2D and 3D Face Analysis for Ticketless Rail Travel

Conference Paper · May 2018

3 authors:

Lyndon Neal Smith
University of the West of England, Bristol

Wenhao Zhang
University of the West of England, Bristol

Melvyn L Smith
University of the West of England, Bristol

2D and 3D Face Analysis for Ticketless Rail Travel

L. N. Smith, W. Zhang, and M. L. Smith
Centre for Machine Vision, BRL, Department of Engineering Design and Mathematics
University of the West of England, Bristol BS16 1QY UK

Abstract - *Research is reported into the design, implementation and functionalities of a vision system that employs the human face as a biometric for enabling ticketless rail travel. The system has been developed to optimise performance in the relatively unstructured railway station scenario. In addition to establishing the working vision system, major outputs of the work have included demonstration that 3D face recovery prevents 'spoofing' by use of photographs; and also the finding that incorporation of 3D data into facial modelling has enabled a 6% improvement in face classification. Our conclusion is that 3D data increase face recognition reliability significantly and will be the enabling factor for ensuring revenue protection when employing vision systems for implementation of ticketless rail travel.*

Keywords: 2D 3D face analysis biometric

1 Introduction

Currently many rail transportation companies ensure revenue protection by issuing tickets (or electronic swipe cards) that enable customers to access platform areas by passing through paddle gates. While this approach works relatively well during most times of the day, it creates some quite serious problems at peak periods – in the form of pronounced bottlenecks and associated passenger throughput delays. If the paddle gates could be eliminated, whilst still ensuring revenue protection, the experience of the traveller could be significantly improved, in terms of their journeys being made easier, quicker and more pleasant. Also, the rail operators might benefit from reduced costs arising from not having to operate and maintain large amounts of relatively expensive and complex electro-mechanical paddle gate equipment. The ultimate long-term aim is therefore to replace the currently employed sets of paddle gates with 'gateless gatelines', where there is one large open entrance which passengers could pass through while the system automatically detects that fare has been paid, thereby producing minimal obstruction and inconvenience and ensuring maximum passenger flow through stations. Machine vision provides a convenient and useful means of implementing such an entrance, without the need for passengers to carry active items such as RFID cards or Smart Phones. This can be achieved by using the face as a useful biometric and one that cannot easily be hid from view (this would itself generate suspicion). Also, unlike most biometric-based technologies, face recognition (FR) does not require particular actions by the users, such as having to place hands/fingers on designated areas for palm/fingerprint recognition. However, the relatively unstructured nature of the situation, and the need to use vision to identify a reliable biometric, so that passengers can be reliably identified and repeat fare evaders detected, means that face analysis for ticketless rail travel is a quite challenging vision application. The situation of rail passengers casually passing through a large entrance area in groups is considerably more involved than, for example, face authentication for a person standing in front of a door. Reasons for this include the fact that in the rush hour, passengers are likely to be moving quite quickly and to be close together and the large entrance allows a number of people to pass through simultaneously. This may result in occlusion, the same face appearing more than once, non-uniform illuminations, low resolution (due to the size of the entrance and the consequent relatively wide camera field-of-view), and/or significant variation in poses. Therefore, reliable FR for this application requires relatively advanced machine vision hardware and software implementations for generating the needed rich data sets and in-depth modelling. Implementing this has required research investigation of:

• Major challenges associated with the application context, including incomplete/noisy data, - such as significant changes in illumination and passenger behaviours, facial occlusion, a relatively long imaging distance, variable walking paces of passengers and space limitations in a station environment.

• A preliminary 2D facial recognition system. This includes the study of feature extraction and classification, as well as evaluation of system performance when subjecting it to a realistic real-world environment. While this 2D system has yielded promising results, it confirmed the need to explore 3D features to ensure robust and reliable operation in this kind of relatively unstructured application environment.

• Analysis of the face in 3D. A 3D system, based on recovery of facial surface normals using a new type of two-source photometric stereo, is described. The accuracy of surface gradient recovery is found to be directly linked to facial recognition accuracy.

A significant finding that emerged from the above investigations was that capture and analysis of 3D face data, in addition to 2D data, is required to provide reliability in face detection and recognition in the rail travel situation. The investigations were preceded by a detailed literature review on FR research, the full details of which are beyond the scope of this paper. However, below the main approaches that have been employed are identified, and illustrated by reference to various researchers, and the benefits that can be realised by employing 3D data in the analysis, are explained.

2 Background

2.1 2D face recognition

Here conventional digital images are processed by direct analysis of pixel intensities, where one approach is to employ feature-based methods to process an image to identify, extract, measure, and compare facial features. Relevant features include eyes, nose, mouth, ears, and/or other fiducial marks. The aim is to establish geometric relationships between facial points, by measuring the distance between the eyes, lips, chin, and eye brows, with statistical techniques then being used to recognize faces using relationships between these measurements. An example of a recent feature-based approach is the work by Thomas and Peter [1], where they used Part-based One vs One Features for bird species identification using the Caltech UCSD Birds (CUB) dataset and on Labelled Faces in the Wild (LFW) dataset. They reported an average accuracy of 73.30% on CUB and 93.13% on the LFW datasets. These feature-based approaches have the advantage of high-speed matching due to their compact representation of the face images, as well as invariance to size and illumination. Disadvantages include errors that occur when the face is presented at an angle to the camera (metrics can be applied with the aim of reducing the magnitudes of these errors, but some errors are likely to remain) and the difficulty associated with reliable automatic feature detection.

The other major approach for 2D FR are the global methods. Here analysis is performed on the entire image of the face rather than on local features so that similarities of the whole face are compared, while ignoring individual features like eyes, mouth, nose etc. In global methods, analysis is generally undertaken by application of statistical or artificial intelligence methods. Statistical methods involve a mask to iteratively compare a face image to all other faces in the database. Such approaches have been shown to work efficiently under controlled environments. For example, Yang and Zhou [2] employed a statistical method known as Augmented Lagrangian Methods (ALM) to recognise faces in images corrupted by disguise, illumination, and pose – an accuracy of up to 82% was reported. Artificial Intelligence global approaches utilize tools such as neural networks and machine learning techniques to recognize faces in 2D texture data. Examples of this approach include: Neural Networks (NN), Support Vector Machines (SVM), Hidden Markov Model (HMM) and Local Binary Pattern (LBP). Although such methods can produce better recognition results than feature-based techniques, they are computationally intensive since they process each facial pixel and they do not perform effectively under large variations in pose, scale and illumination. In order to address this, Marisco et al. [3] proposed a scheme that aimed to quantitatively assessed pose and illumination changes. Experiments conducted on LFW, FERET, and SCface databases revealed an accuracy of 61%, 95%, and 89%, respectively. To summarise, 2D global methods can provide good functionality but lack the robustness needed for reliable operation in unstructured situations such as that of the ticketless rail travel application.

2.2 3D face recognition

The authors have investigated 3D face reconstruction methods that utilize 3D features for achieving higher accuracy and robustness in face analysis [4]. 3D facial features reveal facial topology by providing geodesic distances and surface curvatures. They have thus shown promise for bringing higher accuracy to face recognition and improved robustness to practical applications where scenes are complex and dynamic. However, the exploitation of 3D vision is not currently sufficient to enable a wide array of 3D vision based applications. This is mainly due to 3D reconstruction techniques being associated with 3D imaging systems that are commonly perceived as being slow in operation, having limited work envelopes and requiring bulky and/or expensive setups. Consequently, many algorithms struggle to find their way into real-world scenarios. To address this, we introduced a variation on the standard three or four light photometric stereo (PS) method, where two light sources are employed [4]; and, for the first time, we apply it to various types of realistic data.

2.3 Photometric stereo principles

Photometric stereo (PS) enables calculation of surface normals from reflectance maps obtained from images of the same object captured under different illumination directions. As shown in Fig. 1, the object is viewed from a single camera and imaged sequentially by a number of structured lights in known locations. Four lights (and so four views) are shown in Fig. 1, but Woodham [5] showed that three views are sufficient to uniquely determine surface normal as well as albedo at each image point, provided that the directions of incident illumination are not collinear in azimuth. PS techniques are superior in capturing detailed high-frequency 3D textures and are less affected by image noise compared to triangulation-based techniques [6], which recover distance to a surface rather than surface orientations and may require the use of a projected pattern. The latter will often place constraints upon the system’s range and/or work envelope, while the camera interface bandwidth and the resolution of the projected pattern will limit the resolution with which triangulation systems can measure object distance, hence introducing noise and artefacts and limiting their ability to resolve 3D surface textures. In contrast, PS methods can directly recover 3D surface textures with a high resolution that is only limited by the camera sensor and lens, while normally requiring only one camera for image capture - thereby simplifying the calibration process and allowing for high efficiency, a greater depth of field and a large work envelope.

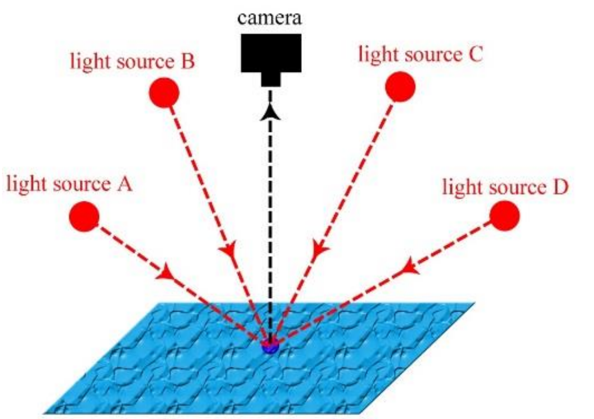


Fig 1. The principle of photometric stereo, which employs a single camera to capture multiple images of a surface illuminated by multiple light sources.

3 Method

3.1 Two-source photometric stereo

Although the standard 4-source PS method is highly accurate and relatively efficient in recovering 3D surface normals, it is not ideal for facilitating real-world applications. Generally, its implementation is prohibited by the need for capturing at least 3 (or more commonly 4 or more) images at high frame rate for every full 3D object reconstruction. Further limitations can arise from the complexity of the data capture system. It is usual for a large set of light sources to be deployed, that result in a set up that is more expensive and requires more space, which can be inconvenient in some cases. We propose to simplify both data capture and hardware design by employing a two-source PS variation where only two light sources are required and therefore only two images need to be captured per reconstruction.

Our aim is to provide a 3D reconstruction algorithm, together with a stereo imaging system, for robust face recognition in unstructured environments. We combine analysis of the 3D data from this system with our 2D global methods, in order to obtain significant increases in the reliability of our face recognition system. To facilitate accurate and robust facial recognition technology for ticketless travel, we developed a two-source Photometric Stereo (PS) system – the first prototype of which is shown in Fig. 2.

Fig. 2. System structure for the 2D+3D face recognition system for ticketless rail travel.

The PS system employs a single camera and two near infra-red illumination sources to capture greyscale images of an object illuminated from different angles. The greyscale images are then used to reconstruct surface normals as a type of 3D feature. The uncertainty associated with the two-source Photometric Stereo problem was compensated for by having two lights symmetrically located on both sides of the camera. This enables us to recover a one-directional surface normal, i.e. horizontal gradients, independent from gradient in the perpendicular direction. Uniform local binary patterns are then employed to further encode surface gradient information (as opposed to conventional intensity analysis). Various classification methods have been implemented and a number were found to achieve desirable performance, including K-nearest neighbour and Support Vector Machine. This is due to the features being more robust than those from intensity images and therefore having higher discriminability. Interestingly, similar results were also achieved when local binary patterns were not employed, but when normalised surface gradient data were used directly for classification.

A representative PS based facial recognition system developed in our CMV laboratory – the Photoface system [7] employed four illumination sources and realised high recognition rates. However, the system shown in Fig 2 only uses two light sources for reconstruction of (horizontal) surface gradients, such that the practicability of the system is sufficient to suit railway station scenarios. Near Infrared (NIR) lights are to be employed to make the system more covert and less intrusive. The surface gradient (3D) features have the advantages that they are independent of illumination and are robust to head pose changes to a large extent, meaning that they are most appropriate for accommodating dynamic passenger behaviours. Furthermore, in being able to capture

PS images and to reconstruct potentially high-resolution 3D features at a high frame rate, this system is superior to stereo vision systems in fulfilling the task of increasing the throughput in station gate scenarios. A 3D facial recognition system also has the advantage that it is not prone to 'spoofing' by a 2D photograph that cannot be differentiated from a real face by conventional 2D feature based systems.

3.2 Construction of face databases

To facilitate ticketless rail travel, two types of face databases need to be constructed, in order that enrolled or registered users/passengers can be differentiated from those who are unregistered. A global face database is constructed by passengers going to any station where an enrolment booth is available. Each passenger needs to be enrolled once only in order to use the system. The hardware structure of the enrolment system is similar to that of the facial recognition system, subject to minor adjustments (e.g. smaller in scale). This database is updated when a passenger gets registered or deregistered. Each local (per-station) face database is also constructed as a subset of the global face database. A local face database is updated when a passenger arrives at a station and 'taps in' (i.e. is added to the database) or a passenger is successfully recognised by the facial recognition system while departing from a station (i.e. removed from the database).

3.3 Conditions for a gateless gateline scenario

The conditions listed below allow design of a facial recognition system that can be more specific to the gateless gateline scenario, thereby allowing maximised performance for this application:

1) No direct sunlight (ambient sunlight is allowed).
2) Co-operative passengers, i.e. the passenger not intentionally avoiding the camera (i.e. no extreme head pose). Note that a passenger will not be required to stop and look up to the camera.
3) No severe occlusion of the face to be recognised.
4) In the initial system, there will be a single-file channel for each gate, i.e. no more than one passenger walking through the gate at a given time.
5) Persons will walk steadily through the gate (e.g. at a fast walking pace), rather than running/moving very quickly.

3.4 Facial Recognition Algorithm Chain

This facial recognition system realises its functionalities by following five main stages:
1) Data capture: the system achieves sequential high-speed synchronisation of a camera (referred to as camera A in the rest of the report) with two NIR light sources. Pairs of greyscale images of a gate corridor scene are captured - with one image illuminated by one of the NIR lights and the other image by the other light.
2) Face detection: the system then searches to detect face regions in the greyscale images.
3) Surface gradient recovery: for every face image present in a pair of PS images, the system performs a two-source PS algorithm to recover (horizontal) surface gradient features.
4) Feature processing: gradient features are then processed in order to provide enhanced discriminative power.
5) Facial recognition: discriminative features are compared to those in the local face database, and a face is classified as belonging to a particular passenger (i.e. a ticket holder) or otherwise an unenrolled/unregistered passenger (i.e. a potential fare evader).

3.5 Facial recognition hardware configuration

Given that the aforementioned two-source PS system will fulfil the role of a facial recognition system, it is worth exploring the influence of the system parameters (e.g. camera height, camera angle, etc.) at this stage, such that the implementation of the system can be tailored for use in a railway station. We therefore sought to obtain the optimal system parameters to shed light upon our initial design choice. This was achieved by a series of simulations, trigonometry calculations and trial experiments. A camera (designated as 'Camera A'), was employed to record PS images for 3D reconstruction and facial recognition, while the utilisation of a second vertically mounted camera was intended for passenger tracking from an overhead viewpoint. Although the investigation of tracking methods was out of the scope of the authors' work on this project, integration of video tracking with the face recognition system offers potential to strengthen revenue protection by helping to avoid face occlusion/overlapping and reducing imaging variations caused by extreme dynamic passenger behaviours.

By incorporating the popular Viola-Jones face detector [8] for face detection and the uniform Local Binary Pattern (LBP) histogram features for facial recognition [9], this system can image and detect the face (without severe head poses present) of a passenger of average height (e.g. 1.75m in the experiments). In the experiments the camera was positioned at a height of 2.35m, which is the minimum clearance according to the UK National Rail standard. It was observed that, at an imaging distance between 1.2m and 2.8m, this provided good-quality facial data – the faces were not severely occluded, they were mostly included in the camera's field of view, and they were of sufficient image sizes (over 200×200 pixels). Although results indicated that this can provide face recognition capability in many gateless gateline scenarios, it was observed that when lighting conditions changed, the recognition accuracy plunged drastically. This again proved that the incorporation of 3D features (i.e. surface gradients from the PS system) would be very significant for ensuring robust facial recognition. For 3D imaging, the distance between the two NIR lights also needs to be determined in order to give accurate surface gradient reconstructions. Increasing the between-light distance (L) has a positive effect on the reconstruction accuracy, according to our physical experiments with an L of 0.6m, 0.8m, 1.0m and 1.2m, respectively. L was eventually set to 0.775m – the same as the width of a UK E2 gate.

4 Results

4.1 2D Feature Based Facial Recognition

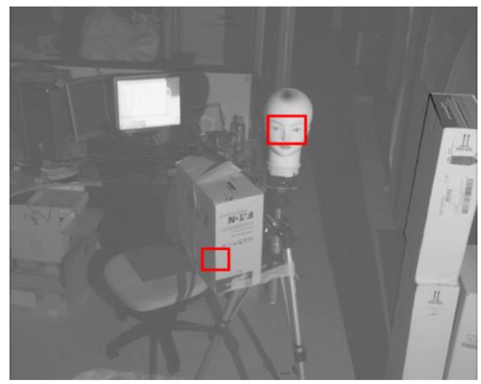
The 2D facial recognition system was developed on the Windows 7 64-bit operating system (OS). Software tools and libraries used include C++ Visual Studio 2013, OpenCV 3.1 and the Pointgrey libraries. After development was completed, the software package was deployed on the Windows 7 64-bit OS. The system exhibited a high performance with regard to accuracy, robustness and efficiency in physical experiments. Fig. 3 shows that three 'passengers' walked past the gate in a natural manner in less than two seconds and were all successfully recognised.

Fig. 3. A demonstration where three passengers were all recognised by the 2D facial recognition system. Upon successful recognition, the name for each passenger was displayed in green on top of an automatically fitted face bounding box.

4.2 Surface Gradient Reconstruction

The distance between the two NIR lights (i.e. L) can be varied in a PS based facial recognition system. Experiments were conducted to investigate the effects of changing this distance, where different values of L were tested. The system setting and one of the recorded images can be seen in Fig. 4.

(a)

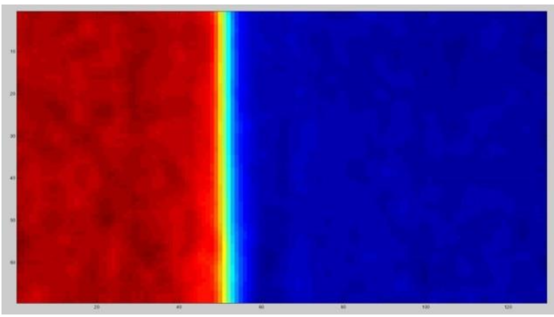


(b)

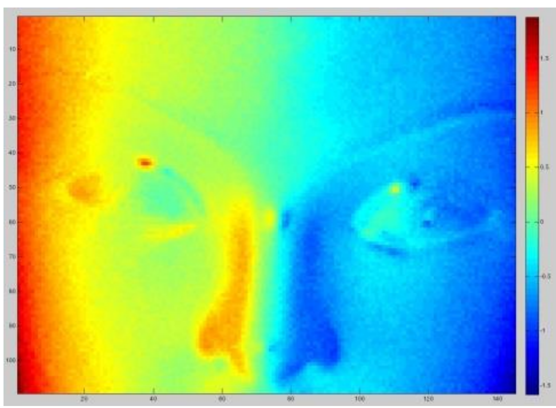
Fig. 4. A demonstration of the PS experiment exploring the impact of different values of L. (a) An illustration of a two-source PS system structure. (b) One of the recorded raw images where the red boxes represent the regions of interest.

Reconstruction results of surface gradients (the horizontal component) for the regions of interest are shown in Fig. 5, where red areas indicate the pixels are oriented to the left, and blue areas indicate the pixels are oriented to the right.

It was found experimentally that the reconstruction errors decrease when L was increased from 0.6m to 1.2m. Therefore, the largest possible value for L of 1.2m is to be employed for design of this PS based facial recognition system. The 3D face recognition system was tested in the authors' machine vision laboratory in Bristol with a large number of persons, in unstructured situations that were intended to simulate a gateless gateline, and average recognition accuracies of up to 95% were obtained. Results also confirmed that recovery and analysis of the 3D data prevents 'spoofing' of the system by holding a photograph of someone else in front of the face – this is described in more detail below. An improvement of up to 6% over use of only 2D data was observed when 3D and 2D data were combined for facial analysis. The potential importance of this improvement was illustrated in February 2018, when the system appeared on UK national prime time television, in a video sequence for the BBC programme 'The One Show' [10]. Here the task was to automatically recognise and so distinguish between two very similar looking people. These two, referred to as "doppelgangers" in the show, looked very similar to the naked eye and reported that they were often mistaken for each other in Bristol, even though they were not related. Interestingly, the system could only differentiate between them when employing 3D data.



(a)



(b)

Fig. 5. Experimental results for surface gradient reconstruction where red areas indicate the pixels are oriented to the left, and blue areas indicate the pixels are oriented to the right. (a) Surface gradients for a region on the box. (b) Surface gradients for a region on the face.

As mentioned above, spoofing of conventional 2D face recognition systems can easily be achieved by holding an image of another person in front of the face. The authors have been asked whether there is any possibility of spoofing the 3D system; and the answer is that this would not be at all easy to do. The system is resilient to simple attempts to spoof it by showing it a photograph that has been curved into a convex shape. The reason for this is that the system is set to detect the shape of the face in 3D; and this is relatively complex compared to the paper, which curves in only direction at a time. One can bend paper into a cylinder, or into a cone, but an attempt to use it to closely cover the morphology of the face would result crumpling and/or distortion. Another approach to spoofing could be to employ a commercial scanner to scan a face, and then to reconstruct it using rapid prototyping technology. In this way a relatively accurate 3D model of a face could be made, but while this would satisfy the 3D analysis, it would not incorporate the 2D facial data and there is no easy or convenient way to incorporate such data. Therefore, it is the combination of 3D with 2D data that provides the high degree of security that makes such a system very difficult to spoof. If however such spoofing were in some way possible, the amount of work and expense involved would be very likely to place it beyond the capabilities and/or realistic expenditure of a potential rail fare evader. It is also likely that the spoofing operation itself would constitute suspicious-looking behaviour that a typical would-be fare evader would not wish to engage in. These considerations have led us to the conclusion that that the described approach is an effective means of preventing spoofing and that it is well suited to enabling ticketless rail travel.

5 Conclusion

A prototype facial recognition system for introducing a gateless gateline in a railway station has been implemented, which features the following:

• A PS based approach that can recover surface gradients for accurate and robust facial recognition. Only two light sources are used for PS reconstructions in order to simplify system structure and to enhance real-time imaging.

• Demonstration hardware system, where a number of parameters can be adjusted within possible ranges, in order to optimise the system configuration.

• A software framework similar to that of the preliminary 2D facial recognition system, with the addition of richer surface gradient features (3D).

A combination of theoretical studies, physical experiments and performance evaluations have ensured that the proposed hardware design is highly compatible with the software system. This has resulted in a prototype system that demonstrates the potential that 2D, and particularly 3D, machine vision recognition systems offer for employing the face as a reliable biometric for ticketless rail travel. The benefits of introducing vision based gateless gatelines at railway stations are, therefore, clear: this technology can reduce crowding problems at stations, particularly at peak times. This will reduce delays experienced by passengers and thus enhance their travel experience, while simultaneously helping operators to increase passenger throughput, thereby increasing revenues. The next stage in achieving these goals will be to test the technology in the real and dynamic environment of a busy railway station.

6 References

[1] B. Thomas and B. Peter, "POOF: Part-based one vs. one features for fine-grained categorization, face verification, and attribute estimation", Computer Vision and Pattern Recognition, (955–962), (2013).

[2] A. Y. Yang and Z. Zhou, "Fast l-1-minimization algorithms for robust face recognition", IEEE Trans. Image Process. Vol. 22 No.8, (1598–1603), (2013).

[3] M. Marsico, M. Nappi, D. Riccio and H. Wechsler, "Robust face recognition for uncontrolled poseand illumination changes", (IEEE Trans. Syst. Man. Cybern.), Vol. 43, No. 1, (149–163), (2013).

[4] W. Zhang, M. Smith, L. Smith and A. Farooq, "Gender recognition from facial images: 2D or 3D?", Journal of the Optical Society of America A, Vol. 33, No. 3, (333-344). ISSN 1084-7529, (2016).

[5] R. J. Woodham, "Photometric method for determining surface orientation from multiple images", Optical engineering, Vol. 19, (139-144), (1980).

[6] S. Herbort, and C. Wöhler, "An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods," 3D Research, Vol. 2, No. 3, (1-17), (2011).

[7] M. F. Hansen, G. A. Atkinson, L. N. Smith and M. L. Smith, "3D face reconstructions from photometric stereo using near infrared and visible light", Computer Vision and Image Understanding, Vol. 114, No. 8, (942-951), (2010).

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", Computer Vision and Pattern Recognition, 2001. CVPR 2001, (proceedings of the 2001 IEEE Computer Society Conference), Vol. 1, (I-511), (2001).

[9] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition", IEEE transactions on pattern analysis and machine intelligence, Vo. 28, No. 12, (2037-2041), (2006).

[10] The One Show, BBC One, British Broadcasting Corporation, 19:00 GMT, 14 February (2018).