FACE - Face At Classroom Environment: Dataset and Exploration

Oscar Karnalim*, Setia Budi*, Sulaeman Santoso*, Erico D. Handoyo*, Hapnes Toba*, Huyen Nguyen†, Vishv Malhotra‡

*Faculty of Information Technology, Maranatha Christian University, Indonesia
{Oscar.Karnalim, Setia.Budi, Sulaeman.Santoso, Erico.DH, HapnesToba}@it.maranatha.edu

†UNSW Art & Design, University of New South Wales Sydney, Australia
Huyen.Nguyen@unsw.edu.au

‡Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, India
vmm@iitg.ac.in

***Abstract*—The rapid development in face detection study has been greatly supported by the availability of large image datasets, which provide detailed annotations of faces on images. However, among a number of publicly accessible datasets, to our best knowledge, none of them are specifically created for academic applications. In this paper, we propose a systematic method in forming an image dataset tailored for classroom environment. We also made our dataset and its exploratory analyses publicly available. Studies in computer vision for academic application, such as an automated student attendance system, would benefit from our dataset.**

***Index Terms*—computer vision, face detection, face recognition, image dataset, data collection, automated attendance system, educational data mining**

I. Introduction

Face detection can be defined as a task to determine whether or not there are human faces in a given image; if present, the location of each face is then yielded. In other words, face detection is a task of localising and extracting face region from an image [1], [2]. It is one of the most studied topics in computer vision domain and has proven to be a difficult computational problem [3]–[6].

Among number of different approaches in face detection study, HAAR Cascade (also known as Viola-Jones detector) [7]–[9] is one of the most notable techniques. It is the first technique which makes face detection practically feasible for real-world applications. Its open source implementation is included within the OpenCV library [10], enabling this technique to be widely used in various applications.

The availability of large and properly labelled datasets play a crucial role in the development of object detection, especially face detection. These publicly available datasets are useful in investigating and analysing different approaches in face detection. The fact that these datasets are often used to train and evaluate face detection models also implies that the existing automated face detection techniques are inadequate and human involvement is still required in the labelling process [11]. Such a laborious process results in the slow formation of datasets.

Following are some notable datasets in face detection study. ImageNet [12], [13] is a large scale object detection/recognition dataset which started in 2010 and currently continues to be one of the most popular datasets in computer vision [11]. Though it is not a dedicated dataset to face detection, person labelling is covered by the dataset. By the time of the manuscript writing, ImageNet has 952K images under person category. PASCAL VOC [14] is another well-known dataset in object detection/recognition. Similar to ImageNet, person images are also covered. INRIA person dataset [15], [16] contains images of humans cropped from various personal photos. It is a popular dataset and has significant contribution in the area of pedestrian detection. More specific datasets on face detection/recognition could be found in Sheffield Face Database [17], UCI's Annotated Faces in the Wild [18], and UMass Face Detection Dataset and Benchmark [19].

Automated student attendance systems for schools and universities are one of many interesting applications of face detection (and recognition) [20]–[23]. It is argued that conventional ways in recording student attendance, e.g., roll call and attendance sheet signing, are considered inefficient and prone to bogus attendance. Such application of face detection is attractive and it fits with our initiative to build a computer vision system for our department. We then assessed the possibility of implementing face detection in our laboratory classroom environment.

Despite a growing number in face detection applications for academic purposes, only a few of the datasets have been made publicly accessible. In addition, among the available datasets (to our best knowledge), none of them were primarily crafted from an academic environment. Our work aims to fill this gap by proposing a systematic method in forming an image dataset that is specifically tailored for classroom environments. We also made our dataset, equipped with its quantitative and qualitative analyses, publicly available. Our dataset is envisioned

978-1-5386-6428-5/18/$31.00 ©2018 IEEE

to benefit scientists exploring face detection applications for the academic environment and also educational experts in educational data mining.

II. Methodology

We divided our work into five stages (see Fig. 1); where a *gold standard dataset* with its exploratory analyses are the primary outcomes of this study. The dataset consists of images of students sitting in a laboratory classroom, where each student's face is tagged. Exploratory evaluation of the dataset would cover both qualitative and quantitative analyses.

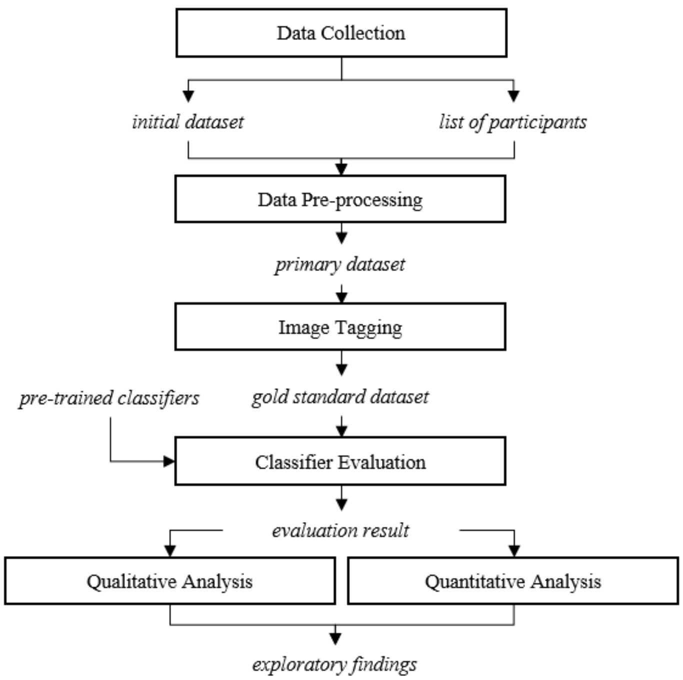

Fig. 1. Proposed methodology to generate dataset and exploratory findings. The dataset is generated and is used later to infer exploratory findings by evaluating *pre-trained classifiers* toward the *gold standard dataset*.

A. Data Collection

Data collection stage involved taking images of students sitting in a laboratory classroom. We use the term "laboratory classroom" to describe our classroom environment, which is in the form of a computer laboratory (as illustrated in Fig. 2). Five different classrooms, each accommodating up to 40 students, were under our observation. Images were taken weekly within the period of 11 weeks; two different smartphones were used to take the images: BlackBerry Passport and Asus Zenfone 2 Selfie.

Participant consent forms were distributed to students during the data collection stage. The forms were used to identify students who were willing to participate in the study.

Two types of data were collected from this stage: images of students sitting in the classrooms and a list of participants. In total, there were 90 students participating in this study and 194 images were assembled. We then referred to this image collection as the *initial dataset*.

Fig. 2. Typical laboratory classroom environment under this study. Data collection involved taking images of students sitting in the classroom. Weekly images were taken column-wise at the front.

B. Data Pre-processing

Prior to the image tagging stage, a data pre-processing was required. This involved two pre-processing procedures: image resizing and data cleansing. We standardised image size in our dataset by proportionally resizing each image with 1,024 pixels as the widest size. Such standardised size would make these images more comparable, considering they were taken with two different devices and were varied in size. In addition, image size reduction may improve efficiency in the face detection process because it reduces the detection area. We then proceeded with data cleansing phase to exclude non-participants' faces from the *initial dataset*. It was done manually by covering each non-participant's face with a red rectangle. We named the outcome of this pre-processing stage the *primary dataset*.

C. Image Tagging

This stage aimed to transform our *primary dataset* into a *gold standard dataset* that contains properly tagged faces of students. We are aware that a fully manual tagging procedure would be highly laborious. In order to lighten such laborious procedure, *gold standard dataset* in our study was obtained in twofold. First, an automated face detection was applied by using a pre-trained classifier [24] provided in the OpenCV library. Second, we then manually inspected and corrected the automated detection result. The correction involved removing mistakenly detected objects as human faces (i.e., false positive detections) and tagging faces which failed to be detected as human faces (i.e., false negative detections). For quality assurance purpose, this manual inspection was conducted iteratively where four staff were involved.

Each detected face was tagged as a rectangle. Four variables were used to represent each rectangle: x, y, width, and height. The meta-data that described face detection for each image was stored in JSON format. The collection of images paired with their meta-data formed our *gold standard dataset*.

D. Classifiers Evaluation

Four pre-trained classifiers for frontal face detection [25] were evaluated in this stage toward our *gold standard dataset*.

These classifiers were formed based on the Viola-Jones detector [7], [8] where each classifier was pre-trained with different training datasets. These pre-trained classifiers are publicly available within the OpenCV library. In our work, we adopted these classifiers and referred to each of them as *al2*, *alt*, *def*, and *tre*. They were then applied to locate faces in all the images of our dataset. We used the same rectangle representation as in our *gold standard dataset* to describe the meta-data of detected faces. We named the collection of meta-data for all classifiers as the *evaluation result*.

E. Quantitative and Qualitative Analyses

We further employed both qualitative and quantitative analyses on the *evaluation result*. Qualitative analysis was conducted by manually inspecting each detection performed by each classifier. Some notable factors which may influence the detection performance were then highlighted.

$$\begin{aligned} sensitivity &= \frac{TP}{TP + FN} \times 100 \\ precision &= \frac{TP}{TP + FP} \times 100 \end{aligned} \tag{1}$$

We performed quantitative analysis by first quantifying the number of true positive (TP), false positive (FP), and false negative (FN) detections for each detection result; compared toward our *gold standard dataset*. Two metrics were employed to measure the detection performance: sensitivity [26] (i.e., also known as true positive rate: among all faces, how many were detected) and precision [27] (i.e., also known as positive predictive value: among all detected faces, how many were real faces). Both metrics were calculated as a percentage (see Eq. 1).

III. Results and Discussion

A. Dataset

Dataset is the primary contribution of this study and it is manifested in two forms. The first is in the form of images: each image capturing number of students sitting in a laboratory classroom environment. The second is in the form of meta-data recorded and stored in JSON format. Each meta-data serves as an annotation of faces present in every image. It is produced through a laborious procedure of image tagging. These images equipped with their meta-data formed our *gold standard dataset*.

This study also produced another meta-data called *evaluation result*. Such meta-data was generated by employing four pre-trained classifiers toward our *gold standard dataset*. This meta-data was later used to evaluate the performance of each classifier.

Our dataset could be utilised in several different studies. The most prominent one is in the study of face detection/recognition, especially when classrooms are the environment of interest. Scientists could evaluate their proposed method on our *gold standard dataset*. Moreover, studies in computer vision targeted automated student attendance systems would benefit from our dataset. A more general study in face detection/recognition and other studies in educational data mining could also take advantage of our dataset.

In order to maximise the benefit of the dataset, we made our dataset publicly accessible as a Github repository [28]. We plan to maintain the dataset by continuously collecting new data and updating it to our repository. Tool and scripting that we used to compile and analyse our dataset are also included in the repository.

B. Qualitative Findings

In our qualitative analysis, the conditions that can contribute to the detection performance of each pre-trained classifier were identified. We classified these conditions into two categories: favourable and unfavourable conditions. Five factors were taken into account in the analysis: pose, occlusion, lighting condition, image quality, and face-like shape (see Table I). Meta-data from the previously produced *evaluation result* was utilised for our analysis.

From the qualitative analysis results, we discovered that facial pose had effect on the detection performance. Non-frontal faces (e.g., tilted face, horizontally or vertically rotated face) often failed to be detected and located. It is not surprising due to the fact that the pre-trained classifiers used in this study were built for frontal face detection. In addition, we found that some extreme facial expressions may also lead to a detection failure. Fig. 3 presents some examples that show the effect of poses on face detection performance. On the other hand, eye expression (even though it is part of facial expression) has no effect on the detection performance (see Fig. 4).

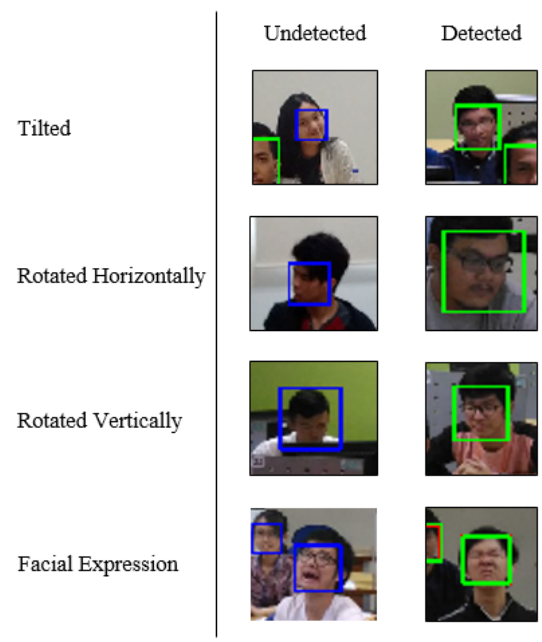

Fig. 3. Examples of false negative and true positive detections, marked with blue and green rectangles respectively, related to pose.

Furthermore, occlusions on key facial features (e.g., forehead, nose, cheek) would significantly degrade the face detection performance. We further categorised objects of occlusion

TABLE I
CONTRIBUTING FACTORS IN FACE DETECTION

Aspect	**Favourable Condition**	**Unfavourable Condition**
Pose	Frontal face, normal facial expression	Non-frontal face, extreme facial expression
Occlusion	Clear key facial features	Occlusions of key facial features by internal facial attribute (e.g., hair-style), external facial attribute (e.g., spectacles, hijab, hat), or other object (e.g., a student sitting in front of other student)
Lighting condition	Sufficient lighting condition with no intrusive light reflection or shadow	Intrusive light reflection or shadow which conceals key facial features or results to face-like pattern
Image quality	Clear image with complete faces	Blurry image where key facial features are concealed, incomplete face
Face-like shape	No face-like shape present	Face-like shape is present

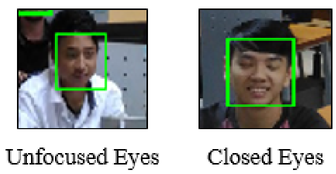

Fig. 4. Examples of true positive detections, marked with green rectangles, related to eye conditions.

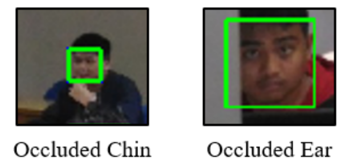

Fig. 6. Examples of true positive detections, marked with green rectangles, related to occlusion on non-key facial features.

in this study into three categories: internal facial attributes, external facial attributes, and other objects (see Fig. 5). We defined internal facial attributes as facial attributes that belong to human body part such as hair-style. External facial attributes cover facial attributes which are not a part of human body (e.g., spectacles, hijab, hat). Objects of occlusion are classified as other objects that are not categorised into the internal or external facial attributes. This may include a student sitting in front of other student, computer monitor, or non-facial body part such as hands. We found that occlusion on non-key facial features, for instance chin or ear, has no effect on the detection result (see Fig. 6).

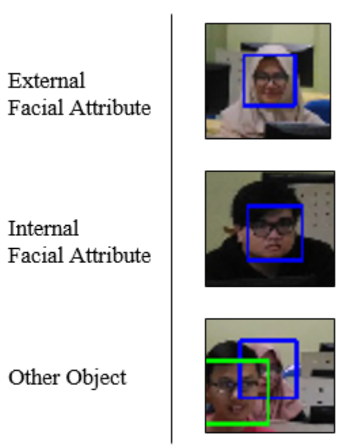

Fig. 5. Examples of false negative detections, marked with blue rectangles, related to occlusions.

Overall, images within our dataset were taken under sufficient lighting conditions. However, we found that lighting related factors could still affect the detection performance (see Fig. 7). Light reflection on spectacles, as an example, often leads to a false negative detection. Similar conditions also occurs on students with oily faces. Such reflections, to some extent, could alter or even hinder some key facial features including the eyes, forehead, or nose. In addition, light reflection on some objects such as the floor or desks may form a face-like shape, which results in false positive detections. Similar result could also be caused by the presence of shadows. When combined with other objects in a classroom, shadows may form a face-like shape. We also discovered that facial attribute, such as hijab or hair, with a colour similar to skin-tone may lead to a false negative detection.

Poor image quality could also degrade the performance of face detection. Even though Viola-Jones detector is robust enough to handle low resolution images, it does not mean that the method is able to deal with incomplete or blurry facial image where key facial features are concealed (see Fig. 8).

False positive detections in our study mainly result from objects whose patterns are similar to the shape of human faces. These patterns are not necessarily apparent for us as humans and the object could be anything ranging from clothes, jackets, desks, computer separators, or the combination of these features (see Fig 9).

Among the four evaluated classifiers in this study, *def* was the most sensitive one. This classifier was able to locate some faces which other classifiers failed to detect. However, such sensitivity also contributed to the increase in false positive

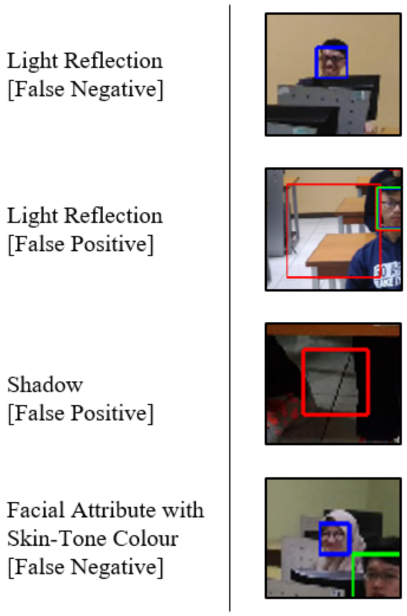

Fig. 7. Examples of false positive and false negative detections, marked with blue and red rectangles respectively, related to lighting condition.

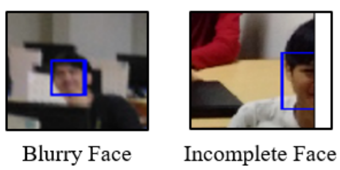

Fig. 8. Examples of false negative detections, marked with blue rectangles, related to image quality.

detections as it often mis-classified face-like patterns as faces. In contrast, *tre* was the least sensitive classifier among the four. The faces that were successfully located by other classifiers were often missed when *tre* was applied. This made *tre* the classifier with the lowest number of true positive detections and the highest number of false negative detections.

C. Quantitative Findings

In our quantitative analysis, number of true positive, false negative, and false positive detections resulted from the four pre-trained classifiers were quantified. Our *gold standard dataset* was applied as the baseline to evaluate the detection results. Figure 10 shows the evaluation results as boxplots. This figure points out that among the four classifiers employed in this study, *tre* has the worst detection performance. This could be inferred from its low true positive detections and its high false negative ones. Although *al2*, *alt*, and *def* yielded a fairly similar degree of true positive detections, *alt* is slightly better as indicated by its shorter whisker.

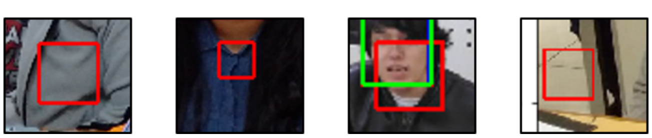

Fig. 9. Examples of false positive detections, marked with red rectangles, related to face-like shape.

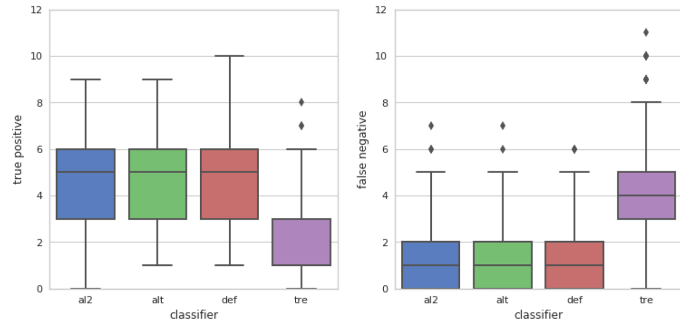

Fig. 10. Comparison of true positive and false negative detections among four pre-trained classifiers toward our *gold standard dataset*.

We then measured the True Positive Rate (TPR) and the Positive Predictive Value (PPV), also known as sensitivity and precision respectively, as presented in Fig. 11. Sensitivity enables us to understand the proportion of the successfully detected faces compared to the actual human faces available within the dataset. Among the four, *tre* yielded the lowest sensitivity. This could be explained by its high false negative detections, where *tre* often failed to detect faces.

Precision helps us to understand the proportion of actual human faces which are successfully detected among all the detected faces. It is clearly shown in Fig. 11 that all four classifiers have approximately similar level of precision; although *def* has a slightly lower precision level. As we discussed in our qualitative analysis section, *def* was able to detect faces which often failed when other classifiers were applied. However, *def* also introduced a high number of false positive detections where face-like patterns were often mis-classified as faces.

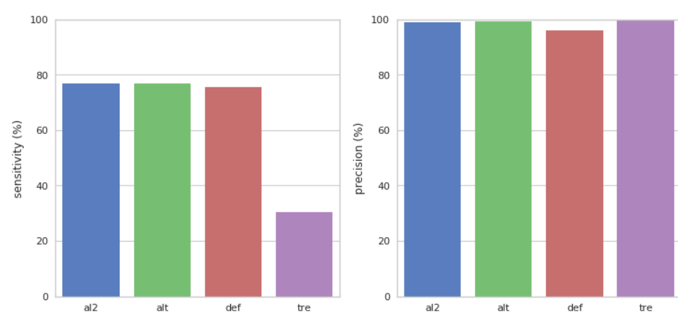

Fig. 11. Measure of sensitivity and precision across the four pre-trained classifiers toward our *gold standard dataset*.

Our quantitative findings suggested that *al2* and *alt* are the favourable pre-trained classifiers for our dataset, as they yielded a comparatively high degree of sensitivity and precision, although *alt* is slightly better compared to *al2* in terms of true positive detection counts. In contrast, *tre* and *def* are the least favourable classifiers as they yielded the lowest level of sensitivity and precision respectively.

IV. Conclusion

The availability of large image datasets has been a contributing factor to the rapid development in face detection study. The dataset is mainly in the form of image collection and is produced through a laborious image tagging procedure. Despite the growing number of face detection/recognition approaches for academic applications, only few datasets have been made publicly accessible. In addition, among the numbers of publicly available datasets, none of them are formed for academic applications.

In this work, we propose a systematic method in forming an image dataset which is specifically crafted for the classroom environment. Each image within the dataset is properly annotated, capturing faces of students sitting in a laboratory classroom environment. The dataset offers various information covering face poses, lighting related factors, and objects of occlusion; providing a base for face detection/recognition approaches to be evaluated. We made our dataset, including its exploratory analyses, available on-line for public access. This would enable scientists and interested parties across the globe to take advantage of the dataset for their research and development. Exploration in automated student attendance systems would benefit from our dataset. Studies in face detection technique could also utilise our dataset as an alternative dataset to train and/or evaluate their proposed method. In addition, we plan to maintain the dataset by continuously adding new data into our repository.

It is also important to note that our dataset has a fundamental limitation in terms of participants, who are mostly Indonesian students. Such limitations might reduce the variability of subjects in the face detection algorithm evaluation. However, it could be overcome by further contribution of images taken in other classroom environments from the global community.

Acknowledgment

The authors acknowledge Ariel Elbert Budiman and Gisela Kurniawati for their contribution in the image tagging procedure. We also thank Peter Kim for proofreading.

References

[1] Ming-Hsuan Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[2] E. Hjelmas and B. K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236–274, 2001.

[3] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Computer Vision and Image Understanding*, vol. 138, pp. 1–24, 2015.

[4] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is Real-World Visual Object Recognition Hard?" *PLoS Computational Biology*, vol. 4, no. 1, p. e27, 2008.

[5] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[6] G. Yang and T. S. Huang, "Human face detection in a complex background," *Pattern Recognition*, vol. 27, no. 1, pp. 53–63, 1994.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*. IEEE, 2001, pp. I–511–I–518.

[8] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[9] A. Kasinski and A. Schmidt, "The architecture and performance of the face and eyes detection system based on the Haar cascade classifiers," *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 197–211, 2010.

[10] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[11] K. Gauen, R. Dailey, J. Laiman, Y. Zi, N. Asokan, Y.-H. Lu, G. K. Thiruvathukal, M.-L. Shyu, and S.-C. Chen, "Comparison of Visual Datasets for Machine Learning," in *Information Reuse and Integration*. IEEE, 2017, pp. 346–355.

[12] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.

[13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[14] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[15] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Computer Vision and Pattern Recognition*. IEEE, 2005, pp. 886–893.

[16] M. Taiana, J. C. Nascimento, and A. Bernardino, "An Improved Labelling for the INRIA Person Data Set for Pedestrian Detection," in *Pattern Recognition and Image Analysis*. Springer, 2013, vol. 7887, pp. 286–295.

[17] D. B. Graham and N. M. Allinson, "Characterising Virtual Eigensignatures for General Purpose Face Recognition," in *Face Recognition*. Springer, 1998, pp. 446–456.

[18] Xiangxin Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2879–2886.

[19] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," University of Massachusetts, Amherst, Tech. Rep. UM-CS-2010-009, 2010.

[20] E. Rekha and P. Ramaprasad, "An efficient automated attendance management system based on Eigen Face recognition," in *Cloud Computing, Data Science and Engineering*. IEEE, 2017, pp. 605–608.

[21] S. Lukas, A. R. Mitra, R. I. Desanti, and D. Krisnadi, "Student attendance system in classroom using face recognition technique," in *Information and Communication Technology Convergence*. IEEE, 2016, pp. 1032–1035.

[22] V. Shehu and A. Dika, "Using real time computer vision algorithms in automatic attendance management systems," in *Information Technology Interfaces*. IEEE, 2010, pp. 397–402.

[23] J. H. Lim, E. Y. Teh, M. H. Geh, and C. H. Lim, "Automated classroom monitoring with connected visioning system," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, 2017, pp. 386–393.

[24] "opencv/haarcascade_frontalface_alt·xml at master · opencv/opencv · GitHub." [Online]. Available: https://goo.gl/gKN6v5

[25] "opencv: Open Source Computer Vision Library," feb 2018. [Online]. Available: https://github.com/opencv/opencv

[26] D. G. Altman and J. M. Bland, "Statistics Notes: Diagnostic tests 1: sensitivity and specificity," *BMJ*, vol. 308, no. 6943, p. 1552, 1994.

[27] D. G. Altman and J. M. Bland, "Statistics Notes: Diagnostic tests 2: predictive values," *BMJ*, vol. 309, no. 6947, p. 102, 1994.

[28] itmaranatha, "Face - face at classroom environment," mar 2018. [Online]. Available: https://github.com/itmaranatha/classroom_dataset