3D Face Recognition using Photometric Stereo and Deep Learning

Bryan Kneis [†]
TBSK Consultancy
Bristol, UK
bryan@tbskconsultancy.com

Wenhao Zhang
Bristol Robotics Laboratory
Bristol, UK
wenhao.zhang@brl.ac.uk

ABSTRACT

Illumination variance is one of the largest real-world problems when deploying face recognition systems. Over the last few years much work has gone into the development of novel 3D face recognition methods to overcome this issue. Photometric stereo is a well-established 3D reconstruction technique capable of recovering the normals and albedo of a surface. Although it provides a way to obtain 3D data, the amount of training data available captured using photometric stereo often does not provide sufficient modelling capacity for training state-of-the-art feature extractors, such as deep convolutional neural networks, from scratch.

In this work we present a novel approach to utilising the lighting apparatus commonly used for photometric stereo to synthesise data that can act as a biometric. Combining this with deep learning techniques not only did we achieve near state-of-the-art results, but it gave insight into the possibility of using photometric stereo without the need of reconstruction. This could not only simplify the face recognition process but avoid unnecessary error that may arise from reconstruction.

Additionally, we utilise the active lighting from photometric stereo to evaluate the effect of illumination on face recognition. We compare our method to the state-of-the-art 3D methods and discuss potential use cases for our system.

CCS CONCEPTS

• Computing methodologies ~ Machine learning ~ Machine learning approaches ~ Neural networks • Computer systems organization ~ Real-time systems ~ Real-time system specification

KEYWORDS

Deep Learning, Photometric Stereo, Face Recognition

1 Introduction

Photometric stereo is a technique capable of capturing 3D information using multiple light sources. Using at least three independent light sources, photometric stereo can recover the surface normal of a point [1]. The method is capable of being used for many different contexts including face recognition [4, 6, 7]. Although much of the research states its applicability to face recognition, little offers insight into the performance of the surface recovery in relation to recognition accuracy. Unlike traditional stereo methods, photometric stereo recovers the surface normal and albedo of a point instead of the depth. Recovering the depth has been widely used for face recognition and performance improvements can be seen over 2D information [2, 8]. The surface normal recovered using photometric stereo is represented as a vector signifying a perpendicular line to the point, while the albedo represents the proportion of incident light being reflected by the surface. Shown in Figure 1 are examples of subject images and their corresponding surface reconstructions.

Photometric stereo was first developed to overcome the problems of traditional stereo techniques, namely the correspondence problem [3]. Traditional stereo techniques, such as binocular stereo vision, need to correspond several points in a pair of stereo images to recover the depth. This introduces a major concern in the context of face recognition, since the amount of time to establish the correspondence can be much larger than the recognition itself, making it a challenge to implement a real-time system. Additionally, a common technique to overcome the correspondence problem is to employ a block matching algorithm [5]. Unfortunately, the algorithm often incurs pixel loss when generating the disparity between the left and right image [9]. The main concern using lower resolution images for face recognition is the loss of texture information. Many feature extractors, such as local binary patterns (LBP), rely on surface texture analysis. Compared to employing low resolution data, providing better resolution images can often increase accuracy.

Face recognition algorithms commonly contain three primary steps: data acquisition, feature extraction and classification. In recent years, much attention has gone into feature extraction, with

multiple variants of traditional methods being developed [10, 11, 12, 13]. It is not until recently that extracting 3D data has been investigated to overcome the limitations of current systems, such as illumination and face expression variations. 2D data acquisitioning methods are direct measures of light irradiance, and therefore the data are influenced heavily by the light conditions of the environment. 3D methods attempt to overcome this by transforming the 2D information in a way such that it is descriptive but contains less dependency on the lighting conditions. In addition to surface gradients and albedo, other values such as BRDF and depth can be exploited [19].

The objective to this research is to determine the effectiveness of transfer learning from data captured using photometric stereo. Additionally, we present a novel approach to synthesizing data using the lighting apparatus to act as a biometric, replacing traditional 3D data found from the surface reconstruction process. Finally, we utilise the unique active lighting apparatus from photometric stereo to perform experiments on the effect of illumination variance on face recognition performance.

Figure 1: Example of Photoface subject images and their corresponding surface reconstructions

2 Related Work

2.1 Photometric Stereo for 3D Face Reconstruction

Ridgefaces introduced in [14] is a variant of the eigenfaces approach [38] in which ridglet transforms are applied to the 3D data recovered using Photometric Stereo. Results were obtained using the Photoface database and showed the method outperforming PCA and fisherfaces by 12%. The paper performed many experiments on the different 3D data and showed the surface gradients along the X and Y axis outperforming the albedo. The best recognition rate was achieved by combining the surface normals with their l1 and l2 norms. Since the ridgefaces paper, notable progress has been made around the capability of automatically identifying features, such as convolutional neural networks. Our work attempts to combine the use of photometric stereo for 3D reconstruction with the performance of state-of-the-art feature learning methods, as opposed to conventional techniques using hand crafted features such as eigenfaces.

Photometric stereo is a technique capable of recovering 3D information from a Lambertian surface. It has been widely adopted as a 3D reconstruction technique since the 1980's [1]. Although much work has been done in its ability to reconstruct a face [6, 7, 15, 16], little research has shown the influence of recognition accuracy. Given that photometric stereo generates several different types of features, more work investigating the features recovered in the context of face recognition could provide better results.

Cheol et al. experimented with face recognition using photometric stereo [20]. PCA, GaborJets and a proposed method was evaluated using the surface normals and albedo. Although this paper gives insight into the use of photometric stereo for face recognition, the paper focused its research on the reconstruction of the face, where the angle of the lights was the focus of examination. Additionally, no experiments were conducted to evaluate which values of the surface normal or albedo performed best. Furthermore, the validation set used was relatively small with only 125 images.

2.2 3D Face Recognition

Using known physical properties, 3D face recognition addresses the problems of 2D face recognition by inferring more descriptive information from an image or images. Using two cameras at a known distance apart, binocular stereo vision captures a stereo pair of images to infer depth. Lao et al. proposed a template matching framework [17] using the depth maps generated from stereo vision to create a pose invariant recognition algorithm. Using the 3D feature vector, the algorithm accurately determines the pose of a face. It does this by searching the face for arcs with radiuses within a certain range, indicating the location of the irises. Using these two iris center coordinates, the model is distorted to transform the face into a canonical pose. Once transformed, recognition is performed to extract the features at the same pose as the other known subjects. In their experiments, varying face poses show only minor differences in recognition accuracy.

Although a lot of literature is available on comparison of 2D to 3D face recognition, little work has been presented in which 2D and 3D face recognition are evaluated using the same database. Comparisons often use qualitative analysis as a means of evaluation [21, 22]. This is due to the lack of data captured containing both 2D and 3D features of the same subject during the same session.

Soltana et al. [23] was one of the first to evaluate the recognition performance of both 2D and 3D features using a database containing both. The 3D images were captured using an infrared camera to produce range images. The paper shows the 3D features outperforming 2D by nearly 10%, especially when there is a pose variation in the subject. To measure the effect of pose and illumination variations on recognition accuracy, the two types of variations need to be isolated from the validation set. Georghiades

et al. [18] performed experiments using five different face recognition techniques where the level of pose and illumination variations were known. Their results showed a correlation between the classification error rate and the angle of the illumination. Natural features of a face, such as a nose, can cause shadows to appear given undesirable illumination and hence an increase in incorrect classifications can occur

3   System Overview

As mentioned, one of the biggest drawbacks to using 3D data is the amount of training data available [21]. Using the Photoface database it would be particularly challenging to train the network from scratch. In order to utilise the vast amount of 2D training data currently available, the network was trained using transfer learning. First the network was initialised using the weights from the popular vggface [30] network and then fine-tuned using the Photoface database. The vggface model is based on the VGG-Very-Deep-16 CNN [31] network architecture and was trained using a combination of the LFW dataset [32] and YouTube face dataset [33]. The network comprises of 11 blocks, each containing a linear operator followed by ReLU. The first 8 blocks are said to be convolutional while the last 3 are fully connected, where the size of the filter matches the size of the input. The resulting feature vector is then passed to a softmax layer to compute the class posterior probabilities.

We have used a much smaller learning rate of 0.00025 for the fine tuning, we found a learning rate higher than this, such as 0.001 was too high and caused the network to overfit to the new data. Lastly, we allowed the network to fine tune all its layers, instead of freezing the convolutional backbone. Although this increased training time, it proved to show even better results than retraining only the classification head and last few convolutional layers. This indicates the appropriateness of the Photoface database for our experiment, since training the network end to end would only be effective when provided enough data.

In order to determine which type of data is more descriptive for our system, we compare the performance of the system using the synthesized data against the albedo and surface normals along the X and Y axis. For information on how the albedo and surface normals are calculated we refer to the method described in the original paper [34].

Figure 2 illustrates the process of using transfer learning to train our network using an existing network trained with similar data. Note that we did not actually freeze the convolutional backbone, instead we opted for a low learning rate to effectively only train the last layers, although the convolutional layers can still then be optimised.
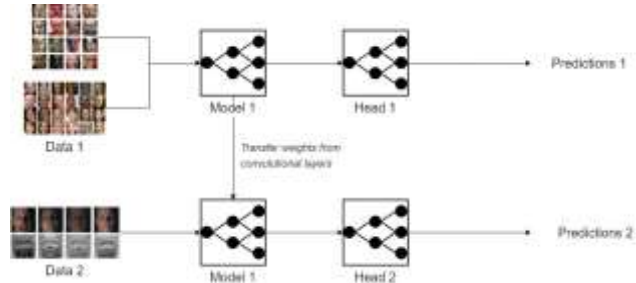

Figure 2: Processing stages of training the network through transfer learning

Shown in Figure 3 is the sequence of our system and its recognition process. There are 5 main entities of the system, the Application, which is responsible for initializing the other 4 entities and ensuring any system resources required are attained. The Camera, which is the software interface to the hardware trigger capturing the actual data. Seen in the diagram is the source images captured by the 4 different cameras. The Image Synthesiser, which combines the source images by averaging each pixel's light vector, to synthesize a new well-lit image. Next is the Face Detector, this is a class implementing the Viola-Jones face detection algorithm and returns a cropped image of only the face region. Lastly we have the Neural Network, this takes the face image and provides a class id and probability of the highest matching classification. Since our network contains fully connected layers after the convolutional layers, we return a prediction directly from the network instead of a feature vector that is then passed to a classifier.
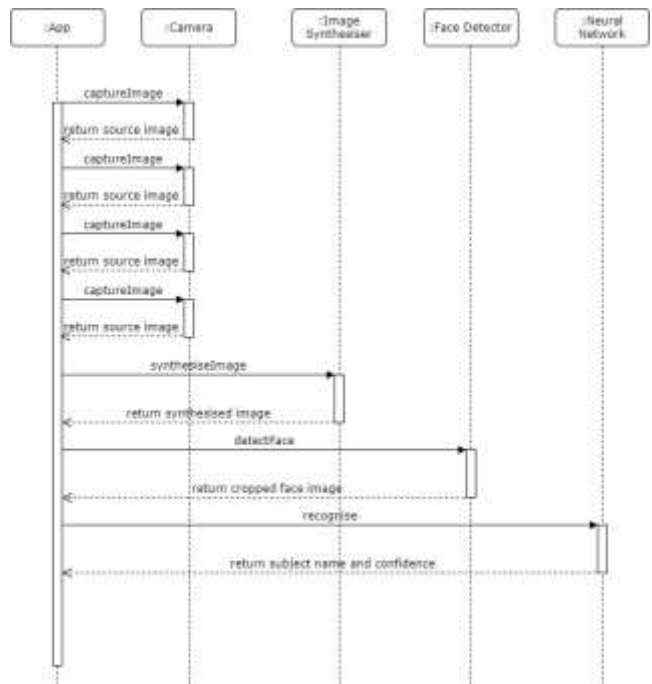

Figure 3: Sequence diagram of the recognition process

4 Method and Experiments

The Photoface database [29] is a dataset containing over 1000 sessions of 260 subjects. The data was captured using a four-source photometric stereo set up, where a camera was set to a fixed position with four independent light sources. The light sources were positioned at the top left, top right, bottom left and bottom right of the camera. In addition to providing a sufficient dataset to validate the photometric stereo techniques, it provides the possibility to use the active lighting to synthesize illumination variation. As the position of the four light sources are fixed and known, we can combine the light intensities of the four images to produce new 2D data. More specifically, by averaging the pixel intensities of all four source images we can create an evenly well-lit image of the subject.

The Photoface database was captured using a Basler 504kc with Camera Link interface operating at 200fps, 1ms exposure time and a 55mm lens. Placed approximately 2 meters from the head of the subject. The light sources were provided by Jessops M100 flashguns spaced 75cm apart [29].

4.1 Synthesizing Data Using Photometric Stereo Lighting

Originally the Photoface database contains four images per subject with varying illumination directions. Traditionally systems using photometric stereo use these four images to reconstruct the surface of the face. Instead our system calculates the mean greyscale values of corresponding pixels across the four images, as shown in Figure 3. This provides the deep neural network optimal lighting conditions of the subject. Furthermore, this allows us to train the neural network on the vast amounts of 2D face data available and then fine tune the network using the synthetic 2D data.
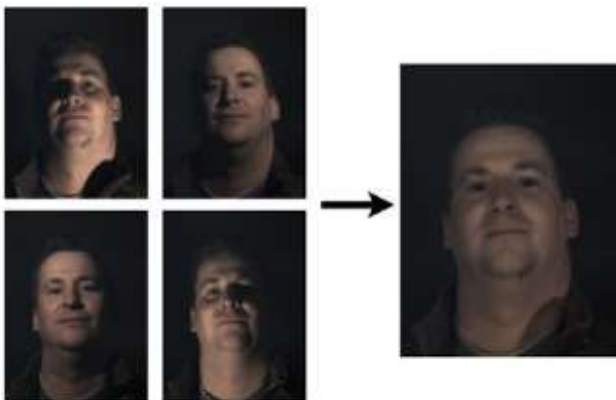

Figure 3: Photometric stereo source images (left) and generated synthetic image (right)

4.2 Effects of Illumination Variation on Face Recognition

Our last experiment measures the effect of illumination variation on face recognition. We compare the recognition accuracy of the

synthetic image used in the first experiment to two additional synthetic images representing illumination distortion. The first image represents a subject containing a directional light source. This was created by averaging the two source images in which the light position was at the top and bottom right to the subject. An example of this can be seen in Figure 4, where clearly the right side of the face contains more illumination.

Figure 4: Photometric stereo source images using top / bottom right light sources (left) and their generated synthetic image

The last image is a randomly selected source image. This represents the worst-case scenario in which there is a single light source illuminating only a part of the face, as shown in Figure 5.
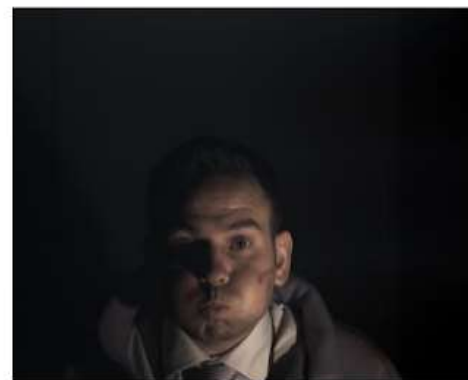

Figure 5: Single photometric stereo source image

5 Results

In this section, we present our experiment's validation results using the Photoface database. The database was pre-processed by removing any subjects with less than three images. Additionally, each face was cropped using the OpenCV implementation of Viola Jones [28] to ensure each method analysed the same face region.

5.1 Performance of our System Against State-of-the-art 3D Methods

Table 1 shows the results of our system using the 3 different types of data, synthesized 2D data, the albedo and surface normals. The percentage of recognition (PR) was used as our primary evaluation metric as it correlates to its effectiveness as a face recognition system [35], moreover the PR was used in the original paper allowing us to create a fair comparison to the state-of-the-art method. Additionally, we show the results of the state-of-the-art method detailed in the original paper for reference.

Method	Recognition Performance
Our system using synthesized 2D data	95%
Our system using albedo	84%
Our system using surface normal (X axis)	77.57%
Our system using surface normal (Y axis)	78.77%
State-of-the-art 3D method	96%

Table 1: Recognition performance of our system and state-of-the-art methods

5.2 Face Recognition Performance Under Varying Illumination

One of the largest problems with 2D face recognition is illumination variation [10, 24, 25]. When illuminating a face, the light can cast shadows that negatively influence the feature vectors extracted and therefore cause classification results to degrade [36]. Using the feature vectors generated previously we can evaluate the effect of illumination variation on recognition accuracy.

Table 2 shows the recognition accuracy for each of three 2D feature vectors we generated. Naturally the synthetic 2D image performs the best, with an improvement of 23.25% to the directional illumination. The synthetic 2D feature vector acts as a baseline for the other two to evaluate against, as it represents the optimal lighting conditions. The worse performing feature vector was the original source image, where an intense illumination variation was present (as can be seen in Figure 5). This feature vector performed the worse with a performance drop of nearly 30%. Based on these results we suggest that the lighting conditions have a strong impact of recognition performance, as discussed in Section 6.

Image	Recognition Performance
Synthetic	95%
Consistent but directional light source	71.75%
Intense lighting variation	54.28%

Table 2: Recognition performance of our system using different training images

6 Discussion

Over the last 10 years much attention has gone into obtaining robust features (hand crafted or automatically identified) including the use of deep neural networks. If we could combine the advantages of using deep neural networks and the recovery of 3D data, even better improvements could be made. The main limitation to this is the acquisition of training data. Typically, deep neural networks use multiple distributions to obtain extremely large data sets [26, 27]. When reconstructing 3D data, the data acquisitioning setup needs to be consistent, such as the four-source photometric stereo rig used for Photoface. This makes gathering the necessary training data for a deep neural network challenging to accomplish.

In our experiment we evaluated the use of the synthetic 2D data, the albedo and the surface normals with the neural network. From our results we clearly see the synthetic data achieving near state-of-the-art performance and outperforms the traditional 3D data such as surface normals, for our system. We believe the neural network can use transfer learning more effectively with the synthetic 2D data since the properties of light between the synthetic and 2D data it was originally trained on are more alike than the 3D data, such as the surface normals.

If the neural network learned robust features when originally trained, that could generalise to new data, then the similarity between the new and original data might have a correlation to the effectiveness of the transfer learning. Since the lower level features are already identified, a smaller number of features would need to be learned, most notably at the higher levels of the neural network. This theory also fits the observation that the albedo outperformed the surface normals, where the albedo shares similar properties to the 2D data, such as representing the ratio of light being reflected from the surface [34].

Although our system did not outperform the state-of-the-art method relative to recognition performance, our system has several benefits that a traditional 3D system might not. Firstly, as we are using the active lighting setup to generate synthetic data to overcome illumination variance, the lighting apparatus is not necessarily required for enrolling subjects into the model. While 3D systems require data acquisitioning to be consistent from enrolment to inference [37], our system would allow a subject to be added using only 2D images, if they do not impose major illumination variance. This could be useful for applications where

enrolment of subjects does not occur where the face recognition system is deployed. Moreover, as mentioned using the Lambertian reflectance model photometric stereo requires at least three lights to reconstruct a surface [34]. Since our system does not perform reconstruction, but instead synthesises the 2D data by averaging the corresponding light intensities, our system could be deployed with two lights. This might especially be useful for mobile applications or areas with limited space, such as small corridors and doorways.

Lastly our experiment has shown that when data captured using photometric stereo has 3D information embedded in them, the reconstruction process might not be a compulsory step for face recognition. This would not only simplify the face recognition process but could avoid unnecessary error that might arise from the surface reconstruction. Furthermore, we believe by using the synthetic data instead of the 3D data, more effective transfer learning can occur from 2D data.

7 Conclusion

In this paper we evaluated our system against the state-of-the-art 3D methods described in the original Photoface paper. Additionally, we measured the effect of illumination variation on recognition accuracy for 2D data by synthesizing illumination using active lighting from the photometric stereo setup. Our results show that performance of a state-of-the-art CNN using the synthesized 2D data achieves near state-of-the-art. Furthermore, at present neural networks for 2D face recognition are receiving a lot more attention from both the research community and industry than 3D [21]. Since our system uses transfer learning based on the weights found from 2D face recognition, our system can continuously appreciate the advances made from other works.

The inspiration for this work was to identify if we could combine the advantages from both 2D and 3D face recognition in a novel way to overcome the shortcomings of existing methods. Another real-world problem with 2D face recognition is spoofing attacks, whereby subjects can use photos or videos of an imposter to trick the system. In future work we intend to extend our system by employing anti-spoofing techniques using the data collected via photometric stereo in order to identify surfaces that do not conform to the usual shape of a face.

Face recognition, like many computer vision tasks is a multi-objective problem. Often works performed on face recognition focus on a single criterion, such as recognition performance. Typically, when optimizing for a specific performance metric, trade-offs needs to be made that can affect other criteria. By combining the benefits found in 2D face recognition, such as its maturity and state-of-the-art feature extractors with the robustness to illumination variance and spoofing attacks from 3D face recognition, a hybrid system could be developed to overcome the challenges found in 2D and 3D face recognition alone.

REFERENCES

[1] Woodham, R. J., Y. Iwahori, and Rob A. Barman.: Photometric stereo: Lambertian reflec-tance and light sources with unknown direction and strength. University of British Colum-bia. Department of Computer Science (1991)

[2] Chang, Kyong I., Kevin W. Bowyer, and Patrick J. Flynn.: An evaluation of multimodal 2D+ 3D face biometrics. In: IEEE transactions on pattern analysis and machine intelligence 27, vol. 4: pp. 619-624 (2005)

[3] Roy, Sébastien, and Ingemar J. Cox.: A maximum-flow formulation of the n-camera stereo correspondence problem. In: IEEE Sixth International Conference Computer Vision, pp. 492-499 (1998).

[4] Roth, Joseph, Yiying Tong, and Xiaoming Liu.: Adaptive 3D face reconstruction from un-constrained photo collections. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4197-4206 (2016)

[5] Brown, Myron Z., Darius Burschka, and Gregory D. Hager.: Advances in computational stereo. In: IEEE transactions on pattern analysis and machine intelligence 25, vol. 8, pp. 993-1008 (2003)

[6] Hansen, Mark F., Gary A. Atkinson, Lyndon N. Smith, and Melvyn L. Smith.: 3D face re-constructions from photometric stereo using near infrared and visible light. In: Computer Vi-sion and Image Understanding 114, vol. 8 pp. 942-951 (2010)

[7] Sun, Yujuan, Junyu Dong, Muwei Jian, and Lin Qi.: Fast 3D face reconstruction based on uncalibrated photometric stereo. In: Multimedia Tools and Applications 74, vol. 11, pp. 3635-3650 (2015)

[8] Chang, K. B. K. I., Kevin Bowyer, and Patrick Flynn.: Face recognition using 2D and 3D facial data. In: ACM Workshop on Multimodal User Authentication, pp. 25-32 (2003)

[9] Sabater, Neus, J-M. Morel, and Andrés Almansa.: How accurate can block matches be in stereo vision?. In: SIAM Journal on Imaging Sciences 4, vol. 1, pp. 472-500 (2011)

[10] Sellahewa, Harin, and Sabah A. Jassim.: Image-quality-based adaptive face recognition. In: IEEE Transactions on Instrumentation and measurement 59, vol. 4, pp. 805-813 (2010)

[11] Brahnam, Sheryl.: Local binary patterns: new variants and applications. Springer Berlin Hei-delberg (2016)

[12] Yang, Meng, and Lei Zhang.: Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In: European conference on computer vision, pp. 448-461 (2010)

[13] Jing, Xiao-Yuan, David Zhang, and Yuan-Yan Tang.: An improved LDA approach. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34, vol. 5, pp. 1942-1951 (2004)

[14] Kautkar, Satyajit N., Gary A. Atkinson, and Melvyn L. Smith.: Face recognition in 2D and 2.5 D using ridgelets and photometric stereo. In: Pattern recognition 45, vol. 9, pp. 3317-3327 (2012)

[15] Villarini, Barbara, Athanasios Gkelias, and Vasilios Argyriou.: Photometric Stereo for 3D face reconstruction using non linear illumination models. In: IAPR Workshop on Multimod-al Pattern Recognition of Social Signals in Human-Computer Interaction, pp. 140-152 (2016)

[16] Lee, Sang-Woong, Patrick SP Wang, Svetlana N. Yanushkevich, and Seong-Whan Lee.: Noniterative 3D face reconstruction based on photometric stereo. International Journal of Pattern Recognition and Artificial Intelligence 22, vol. 03, pp. 389-410 (2008)

[17] Lao, Shihong, Yasushi Sumi, Masato Kawade, and Fumiaki Tomita.: 3D template matching for pose invariant face recognition using 3D facial model built with isoluminance line based stereo vision. In: Proceedings 15th International Conference on Pattern Recognition, vol. 2, pp. 911-916 (2000)

[18] Georghiades, Athinodoros S., Peter N. Belhumeur, and David J. Kriegman.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE transactions on pattern analysis and machine intelligence 23, vol. 6, pp. 643-660 (2001)

[19] Hansen, Mark F., Gary A. Atkinson, and Melvyn L. Smith.: BRDF estimation for faces from a sparse dataset using a neural network. In: International Conference on Computer Analysis of Images and Patterns, pp. 212-220 (2013)

[20] Kee, Seok Cheol, Kyoung Mu Lee, and Sang Uk Lee.: Illumination invariant face recogni-tion using photometric stereo. IEICE TRANSACTIONS on Information and Systems 83, vol. 7, pp. 1466-1474 (2000)

[21] Abate, Andrea F., Michele Nappi, Daniel Riccio, and Gabriele Sabatino.: 2D and 3D face recognition: A survey. Pattern recognition letters 28, vol. 14, pp. 1885-1906 (2007)

[22] Paysan, Pascal, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter.: A 3D face model for pose and illumination invariant face recognition. In: AVSS'09. Sixth IEEE International Conference on Advanced video and signal based surveillance, pp. 296-301 (2009)

[23] Soltana, W. Ben, Di Huang, Mohsen Ardabilian, Liming Chen, and C. Ben Amar.: Compar-ison of 2D/3D features and their adaptive score level fusion for 3D face recognition. In: 3D Data Processing, Visualization and Transmission (2010)

[24] Adini, Yael, Yael Moses, and Shimon Ullman.: Face recognition: The problem of compen-sating for changes in illumination direction. In: IEEE Transactions on pattern analysis and machine intelligence 19, vol. 7, pp. 721-732 (1997)

[25] Zhao, Wenyi, Rama Chellappa, P. Jonathon Phillips, and Azriel Rosenfeld.: Face recogni-tion: A literature survey. ACM computing surveys (CSUR) 35, vol. 4, pp. 399-458 (2003)

[26] Baltrušaitis, Tadas, Peter Robinson, and Louis-Philippe Morency.: Openface: an open source facial behavior analysis toolkit. In: Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, pp. 1-10 (2016)

[27] Schroff, Florian, Dmitry Kalenichenko, and James Philbin.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vi-sion and pattern recognition, pp. 815-823 (2015)

[28] Viola, Paul, and Michael J. Jones.: Robust real-time face detection. International journal of computer vision 57, vol. 2, pp. 137-154 (2004)

[29] Zafeiriou, Stefanos, Mark Hansen, Gary Atkinson, Vasileios Argyriou, Maria Petrou, Melvyn Smith, and Lyndon Smith. The photoface database. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on, pp. 132-139 (2011)

[30] VGGFace Homepage, http://www.robots.ox.ac.uk/~vgg/software/vgg_face/, last accessed 2019/02/03

[31] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In: BMVC, vol. 1, pp. 6 (2015)

[32] LFW Dataset Homepage, http://vis-www.cs.umass.edu/lfw/, last accessed 2019/01/19

[33] YouTube Faces Dataset Homepage, https://www.cs.tau.ac.il/~wolf/ytfaces/, last accessed 2019/01/17

[34] Zafeiriou, Stefanos, Gary A. Atkinson, Mark F. Hansen, William AP Smith, Vasileios Ar-gyriou, Maria Petrou, Melvyn L. Smith, and Lyndon N. Smith.: Face recognition and verifi-cation using photometric stereo: The photoface database and a comprehensive evaluation. IEEE transactions on information forensics and security 8, vol. 1 pp. 121-135 (2013)

[35] Kasturi, Rangachar, Dmitry Goldgof, Padmanabhan Soundararajan, Vasant Manohar, John Garofolo, Rachel Bowers, Matthew Boonstra, Valentina Korzhova, and Jing Zhang.: Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. IEEE Transactions on Pattern Analysis and Machine In-telligence 31, vol. 2, pp.319-336 (2009)

[36] Huang, Shih-Ming, and Jar-Ferr Yang. Improved principal component regression for face recognition under illumination variations. IEEE signal processing letters 19, vol. 4, pp.179-182 (2012)

[37] Paysan, Pascal, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter.: A 3D face model for pose and illumination invariant face recognition. In: Sixth IEEE Interna-tional Conference on Advanced Video and Signal Based Surveillance, pp. 296-301 (2009)

[38] Turk, Matthew A., and Alex P. Pentland. Face recognition using eigenfaces. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586-591 (1991)