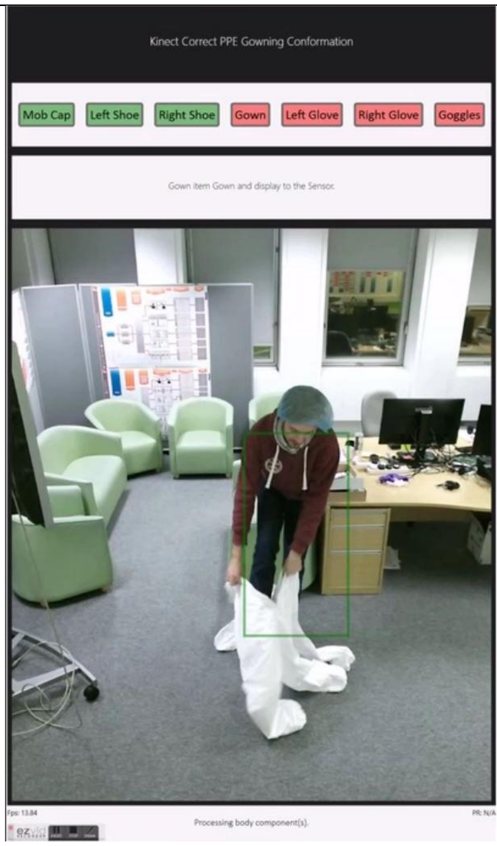
Initialisation. The recognition module must first be targeted at the gowning procedure file which defines the gowning process for this high-risk facility. When the recognition module is initialised, this file is first deserialised and interpreted to construct the process that will be followed.	
User Identification. User is asked to adopt a particular gesture that is recognised by the gesture detection algorithm described in the preliminary work section. Note that the white circle monitors the progress of the gesture to require that the user elevates their arm for three seconds to trigger the event.	

<table>
<tr>
<td>Beard Detection (if the procedure contains a beard snood item). If the procedure contains a beard snood item, the beard detection process will be run (otherwise this step is omitted):

Beard detection GSVM file (part of installation directory) read in.
Applied to each part of the user’s face – predictions run for each part.

If beard is not detected, remove the beard snood from this user’s procedure.</td>
<td>
</td>
</tr>
<tr>
<td>Update UI. The application’s UI is then updated to reflect the procedure defined in the target GPD file. Note that a beard snood would be omitted from this list if a user was found to not have a beard.</td>
<td>
</td>
</tr>
</table>

<table>
<tr>
<td>Ungowned Presentation. The application waits until the user adopts the presentation gesture and simultaneously applies every GSVM classifier extracted from the GPD file. The application moves on when all classifiers return an ‘ungowned’ result, to ensure the user begins the procedure in a completely ungowned state.</td>
<td>


</td>
</tr>
<tr>
<td>Sequential Gowning. Each body region is analysed in the order specified by the GPD file. The item classifier extracted from the GSVM file for the relevant PPE item is run against each pre-processed input frame.</td>
<td>


</td>
</tr>
</table>

Gowned Presentation. The user is again asked to adopt the presentation gesture, to allow the application to ensure all relevant classifiers return a positive result, indicating that no gowned item became ungowned mid-procedure.	
Access Granted. Having determined that the user has correctly completed the specified gowning procedure, the application provides access to the facility using a supplied API.	

Table 20: Recognition Module Implementation

9.3 Notable Features

9.3.1 External Packages

Due to the wide-ranging academic and industrial use of the OpenCV processing package (Itseez, 2015), work begin using EmguCV (Emgu CV Foundation, 2015) – a popular .NET wrapper. Unfortunately, an alternative had to be identified after encountering a number of implementation bugs that prevented the wrapper from processing high-volume data. The MATLAB .NET interface (MathWorks, 2015b) was also considered as a processing option which would have allowed the existing MATLAB code to be called from the C# Kinect Gowning Application, although a number of online articles suggested efficiency concerns (MathWorks, 2009) Although not as significant, a MATLAB implementation would also have required the installation of the MATLAB runtime, causing a slight deployment overhead.

Accord.NET (Souza, 2015) has been used due to its native .NET implementation, support for generically-defined bag of words objects alongside standard SURF and HOG versions. The colour extractor required careful implementation, making use of the Magick.NET package (CodePlex, 2016) to perform the necessary image manipulation and the ColorMine package (GitHub, 2015) for non-trivial LAB conversion.

9.3.2 Efficiency Consideration

In order to maintain a responsive user-interface during the processing-intensive prediction algorithm, the .NET background worker object is used to run these tasks asynchronously (Microsoft Corporation, 2015a). Figure 40 shows the high-level description of this scheme.

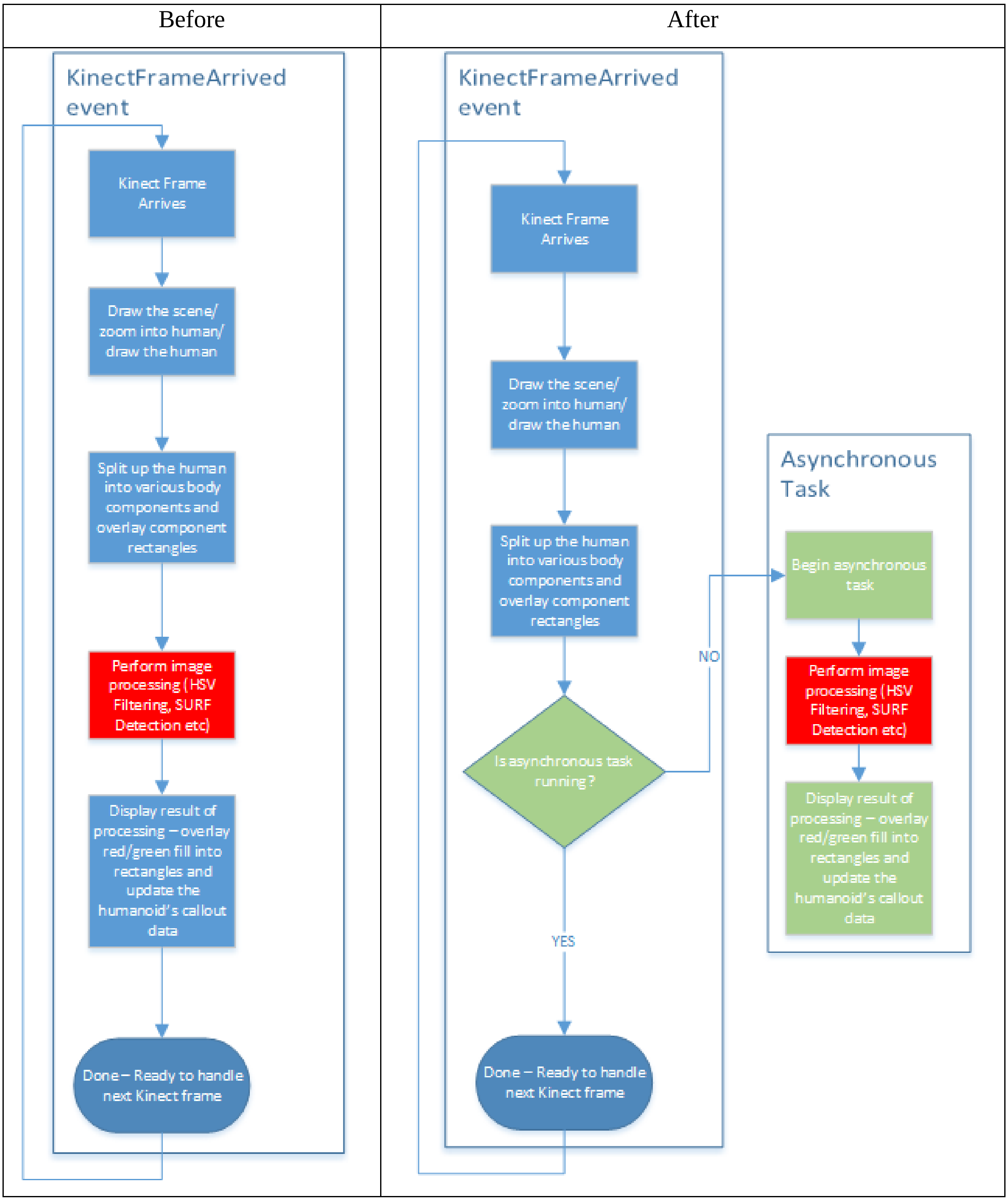


Figure 40: Asynchronous Image Processing (Before/After)

9.4 Kinect Gowning Application Solution Testing

9.4.1 User Acceptance and Validation Testing

At the time of writing, the Kinect Gowning Application is undergoing a formal validation process by the industrial sponsor. Initial demonstrations and trial runs have been encouraging, although formal details cannot be provided until the validation report is returned.

9.4.2 Evaluation of General Requirements

The initial project requirements can be evaluated according to the Kinect Gowning Application's known capabilities:

R1. Application should be able to recognise specific PPE items when worn by a human subject. **(Critical)**

- Application must prove to be extremely accurate at accepting/rejecting users' attempts to perform the relevant gowning procedure.
    - ***Met** - Recognition Module applies appropriate SVM classifier to Kinect frames that feature a live gowning procedure.*
- Application should restrict access to the high-risk facility until the operator has correctly followed the specified gowning procedure.
    - ***Met** - ReleaseDoor function of the industry-supplied API is executed during the 'Gowning Procedure Completed' event, releasing the facility door.*

R2. Application should be able to verify the order in which the PPE items were gowned. **(Critical)**

- ***Met** – User interface only moves on when current PPE item has is detected.*

R3. Application should include a training module that allows a user to teach the system to recognise new gowning procedures. **(High)**

- This will involve providing the user with a mechanism for teaching the application to recognise PPE items and define new orderings on known PPE items.
    - ***Met** - Training module applies the Max-Rule scheme to gowned and ungowned images to create GSVM item classifiers. New orderings are defined by selecting a sequence of pre-trained items, and exported to a GPD file.*
- The training module should be suitably designed for an unspecialised user.
    - ***Met** - Training module allows user to target a directory of images taken with unspecialised equipment and provides a button that performs the entire process.*

R4. Application should include a mechanism for identifying if a user has sufficiently dense facial hair to deem the beard snood necessary. **(High)**

- This will be used when validating procedures that include a beard snood. If a user does not have a beard, the beard snood should not be required.
    - ***Met*** *- Beard detector constructed and integrated with the recognition module.*

R5. The recognition part of the application should not require human contact in its general operating mode. **(High)**

- ***Met*** *- Gesture-based user interface allows completely contactless recognition module.*

R6. Recognition module must be suitably designed for large, high-resolution display in a portrait orientation. **(Medium)**

- ***Met*** *- WPF techniques used to allow resizing and display in both landscape and portrait orientations.*

R7. Recognition module must provide the user with continuous feedback describing the next item to be gowned or reporting an error in the existing sequence. **(Medium)**

- ***Met*** *- Feedback displayed in the top bar of the user interface. Violations are reported as they are encountered.*

R8. Recognition module must be responsive and maintain a 'live' passive feel. **(Medium)**

- ***Met*** *- Asynchronous techniques used to maintain efficiency.*

R9. Recognition module must be extremely simple to use. **(High)**

- ***Met*** *- Simple instructions given to user, who simply carries out existing procedure.*

R10. Recognition module must not significantly increase the time or effort required by an operator attempting to access the lab. **(High)**

- ***Met*** *- Aside from adopting entrance and two presentation gesture poses, the user performs their gowning procedure as normal. Time increased by less than a minute.*

R11. The project should remain within a reasonable budget. **(High)**

- ***Met*** *- Complete Kinect set-up purchased for under £200, remaining within budget.*

R12. Application must not impose a significant overhead cost to an adoptive company. **(High)**

- ***Met*** *- No mandatory modification required although recommendations are made over future equipment purchases.*

10 Conclusion

This project has met its objectives by constructing an automated process to verify that user correctly follows the in-place gowning procedure before entrance to a high-risk facility is permitted. The application also contains a training module, allowing the introduction of new PPE equipment specification of new gowning procedures by a lab administrator who may not have specific computer vision skills. Finally, to provide a quantitative measure of whether a user would wear a beard snood, should one be defined by the procedure, a beard detection module has been constructed by employing the Max-Rule scheme.

10.1 Evaluation of Project Outcomes

10.1.1 Research Phase

This project has demonstrated a research phase in which several PPE item classifier training schemes were considered to find a generic, optimal process for generating a series of classifiers where each determines the presence of its associated PPE item on images that depict the relevant body region. The techniques were demonstrated by MATLAB implementations and evaluated with bespoke live Kinect .NET software. Formal testing was then conducted by executing a compiled MATLAB application that encompassed six variations on these schemes providing comparable results. The result of this procedure indicated a strong benefit to the Max-Rule approach, satisfactorily confirming the original motivation for combining multiple features.

A great success of this project has been the discovery of a visual approach that requires no mandatory modification to the existing PPE, which otherwise would have resulted in a huge overhead cost to any adoptive company. However, by analysing the performance of each PPE item, evidence-based recommendations can be provided to companies wishing to adopt the procedure when they next procure their PPE items.

10.1.2 Final Implementation

The Kinect Gowning Application produced in the final stage of this project demonstrates a viable technique for regulated companies to ensure that they meet their obligations under EU law. The produced application's training module is suitable for GSK and external organisations to teach the application to recognise their specific equipment and define an order in which this equipment should be put on. The process is well-designed for an unspecialised user, who can invoke a complex process by pointing the application to a directory containing photographs of their new item as worn by multiple volunteers. The gowning procedure file that this process exports is suitably formatted for the accompanying recognition module, which validates a live gowning procedure to ensure an operator wishing to gain access to the facility has first put on the proper clothing in the correct order.

The recognition module is also able to determine whether a user has facial hair of sufficient length to require a beard snood, should one be included in the procedure, by running a bespoke beard detector. By integrating this component, the system provides a measure of whether a user has sufficient facial hair to pose a risk to the product, replacing the previous self-determination process that was often unfollowed. By making intelligent design choices, the application uses contactless HCI techniques to ensure suitability for 'dirty zones' with an identified contamination risk. Asynchronous techniques have been implemented maintain a high frame rate, yielding a live, observatory feel by performing processing-intensive work on an independent thread of execution.

10.1.3 Industrial Standard Testing

As detailed in the previous section, the developed Kinect Gowning Application has been shown to satisfy all listed requirements that were set through a negotiation phase involving the project supervisor and the industrial sponsor. The application has now been submitted to GSK for an industry-standard validation process whereby the given testing and accuracy figures will be analysed alongside consideration over the application's stability and robustness. Although not a quantitative measure, the application continues to receive a great deal of praise during an ongoing user acceptance testing procedure and in executive-level demonstrations.

10.2 Evaluation of Project Management

Despite a successful project outcome, a number of re-orderings were made to the initial timeline of identified objectives. As expected, a number of these modifications occurred when new strategies or techniques were identified, but a significant number were caused by other commitments and tasks being made of the developer that were not apparent at the project's conception. Initial, mid-term and final timelines, in the form of Gantt charts, are shown in Appendix B – Timelines.

A major challenge during the project was the differing requirement sets given by the industrial sponsor and those required by the academic institution. To satisfy the industrial sponsor's need to demonstrate a progressing version of the application in an executive-level meeting, a secondary application was independently developed during the initial stages of this project that fully implemented the inferior (but simpler) HSV Thresholding technique. Although occasional challenges that nonetheless would have required solving were considered and resolved during this process, as an implementation of a provably inferior technique, the fully-operational C# .NET HSV Thresholding application did not significantly feature in this report. Despite the time that was depleted by constructing this robust and reliable application, the executive demonstration resulted in the guarantee of further sponsorship to cover the remaining term, allowing the continued use of company-purchased equipment.

A significant adaptation to the project was the decision to conduct the entire research phase in MATLAB. After tackling a number of .NET packages in an effort to construct a framework to learn advanced image processing techniques, the often obscure naming conventions and frequent software bugs led the developer to turn to a more commonly-used learning tool. Although the test-development cycle now involved an extra stage, progress was far quicker due to the substantial online MATLAB support and code examples.

10.3 Author’s Assessment of Project

Despite initial concerns towards the feasibility of conducting a complete computer vision research project alongside a robust and production-ready Kinect Gowning Application, I am delighted to have produced a body of research and a full implementation that demonstrates a feasible solution to a number of unsolved and ambitious challenges. By following a coherent investigative strategy and determinately proving the feasibility of each new idea with informative, demonstrative applications, the project has been able to produce a single algorithmic scheme that enables both custom clothing detection and facial hair recognition. By careful solution and hardware consideration, the produced application has remained within budget and demands no substantial overhead cost to an adoptive customer.

This project’s engagement with a supportive industrial sponsor has been of significant benefit. Due to the initial commitment towards the negotiated requirement set, this project has remained delivery-focused and determined to include all features necessary to conclude with an implementable solution. Through this partnership, the necessary equipment was procured at no cost to the academic institution and dataset volunteers could be easily identified through internal email communication. By virtue of a willing customer, the Kinect Gowning Application will benefit from ongoing user feedback during demonstrations in a replica changing room fitted with a full high-resolution monitor in the set portrait orientation.

A number of major challenges were overcome during the research and development of this project. In particular, the lack of prior experience towards image processing techniques, developing Kinect-enabled software and using WPF as the overall design framework imposed a steep learning curve on almost every major aspect of this project. Although supported by Dr Abhir Bhalerao’s CS413 Image and Video Analysis in-depth lecture series (Bhalerao, 2015), a great number of techniques covered – particularly the use of bag of visual words for dataset standardisation – are complex and were therefore left to the penultimate chapter of the module. The concepts and justification of dataset classification techniques taught in Dr Theo Damoulas’ CS342 Machine Learning module (Damoulas, 2016) have also been influential towards the testing and explanatory sections of this report, although the academic timetable did not provide this teaching during the research and design phase. Despite the challenges, proactive research and frequent correspondence with both module organisers enabled the development of a well-designed classification scheme and a robust testing framework.

Owing to the relatively sparse use of commercial Kinect development, it has been challenging to find online documentation or forum examples other than those included in the ‘Samples’ section of the SDK download. To give further indication, an official Kinect v2 SDK book does not exist at the time of writing. It has also been challenging to get to grips with the WPF as the recommended design technology. Although frequently heralded as a fantastic tool, the stark differences to most other UI packages impose a steep learning curve to new adopters. Although every effort was made to conform to the overarching design paradigm, time restrictions would not allow proper investigation into some advanced techniques which may have shortened the codebase, such as implementing user control dependency properties.

A number of previously-unexplored .NET features were also covered, including the use of *async await* for constructing well-architected asynchronous routines while dealing with standard multi-threading challenges (such as cross-thread communication) and employing .NET's rich serialisation framework to enable the GSVM and GPD files to be exported.

Through conducting this project, I have been able to develop a number of technical and management skills that have been of substantial benefit to my future career aims. I have been fortunate enough to demonstrate an initial release candidate in a number of executive-level meetings and have featured details of the project's progression in numerous career applications. I strongly believe that these successes owe a great deal to the initial decision to adopt an ambitious project with an even more ambitious requirement set, following a considered and flexible approach towards the research strategy and by ongoing engagement with the host institution and industrial sponsor.

Having absorbed a great deal of my life for a number of months, I am delighted with the result of this project. The skills and knowledge I have gained throughout this endeavour in project management are invaluable and continue to benefit my approach to investigative research. At the time of writing, I have agreed to lead an industrialisation process for this project, which will target external vendors to construct a multi-camera version for use in a new, purpose-built laboratory at GSK Stevenage.

10.4 Future Work

The following section describes a set of potential enhancements that could be made to training and recognition modules.

10.4.1 Training Module

10.4.1.1 Data Capture

Although the training module has been made suitable for an unspecialised user through its ability to construct item classifiers from standard RGB images, further benefit could be achieved by designing a separate 'Training Data Capture' window. A user could then invite gowned volunteers to stand in front of the same Kinect Sensor used for the recognition module and allow the application to construct the necessary training set directories by employing the pre-processing technique to capture images of each gowned body region from each volunteer. After the final volunteer is processed, a suitable PPE classifier for this item could then be automatically trained.

As well as providing a much faster and more user-friendly training process, the captured images would be standardised and bare stronger resemblance to those that the produced trained item classifiers are subsequently required to categorise. Owing to the increased processing time, adopting a high-performance, cloud-based model for asynchronous classifier generation may also be valuable, although would likely come at a maintenance overhead.

Should this new process be implemented, the training module should also have an uplifted design with similar stylistic choices to those chosen for the recognition module.

10.4.1.2 Image Processing and Machine Learning Techniques

As longer-term objectives, further work should be invested in investigating extra image feature representations that may benefit the selected Max-Rule training scheme. As an initial suggestion, should the adapted Kinect-based training procedure be developed, it would be possible to generate an infrared image representation to indicate a measure of a PPE item's reflective index, which may enhance the accuracy of the texture classifier. Consideration should also be put towards the additional texture mapping techniques cited in (Yang & Yu, 2011). In particular, the wavelet classifier has been shown to be particularly useful.

Although hypothesised during this research, further techniques for combining multiple feature representations should be investigated and evaluated against the same image dataset. A number of alternative supervised models, such as neural networks, random forests etc. could be considered alongside a number of ensemble learning techniques. This report has previously discussed the potential benefit of replacing the SVM with a suitable probabilistic model to generate a customisable weighting to each feature representation, and this could certainly be pursued further.

As a final suggestion, a similar approach might be adopted to that taken by the Netflix company, by publishing a detailed dataset to a machine learning forum (such as Kaggle), and offering a prize to anyone who can offer a higher classification accuracy to the solution offered by this report.

10.4.2 Recognition Module

It is likely that accuracy improvements could be obtained by adopting a multi-hardware recognition solution, such as the combined Kinect and Light Field camera set-up discussed in the 'Potential Solution Designs' section of this report. Through this adaption, extremely high-resolution images of each body region could be used for recognition, although significant effort would be required to coordinate the Kinect joint information and the Light Field camera feed. Although this technique was not covered by this project, the industrial sponsor has shown willingness towards the future purchase of enhanced hardware for use in the planned industrialisation phase.

Through another extension, it may be possible to incorporate automatic verification of non-gowning instructions defined in the document for standard operation procedure SOP-PDK-0012. Through designing a ‘movement gesture’, it may be possible to determine if a user has correctly stepped across the cross-partition before gowning their shoes. This has not been achieved in this project owing to the likelihood that the user’s movement over such a partition would require the development of a multi-sensor solution, potentially requiring multiple development machines, owing to the SDK’s one-Kinect-to-one-PC restriction.

Bibliography

Betaface, 2015. *Betaface API.* [Online]
Available at: http://betafaceapi.com/
[Accessed 20 02 2016].

Bhalerao, A., 2015. *CS413 Image and Video Analysis.* [Online]
Available at: https://www2.warwick.ac.uk/fac/sci/dcs/teaching/modules/cs413/
[Accessed 24 November 2015].

CodePlex, 2016. *Magick.NET image manipulation.* [Online]
Available at: https://magick.codeplex.com/
[Accessed 01 01 2016].

Csurka, G. et al., 2004. Visual Categorization with Bags of Keypoints. *Workshop on Statistical Learning in Computer Vision,* 1(1), pp. 1-22.

Damoulas, T., 2016. *CS342 Machine Learning.* [Online]
Available at: https://www2.warwick.ac.uk/fac/sci/dcs/teaching/material/cs342/
[Accessed 18 01 2016].

Ebrahimzadeh, R. & Jampour, M., 2014. Efficient handwritten digit recognition based on histogram of oriented gradients and svm. *International Journal of Computer Applications,* 104(9).

Ellis, D., Sommerlade, E. & Reid, I., 2009. *Visual analysis of pedestrian motion*, Oxford: University of Oxford.

Emgu CV Foundation, 2015. *Emgu CV Documentation.* [Online]
Available at: http://www.emgu.com/wiki/index.php/Documentation
[Accessed 1 November 2015].

Fire Product Search, 2016. *RescueFlex.* [Online]
Available at: http://fireproductsearch.com/products/firefighter-clothing/technical-rescue-suits/rescueflex/
[Accessed 26 04 2016].

GitHub, 2015. *ColorMine.* [Online]
Available at: https://github.com/THEjoezack/ColorMine
[Accessed 02 01 2016].

GlaxoSmithKline, 2014. *Dress and Conduct for Entry into IMO Secondary Pilot Plants, Building H89 & H92 Harlow & Building 5 Ware and GDS Packaging, Harlow, IMS UK (SOP-PDK-0012).* s.l.:s.n.

Horvath, M., 2010. *Wikimedia Commons.* [Online]
Available at:
https://commons.wikimedia.org/wiki/File:HSV_color_solid_cylinder_alpha_lowgamma.png
[Accessed 20 04 2016].

Itseez, 2015. *OpenCV API Reference.* [Online]
Available at: http://docs.opencv.org/2.4/modules/refman.html
[Accessed 3 November 2015].

Lazebnik, S., Schmid, C. & Ponce, J., 2005. A Sparse Texture Representation Using Local Affine Regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 27(8), pp. 1265-1278.

Lytro, 2015. *What is Light Field?.* [Online]
Available at: http://blog.lytro.com/post/132599659620/what-is-light-field
[Accessed 27 02 2016].

MathWorks, 2009. *MATLAB Newsgroup.* [Online]
Available at: http://uk.mathworks.com/matlabcentral/newsreader/view_thread/259806
[Accessed 04 04 2016].

MathWorks, 2014. *Color Thresholder.* [Online]
Available at: http://uk.mathworks.com/help/images/ref/colorthresholder-app.html
[Accessed 12 03 2016].

MathWorks, 2015a. *Image Processing Toolbox.* [Online]
Available at: http://uk.mathworks.com/products/image/
[Accessed 05 10 2015].

MathWorks, 2015b. *Using .NET from MATLAB.* [Online]
Available at: http://uk.mathworks.com/help/matlab/matlab_external/using-net-from-matlab-an-overview.html
[Accessed 23 12 2015].

MathWorks, 2016a. *bagOfFeatures class.* [Online]
Available at: http://uk.mathworks.com/help/vision/ref/bagoffeatures-class.html#buo5kpa
[Accessed 01 10 2015].

MathWorks, 2016b. *MATLAB Compiler.* [Online]
Available at: http://uk.mathworks.com/products/compiler/
[Accessed 24 01 2016].

MathWorks, 2016c. *Representing color with the Lab color space.* [Online]
Available at: http://uk.mathworks.com/discovery/lab-color.html
[Accessed 20 10 2015].

Microsoft Corporation, 2015a. *Asynchronous Programming with Async and Await (C# and Visual Basic).* [Online]
Available at: https://msdn.microsoft.com/en-us/library/hh191443.aspx
[Accessed 3 November 2015].

Microsoft Corporation, 2015b. *Kinect for Windows SDK 2.0.* [Online]
Available at: https://www.microsoft.com/en-us/download/details.aspx?id=44561

Microsoft Corporation, 2015c. *Visual Gesture Builder (VGB).* [Online]
Available at: https://msdn.microsoft.com/en-us/library/dn785304.aspx
[Accessed 25 October 2015].

Microsoft Corporation, 2016a. *BodyIndexFrame Class.* [Online]
Available at: https://msdn.microsoft.com/en-us/library/windowspreview.kinect.bodyindexframe.aspx
[Accessed 02 11 2015].

Microsoft Corporation, 2016b. *Coordinate Mapping.* [Online]
Available at: https://msdn.microsoft.com/en-gb/library/dn785530.aspx
[Accessed 21 12 2015].

Microsoft Corporation, 2016c. *FaceFrame Class.* [Online]
Available at: https://msdn.microsoft.com/en-us/library/microsoft.kinect.face.faceframe.aspx
[Accessed 20 October 2015].

Microsoft Corporation, 2016d. *HighDefinitionFaceFrame Class.* [Online]
Available at: https://msdn.microsoft.com/en-us/library/microsoft.kinect.face.highdefinitionfaceframe.aspx
[Accessed 20 October 2015].

Microsoft Corporation, 2016e. *JointType Enumeration.* [Online]
Available at: https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx
[Accessed 15 01 2016].

Microsoft Corporation, 2016f. *Kinect hardware requirements and sensor setup.* [Online]
Available at: https://developer.microsoft.com/en-us/windows/kinect/hardware-setup
[Accessed 25 04 2016].

Microsoft Corporation, 2016g. *LifeCam Studio.* [Online]
Available at: https://www.microsoft.com/accessories/en-us/products/webcams/lifecam-studio/q2f-00013
[Accessed 12 04 2016].

Microsoft Corporation, 2016h. *Microsoft Azure.* [Online]
Available at: https://azure.microsoft.com/en-gb/?WT.srch=1&WT.mc_ID=SEM_WS1bucqG
[Accessed 01 04 2015].

Movember Foundation, 2016. *Movember Foundation.* [Online]
Available at: https://uk.movember.com/
[Accessed 15 10 2015].

Nathan, A., 2013. *WPF 4.5 Unleashed.* Indiana: SAMS.

Pterneas, V., 2014a. *How to Use Kinect V2 Face Basics.* [Online]
Available at: http://pterneas.com/2014/12/21/kinect-2-face-basics/
[Accessed 01 02 2016].

Pterneas, V., 2014b. *Kinect for Windows Version 2: Body Tracking.* [Online]
Available at: http://pterneas.com/2014/03/13/kinect-for-windows-version-2-body-tracking/
[Accessed 01 11 2015].

Pterneas, V., 2015. *How to Use Kinect HD Face.* [Online]
Available at: http://pterneas.com/2015/06/06/kinect-hd-face/
[Accessed 14 03 2016].

Qian, C. et al., 2014. *Realtime and robust hand tracking from depth..* s.l., Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1106-1113.

RFID Journal, 2016. *How much does an RFID tag cost today?.* [Online]
Available at: https://www.rfidjournal.com/faq/show?85
[Accessed 19 03 2016].

SI 1994 / 2326, 1994. *Personal Protective Equipment (EC Directive) (Amendment) Regulations 1994.* s.l.:s.n.

Souza, C., 2015. *Accord.NET Framework.* [Online]
Available at: http://accord-framework.net/index.html
[Accessed 25 November 2015].

StileEye, 2013. *What is StileEye?.* [Online]
Available at: http://www.stileeye.com/what_is_stileeye.html
[Accessed 12 03 2016].

Yang, M. & Yu, K., 2011. *Real-time clothing recognition in surveillance videos.* s.l., IEEE, pp. 2937-2940.

Appendix A – Kinect Gowning Application Key Functions

The ProcessImage function in Figure 41 contains the .NET algorithm used to extract local colour features from an input bitmap image.

```
public List<ColorFeaturePoint> ProcessImage(Bitmap image)
{
    using (MagickImage magickImage = new MagickImage(image))
    {
        double newWidth = Math.Ceiling(magickImage.Width / 16.0); // This is efficient way to scale to 16x16 cell blocks
        double newHeight = Math.Ceiling(magickImage.Height / 16.0);
        magickImage.Scale((int)newWidth, (int)newHeight); // Scale the input image
        Bitmap scaledBitmap = magickImage.ToBitmap();
        List<ColorFeaturePoint> colorFeaturePoints = new List<ColorFeaturePoint>();
        for (int x = 0; x < scaledBitmap.Width; x++) // Run over every pixel in the new, scaled image
        {
                for (int y = 0; y < scaledBitmap.Height; y++)
                {
                        double normalizedX = -0.5 + (x / (double)scaledBitmap.Width);
                        double normalizedY = -0.5 + (y / (double)scaledBitmap.Height);
                        Color myColor = scaledBitmap.GetPixel(x, y);
                        Rgb myColorRgb = new Rgb() { R = myColor.R, G = myColor.G, B = myColor.B };
                        Lab myColorLab = myColorRgb.To<Lab>();
                        double sumOfSquares = Math.Pow(myColorLab.L, 2) + Math.Pow(myColorLab.A, 2) + Math.Pow(myColorLab.B, 2);
                        double rowNorm = Math.Sqrt(sumOfSquares);

                        myColorLab.L = myColorLab.L / rowNorm;
                        myColorLab.A = myColorLab.A / rowNorm;
                        myColorLab.B = myColorLab.B / rowNorm;

                        colorFeaturePoints.Add(new ColorFeaturePoint(normalizedX, normalizedY, myColorLab));
                }
        }
        return colorFeaturePoints;
    }
}
```

Figure 41: .NET Local Colour Feature Extractor

Making use of the standard Binary Split clustering algorithm, the three bag of words objects are constructed as shown in Figure 42 to Figure 44.

```
public void ConstructSURFBag(bool serialize = true)
{
        _surfBagOfVisualWords = new BagOfVisualWords(_clusteringAlgorithm); // Default is SURF
       _surfBagOfVisualWords.Compute(_imageDataSet.TrainingImages);
       if (serialize)
       {
               _surfBagOfVisualWords.Save(Path.Combine(_basePath, "surfBag.bow"));
        }
}
```

Figure 42: Constructing .NET SURF Bag of Words

```
public void ConstructColorBag(bool serialize = true)
{
        ColorFeatureDetector colorFeatureDetector = new ColorFeatureDetector();
        _colorBagOfVisualWords = new BagOfVisualWords<ColorFeaturePoint, double[]>(colorFeatureDetector, _clusteringAlgorithm);
       _colorBagOfVisualWords.Compute(_imageDataSet.TrainingImages);
       if (serialize)
       {
        _colorBagOfVisualWords.Save(Path.Combine(_basePath, "colorBag.bow"));
        }
}
```

Figure 43: Constructing .NET Colour Bag of Words

```
public void ConstructHOGBag(bool serialize = true)
{
        HOGFeatureDetector hogFeatureDetector = new HOGFeatureDetector();
       _hogBagOfVisualWords = new BagOfVisualWords<HOGFeaturePoint, double[]>(hogFeatureDetector, _clusteringAlgorithm);
       _hogBagOfVisualWords.Compute(_imageDataSet.TrainingImages);
       if (serialize)
       {
        _colorBagOfVisualWords.Save(Path.Combine(_basePath, "hogBag.bow"));
        }
}
```

Figure 44: Constructing .NET HOG Bag of Words

After constructing the necessary training feature vectors, these are combined as per the MATLAB implementation to form a series of normalised, multi-feature vectors as in Figure 45.

```
private double[][] GetMultiFeatureVectors(IEnumerable<TrainingImage> inputSet, int[] expectedLabels)
        {
            double[][] multiFeatureVectors = new double[inputSet.Count()][];

            int trainingImageCount = 0;
            foreach (TrainingImage inputImage in inputSet)
            {
                double[] surfFeatureVector = _surfBagOfVisualWords.GetFeatureVector(inputImage.Image);
                double[] colorFeatureVector = _colorBagOfVisualWords.GetFeatureVector(inputImage.Image);
                double[] textureFeatureVector = _textureBagOfVisualWords.GetFeatureVector(inputImage.Image);

                double[] normalizedSurfFeatureVector = surfFeatureVector.Normalize();
                double[] normalizedColorFeatureVector = colorFeatureVector.Normalize();
                double[] normalizedTextureFeatureVector = textureFeatureVector.Normalize();


                double[] multiFeatureVector = normalizedSurfFeatureVector.Concatenate(normalizedColorFeatureVector)
.Concatenate(normalizedTextureFeatureVector);

                Debug.Print("Multi Feature Vector: {0}", multiFeatureVector.ToString(DefaultArrayFormatProvider.InvariantCulture));

                multiFeatureVectors[trainingImageCount] = multiFeatureVector;
                trainingImageCount++;
            }
            return multiFeatureVectors;
        }
```

Figure 45: Combining Multiple Feature Representations in .NET

The ‘statistics’ module of this library contains a support vector machine implementation that can be trained on the two-dimensional multi-feature vector array. The SVM calculates its classification boundary in a separate step to avoid heavy overhead on the constructor.

```
        public GowningSVMClassifier(IKernel kernel, double[][] trainingFeatureVectors, int[] trainingLabels)
        {
            CustomSVM = new MulticlassSupportVectorMachine(trainingFeatureVectors[0].Length, kernel, 2);
            _customSVMLearner = new MulticlassSupportVectorLearning(CustomSVM, trainingFeatureVectors, trainingLabels);
            _customSVMLearner.Algorithm = (mySvm, classInputs, classOutputs, i, j) =>
                new SequentialMinimalOptimization(mySvm, classInputs, classOutputs.Apply(x => BinariseClasses(x)));
        }

    public double BuildClassifier(string path = null)
        {
            double error = _customSVMLearner.Run();
            if (path != null) //Serialise to disk
            {
                CustomSVM.Save(Path.Combine(path, "gowning.svm"));
            }
            return error;
        }
```

Figure 46: Constructing .NET SVM Classifier

The standard .NET 4.5 framework was used to serialise the necessary class objects to enable GSVM and GPD files to be constructed and exported to disk.

```
        private void btnExportGSVM_Click(object sender, RoutedEventArgs e)
        {
            string currentDateTime = DateTime.Now.ToString("yyyyMMddTHHmmss");
            string formattedDisplayName = ClassifierProperties.DisplayName.Replace(" ", "_").ToLower();
            string fileName = System.IO.Path.ChangeExtension(String.Format("{0}_{1}", currentDateTime, formattedDisplayName), ".gsvm");
            string filePath = System.IO.Path.Combine(EnvironmentVariables.GSVMFilePath, fileName);

            using (FileStream fileStream = new FileStream(filePath, FileMode.CreateNew))
            {
                SerializableGSVMProperties serializableGSVM = new SerializableGSVMProperties(
ClassifierProperties.DisplayName,
ClassifierProperties.ClothingType,
DateTime.Now,
filePath);
                BinaryFormatter bf = new BinaryFormatter();
                bf.Serialize(fileStream, serializableGSVM);
                _gowningSVMClassifier.SerializeSVM(fileStream);
                _bagOfWords.SerializeBags(fileStream);
            }
            CurrentStatus = String.Format("Exported GSVM to: {0}", filePath);
        }
```

Figure 47: Serialising GSVM File in .NET

```
private void btnExportGowningProcedure_Click(object sender, RoutedEventArgs e)
{
    string currentDateTime = DateTime.Now.ToString("yyyyMMddTHHmmss");
    string formattedDisplayName = ProcedureName.Replace(" ", "_").ToLower();
    string fileName = System.IO.Path.ChangeExtension(String.Format("{0}_{1}", currentDateTime, formattedDisplayName), ".gpd");
    string outputPath = System.IO.Path.Combine(EnvironmentVariables.GPDFilePath, fileName);

    BinaryFormatter bf = new BinaryFormatter();

    using (FileStream outputStream = new FileStream(outputPath, FileMode.Create))
    {
        SerializableGPDProperties serializableGPDProperties = new SerializableGPDProperties()
        {
            DisplayName = ProcedureName,
            FilePath = outputPath,
            CreatedTime = DateTime.Now,
            NumberOfItems = ChosenGowningSVMClassifierProperties.Count,
            GSVMFilePaths = ChosenGowningSVMClassifierProperties.Select(x => x.FilePath).ToList(),
        };

        bf.Serialize(outputStream, serializableGPDProperties);
        foreach (SerializableGSVMProperties serializeableGSVMProperties in ChosenGowningSVMClassifierProperties)
        {
            using (FileStream inputStream = new FileStream(serializeableGSVMProperties.FilePath, FileMode.Open))
            {
                SerializableGSVMProperties temp = (SerializableGSVMProperties)bf.Deserialize(inputStream);
                Debug.Print(inputStream.Position.ToString());

                bf.Serialize(outputStream, serializeableGSVMProperties);

                inputStream.CopyTo(outputStream);
            }
            CurrentStatus = String.Format("Written GSVM: {0} to output path: {1}", serializeableGSVMProperties.DisplayName, outputPath);
        }
    }
    CurrentStatus = String.Format("Completed exporting GPD: {0} to output path: {1}", ProcedureName, outputPath);
}
```

Figure 48: Serialising GPD File in .NET

The recognition module performs PPE item classification by running the ‘Compute’ function from the relevant, deserialised GSVM file.

```
public static GSVMProcessingResult RunGowningBowSVMClassifier(BodyComponentProcessingResult bodyComponentProcessingResult,
ClothingItemClassifier clothingItemClassifier)
{
GSVMProcessingResult gsvmProcessingResult = new GSVMProcessingResult();
WriteableBitmap componentWriteableBitmap = bodyComponentProcessingResult.ComponentWriteableBitmap;

UnmanagedImage unmangedImage = new UnmanagedImage(
componentWriteableBitmap.BackBuffer,
              (int)componentWriteableBitmap.Width,
              (int)componentWriteableBitmap.Height,
              componentWriteableBitmap.BackBufferStride,
              System.Drawing.Imaging.PixelFormat.Format32bppArgb);

double[] surfFeatureVector = clothingItemClassifier.SURFBagOfVisualWords.GetFeatureVector(unmangedImage);
       double[] colorFeatureVector = clothingItemClassifier.ColorBagOfVisualWords.GetFeatureVector(unmangedImage);

double[] normalizedSurfFeatureVector = surfFeatureVector.Normalize();
       double[] normalizedColorFeatureVector = colorFeatureVector.Normalize();

double[] multiFeatureVector = normalizedSurfFeatureVector.Concatenate(normalizedColorFeatureVector);
       int classificationResult = clothingItemClassifier.GSVM.Compute(multiFeatureVector);

if (classificationResult == 0)
       {
       gsvmProcessingResult.WearingClothingItem = true;
}
       else
       {
       gsvmProcessingResult.WearingClothingItem = false;
}

            return gsvmProcessingResult;
}
```

Figure 49: Using GSVM File to Run PPE Item Predictions

And similarly for running the beard detector:

```
public static GSVMProcessingResult RunBeardSVMClassifier(BeardComponentProcessingResult beardComponentProcessingResult, BeardItemClassifier
beardItemClassifier)
        {
            GSVMProcessingResult gsvmProcessingResult = new GSVMProcessingResult();
            WriteableBitmap componentWriteableBitmap = beardComponentProcessingResult.ComponentWriteableBitmap;

            UnmanagedImage unmangedImage = new UnmanagedImage(
                componentWriteableBitmap.BackBuffer,
                (int)componentWriteableBitmap.Width,
                (int)componentWriteableBitmap.Height,
                componentWriteableBitmap.BackBufferStride,
                System.Drawing.Imaging.PixelFormat.Format32bppArgb);

            double[] surfFeatureVector = beardItemClassifier.SURFBagOfVisualWords.GetFeatureVector(unmangedImage);
            double[] colorFeatureVector = beardItemClassifier.ColorBagOfVisualWords.GetFeatureVector(unmangedImage);
            double[] hogFeatureVector = beardItemClassifier.HOGBagOfVisualWords.GetFeatureVector(unmangedImage);

            double[] normalizedSurfFeatureVector = surfFeatureVector.Normalize();
            double[] normalizedColorFeatureVector = colorFeatureVector.Normalize();
            double[] normalizedHOGFeatureVector = hogFeatureVector.Normalize();

            double[] multiFeatureVector =
normalizedSurfFeatureVector.Concatenate(normalizedColorFeatureVector).Concatenate(normalizedHOGFeatureVector);
            int classificationResult = beardItemClassifier.GSVM.Compute(multiFeatureVector);

            if (classificationResult == 0)
            {
                gsvmProcessingResult.WearingClothingItem = true;
            }
            else
            {
                gsvmProcessingResult.WearingClothingItem = false;
            }

            return gsvmProcessingResult;
        }
```

Figure 50: Using GSVM File to Run Beard Predictions

Appendix B – Timelines

Initial Timeline

ID	Task Name	Start	Finish	Q4 15			Q1 16			Q2 16
				Oct	Nov	Dec	Jan	Feb	Mar	Apr
1	Buy clothing items, create ethical consent form, identify volunteers and create a process which I will follow to collect data. Perform a trial run on a test subject.	19/10/2015	26/10/2015							
2	Book rooms for recording and organise volunteers. Take recordings of volunteers wearing items.	26/10/2015	30/11/2015							
3	Begin researching classifier algorithms. Begin trying to implement these on the test data and create a few demo applications. Only deal with the head region.	02/11/2015	16/11/2015							
4	Try to generalise the method to create a generic clothes classifier that I could programmatically teach to recognise new clothes.	16/11/2015	30/11/2015							
5	Build an application that can set the parameters of this general classifier when it sees new clothes.	30/11/2015	21/12/2015							
6	Use known algorithms to create a beard detector. Keep improving the basic application – include depth etc?	28/12/2015	25/01/2016							
7	Extend the application to work for the rest of the body. Prove the concept by testing on different kinds of glove.	26/01/2016	12/02/2016							
8	Research WPF in order to create a pretty UI and finish application.	15/02/2016	14/03/2016							
9	GSK to test the completed application.	14/03/2016	21/03/2016							
10	Write report and present draft to supervisor.	14/03/2016	18/04/2016							
11	Finalise and submit report.	18/04/2016	25/04/2016							

Mid-Term Timeline

ID	Task Name	Status	Information	Original Start	Original Finish	Modified Start	Modified Finish	Q4 15			Q1 16			Q2 16		
								Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
1	Buy clothing items, create ethical consent form, identify volunteers and create a process which I will follow to collect data. Perform a trial run on a test subject.	Completed on time.	I have obtained items from GSK and also purchased hats, gloves, glasses, headphones and a ski mask from local stores.	19/10/2015	19/10/2015	19/10/2015	26/10/2015									
2	Book rooms for recording and organise volunteers. Take recordings of volunteers wearing items.	Delayed.	So far, I have only tested the application on myself and a singke colleague who will sign an ethical consent form. This task has been pushed later in the timeline so that I have a strong strategy before requesting time from volunteers. I wish to avoid potentially having to recapture data if the strategy changes.	26/10/2015	30/11/2015	26/01/2016	01/03/2016									
3	Begin researching classifier algorithms. Begin trying to implement these on the test data and create a few demo applications. Only deal with the head region.	Completed on time.	A coherent strategy has been formed and is ready for an implementation attempt.	02/11/2015	16/11/2015	02/11/2015	16/11/2015									
4	Try to generalise the method to create a generic clothes classifier that I could programmatically teach to recognise new clothes.	Complete on time.	The method is already generalised by design – the concept is to detect similarities between gowned and -ungowned images. The actual detected clothing item is irrelevant to the algorithm.	16/11/2015	30/11/2015	16/11/2015	30/11/2015									
5	Build an application that can set the parameters of this general classifier when it sees new clothes.	Incomplete. No delay expected.	This is currently being built.	30/11/2015	21/12/2015	30/11/2015	21/12/2015									
6	Use known algorithms to create a beard detector. Keep improving the basic application – include depth etc?	Incomplete. No delay expected. Initial work has been done.	II have begun investigating this problem and have managed to write an algorithm that can calculate a person's skin tone. Further work is required.	28/12/2015	25/01/2016	28/12/2015	25/01/2016									
7	Extend the application to work for the rest of the body. Prove the concept by testing on different kinds of glove.	Incomplete. No delay expected.	Not started.	26/01/2016	12/02/2016	26/01/2016	12/02/2016									
8	Research WPF in order to create a pretty UI and finish application.	Incomplete. Task brought forward in timeline.	Incomplete.. The paradigm has posed many challenges that I have decided to tackle earlier in the project timeline.	15/02/2016	14/03/2016	26/10/2015	14/03/2016									
9	GSK to test the completed application.	Incomplete. No delay expected.		14/03/2016	21/03/2016	14/03/2016	21/03/2016									
10	Write report and present draft to supervisor.	Incomplete. No delay expected.		14/03/2016	18/04/2016	14/03/2016	18/04/2016									
11	Finalise and submit report.	Incomplete. No delay expected.		18/04/2016	25/04/2016	18/04/2016	25/04/2016									

Final Timeline

ID	Task Name	Status	Information	Original Start	Original Finish	Modified Start	Modified Finish	Q4 15 (Oct, Nov, Dec) / Q1 16 (Jan, Feb, Mar) / Q2 16 (Apr, May, Jun)
1	Buy clothing items, create ethical consent form, identify volunteers and create a process which I will follow to collect data. Perform a trial run on a test subject.	Completed on time.	Items obtained from GSK and additional items purchased from local stores.	19/10/2015	19/10/2015	19/10/2015	26/10/2015	
2	Book rooms for recording and organise volunteers. Take recordings of volunteers wearing items.	Completed with delay.	Data captured from volunteers at GSK.	26/10/2015	30/11/2015	26/01/2016	01/03/2016	
3	Begin researching classifier algorithms. Begin trying to implement these on the test data and create a few demo applications. Only deal with the head region.	Completed on time.	A coherent strategy has been formed and is ready for an implementation attempt.	02/11/2015	16/11/2015	02/11/2015	16/11/2015	
4	Try to generalise the method to create a generic clothes classifier that I could programmatically teach to recognise new clothes.	Complete on time.	The method is already generalised by design – the concept is to detect similarities between gowned and - ungowned images. The actual detected clothing item is irrelevant to the algorithm.	16/11/2015	30/11/2015	16/11/2015	30/11/2015	
5	Build an application that can set the parameters of this general classifier when it sees new clothes.	Completed on tme	MATLAB bag of words trainer completed.	30/11/2015	21/12/2015	30/11/2015	21/12/2015	
6	Use known algorithms to create a beard detector. Keep improving the basic application – include depth etc?	Completed on time.	Initial beard detector written in MATLAB and live demonstration written in C# .NET	28/12/2015	25/01/2016	28/12/2015	25/01/2016	
7	Extend the application to work for the rest of the body. Prove the concept by testing on different kinds of glove.	Competed on time.	Body regions trained as per Preliminary Kinect SDK section.	26/01/2016	12/02/2016	26/01/2016	12/02/2016	
8	Research WPF in order to create a pretty UI and finish application.	Completed on time.	Tutorials for WPF and Kinect SDK completed.	15/02/2016	14/03/2016	26/10/2015	14/03/2016	
9	GSK to test the completed application.	Delayed	Submitted for testing. Not yet returned.	14/03/2016	01/06/2016	14/03/2016	01/06/2016	
10	Write report and present draft to supervisor.	Complete with minor delay.	First draft submtted.	14/03/2016	25/04/2016	14/03/2016	25/04/2016	
11	Finalise and submit report.	Complete with minor delay	Report finalised ad submitted.	18/04/2016	27/04/2016	18/04/2016	27/04/2016	

115

Appendix C – Confidentiality and Intellectual Property Agreement

The following pages contain confidentiality and intellectual property agreements signed by the developer, the project supervisor and the industrial sponsor.

Confidential Information, Intellectual Property Rights and Use of Personal Data

Dear Benjamin Biggs

You will be participating in the GlaxoSmithKline University Dissertation Scheme ("the Scheme") in relation to the dissertation which you are undertaking at The University of Warwick (the "Project").

It is a condition of participating in the Scheme that you sign this Agreement; it sets out the expectations of GlaxoSmithKline Services Unlimited and its affiliated group of companies (collectively "GSK") in relation to any GSK confidential information which you may have access to; the intellectual property rights in relation to the Project; and the use by GSK of your personal data whilst participating in the Scheme.

Confidential information

While you are participating in the Scheme it may be necessary for us to provide you with information of a confidential nature relating to GSK which is not already available to the public (collectively, the 'Confidential Information'). The Information may include, but will not be limited to:
- any technical secrets, confidential research work, technical processes, formulae, inventions, applications for and information about patents;
- transactions, finance or business affairs of GSK or of any customers of GSK; and
- any information, similar to that identified in the three bullet points above relating to any third parties with which GSK has a business relationship

We agree to disclose the Confidential Information to you, and you agree to receive the Confidential Information, on condition that you will:
- use the Confidential Information solely for the purpose of the Project and for no other purpose whatsoever;
- obtain the written agreement of AMT Informatics Manager before disclosing any Confidential Information which is contained in the Project to your university;
- disclose the Confidential Information only to a Abhir Bhalerao who has signed a confidentiality agreement with GlaxoSmithKline Services Unlimited;
- not disclose the Confidential Information in any other manner whatsoever without the prior written consent of AMT Informatics Manager.
- keep the Confidential Information, and all copies of that information, secure, and in such a way as to prevent unauthorised access by a third party.

It is agreed that the limitations on use and disclosure referred to in the preceding paragraph shall not apply to any part of the Confidential Information that:
- you can prove was already known to you prior to the disclosure of the Confidential Information by GSK; or
- is public knowledge or subsequently becomes known to the public other than by a breach of this Agreement; or
- is received by you without restriction from any third party who is entitled to disclose the Confidential Information.

You will deliver to us on request (which we shall be entitled to make at any time), and in any event on completion of the Project, all documents furnished by us or obtained or created by you and comprised in the Confidential Information, together with all copies, abstracts or summaries thereof, whether held on computer or in manual form.

Intellectual property

In consideration of your enrolment in the Scheme, and of any resources provided to you by GSK during your participation, you agree that, on GSK's written request, you will assign your entire right, title and interest (including but not limited to any intellectual property rights) in and to any Project Materials produced in the course of the Project to GSK. For the purposes of this Agreement, "Project Materials" shall mean any idea, invention, technique, modification, process or improvement (whether patentable or not), any industrial design (whether registrable or not), any writings, other works of authorship or derivative works of such writing or works and any other work product created, conceived or developed by you, solely or in conjunction with others. You further agree that you will execute, or procure the execution of, such documents as may be required to file applications and obtain relevant patents in any countries in the name of GSK or its nominees.

Confidentiality Agreement

22/07/15

Abhir Bhalerao
The University of Warwick
Coventry
CV4 7AL

Dear **Abhir Bhalerao**

GlaxoSmithKline Confidential Information

You are acting as **Academic Supervisor** of **Benjamin Biggs**, in relation to the **project** which **he** is completing for the purposes of **his Discrete Mathematics degree.**

During your supervision you may have access to, or have disclosed to you, certain confidential information (the 'Information') relating to the business of the GlaxoSmithKline group of companies, and in particular **Kinect frame data (similar to short videos) captured, depicting members of GSK staff who have agreed (in writing) to take part in this project.**

We ask you to undertake to hold the Information, whether disclosed in writing or verbally, in the strictest confidence and not to disclose the Information or any part of it to any other person or company. We agree that these limitations will not apply to any part of the Information which is wholly in the public domain or is subsequently disclosed to the public other than by breach of this Agreement.

Please sign the enclosed duplicate of this letter to confirm that you accept the above undertaking and return the signed agreement to **Patrick Hyett** at **GlaxoSmithKline, Ware** via email **benjamin.x.biggs@gsk.com** (Cc: WW.UNIVERSITY@gsk.com)

Yours sincerely

Patrick Hyett
For and on behalf of GlaxoSmithKline Services Unlimited

I agree to the terms of the confidentiality agreement set out above

Signed ...... Abhir H. Bhalerao Date ...29/07/2015..................

Abhir Bhalerao

119

22/7/2015

Abhir Bhalerao
University of Warwick
Coventry
CV4 7AL

Dear Abhir Bhalerao

Secrecy Undertaking

You have indicated that you are willing to act as academic supervisor of Benjamin Biggs, an Industrial Placement Student of GlaxoSmithKline, who is doing a project on Kinect Clothes Recognition.

During your supervision/examination you may have access to or have disclosed to you certain confidential information (hereinafter referred to as "the Confidential Information") relating to the above and to our or our Affiliates research and business in general.

We would ask you to undertake:-

(a) to hold the Confidential Information whether disclosed in writing or verbally, in the strictest confidence and not to disclose the Confidential Information or any part thereof to any other person or company; and

(b) to make no commercial use of the Confidential Information without our prior written consent.

It is hereby agreed that the limitations to use and disclosure referred to above shall not apply to any part of the Confidential Information that:

(i) is already known to you prior to its disclosure by us; or

(ii) is in the public domain or is subsequently disclosed to the public other than by breach of this Agreement; or

(iii) is received by you without limitation from a bona fide independent third party; or

(iv) you can demonstrate by written record was developed by yourselves independently of the disclosure of the Confidential Information by us; or

(v) is required to be disclosed by law. For the avoidance of doubt, nothing within this provision is intended to permit disclosures under the Freedom of Information Act 2000, which would, but for this provision, have been covered by an exemption under this Act.

Notwithstanding any provision herein to the contrary, in the event that you hereafter believe that you may be obligated by mandatory applicable law, regulatory rule or judicial or administrative order to disclose the Confidential Information or any portion thereof, you shall immediately notify GSK thereof of each such potential requirement and identify the Confidential Information that may be affected thereby, so that GSK may advise you of whether it agrees that you are so obligated and/or seek an appropriate protective order or other remedy with respect to narrowing the scope of such requirement and/or waive compliance by you with the provisions of this Agreement.

In this Agreement "Affiliate" means any company which controls, is controlled by, or is under common control with, a party. A company shall be regarded as in control of another company for the purposes of this Agreement if it owns or directly or indirectly controls at least fifty percent (50%) of the voting share capital of the other company or, in the absence of the ownership of at least fifty percent (50%) of the voting share capital of the company, if it controls the composition of its Board of Directors.

If the foregoing is acceptable to you, please so indicate by duly signing the enclosed duplicate of this letter and returning it to us.

Yours sincerely
for and on behalf of

GLAXOSMITHKLINE RESEARCH AND DEVELOPMENT LIMITED

Signature: ....

Title: ....HM MANAGER..

Date: ....11 AUG 2015...

Accepted by:- Abhir Bhalerao

Signature .... Abhir H. Bhalerao ....

Date ....29/07/2015...