Children Face Long-term Identification in Classroom: Prototype Proposal

Nikolajs Bumanis[1][a], Gatis Vitols[2][b], Irina Arhipova[2][c] and Inga Meirane[2]
[1]Faculty of Information Technologies, Latvia University of Life Sciences and Technologies, 2 Liela str., Jelgava, Latvia
[2]WeAreDots Ltd., Elizabetes str. 75, Riga, Latvia

Keywords: Face Recognition, Deep Learning, Long-term Identification.

Abstract: A children face automated identification raise additional challenges compared to an adult face automated identification. A long-term identification is used in the environment in which a person must be identified in longer time spans, such as months and years. A long-term identification is present for example in schools where children spend multiple years and, if automated face identification solution is implemented, it must be resilient to recognise face biometrical data in the span of typically up to 9 years. In this proposal, we discuss children face identification available solutions which use deep learning networks, introduce legal constraints that come with privacy of children and propose prototype for a long-term identification of children attendance in their classroom. The solution consists of a developed prototype that is architecturally separated into three layers. The layers encapsulate necessary local and remote hardware, software and interconnectivity solutions between these entities. The protype is intended for implementation into a school's class attendance management system, and should provide sufficient functionality for person's identity management, object detection and person's identification processes. The prototype's processing is based on the model that incorporates the principles of multiple correct biometric pattern versions, providing possibility of a long-term identification. The model uses Single Shot MultiBox Detector for object detection and Siamese neural network for a person identification.

1 INTRODUCTION

Human face recognition automation has been broadly researched especially with the improvement in computational power and the development of AI algorithms and implementations. Various solutions have been developed and trained with different data sets, for example, for age estimation (Anda et al., 2019; Huang, Li, Zhu, & Chen, 2017). A large portion of research dedicated to automated face recognition comes from the realm of a smart city development and a citizen recognition in a real-time for security (Wu, Xu, & Li, 2020).

A long-term children face identification can be a part of more a complex solution for overall management of learning facilities with the emphasis on security, a class emotional climate and the support of an existing school management system, such as automated counting of pupils in the classroom. A short term identification of a pupil can be achieved easier than long-term. Children spend time in school for multiple years and during that time various constraints on identification need to be addressed, such as changes in face, changes of dressing styles, cosmetic variations of faces, etc.

The aim of the research is to propose a prototype for children face long-term identification in learning facilities.

2 CHILDREN FACE IDENTIFICATION CHALLENGES

Automated identification and re-identification of baby, toddler and children faces can raise some

[a] https://orcid.org/0000-0002-1884-7731
[b] https://orcid.org/0000-0002-4131-8635
[c] https://orcid.org/0000-0003-1036-2024

Bumanis, N., Vitols, G., Arhipova, I. and Meirane, I.
Children Face Long-term Identification in Classroom: Prototype Proposal.
DOI: 10.5220/0009414002870293
In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) - Volume 2*, pages 287-293
ISBN: 978-989-758-423-7

challenges which are not present in the face identification of adults. The main challenges are faster changes in certain face biomarkers and stricter privacy control policies for implementation.

2.1 Deep Learning for Automation of Children Face Identification

The research on younger age people face recognition typically focuses on children face recognition for their identification in classroom settings in schools or other establishments and a face emotion recognition (Jaiswal & Nandi, 2019).

Classroom observation automated solutions have been researched in recent years. There are promising solutions (Ramakrishnan, Ottmar, LoCasale-Crouch, & Whitehill, 2019) in detection of an emotional status of a class (positive or negative) with an improved success using ensemble learning using such deep learning architectures as CNN and Bi-LSTM.

In India, children face identification after kidnaping has been improved with deep learning using CNN model and VGG Face Descriptor(Chandran et al., 2018). Another article (Siddiqui, Vatsa, & Singh, 2018) proposes class-based penalties mechanism for CNN to improve the recognition of toddler images.

By testing 8 datasets and using CNN network with reduction of hyperparameters researchers achieved 74% accuracy in children face emotion detection (Jaiswal & Nandi, 2019).

The analysis of children behaviour in the classroom is done by using a security camera in the classroom and by analysing images from cameras (Rothoft, Si, Jiang, & Shen, 2017). To detect if children pay attention to the designed spatial area, distribution of the focus points in two dimensions is performed. After that the identification of anomalous points is identified.

Face recognition for students' attendance has been developed (Prangchumpol, 2019) using Android Face Recognition service which is linked to the database on cloud storage. The information is automatically added to a web-based attendance management system.

The authors propose (Bhattacharya, Nainala, Das, & Routray, 2018) the solution to prevent fake attendance by using CNN network with loss function and tools to improve image quality reaching above 80% accuracy. Authors use 15MP single camera.

Another article (Dalal, Dalal, & Dalal, 2019) for the same attendance identification task propose to use Viola-Jones algorithm for the detection of face region in image and feedforward neural network Extreme learning machines for identification.

A typical recording setting for children identification in the classroom is done with single camera and few participants(Bhattacharya et al., 2018; Lin & Li, 2019). However, less research address issue of more people in the classroom and possible implementation of multiple camera recording and synchronisation of data between cameras.

A recent research shows improved face and face expression identification and verification results with application of Siamese Networks based on CNN(Dwi Putranto & Wahyono, 2019; Hayale, Negi, & Mahoor, 2019)

2.2 Privacy Policies

Children's biometric data is regulated by multiple entities. In Europe, GDPR impact the way data, including biometric data, is handled. Some articles (Sanchez-Reillo, Ortega-Fernandez, Ponce-Hernandez, & C. Quiros-Sandoval, 2018) address biometric data and legal regulation relationships proposing particular procedures to handle analysed data according to the regulations.

Some of the countries implement special acts dedicated to people freedom and privacy. For example, in the United Kingdom there is Protection of Freedom Act 2012 (Parliament of the United Kingdom, 2011) which first chapters address Regulation of biometric data and Regulation of surveillance.

For children biometrical data protection, there is a separate regulation such as "Protection of biometric information of children in schools and colleges" (U.K. Department for Education, 2018) which provide guidance for schools and possible implementation of local school policies. Then also each school has internal rule and regulation document, typically including section about safety (Jules Verne Riga French School council, 2014).

3 PROTOTYPE ARCHITECTURE

The prototype is developed for school attendance management process. A typical management process includes manual observation and fixation of each student's attendance into the electronic system or journal. At the end of each semester, the attendance readings are aggregated and used to determine necessary procedures, for example, to allow a student to take an exam, to prevent a student from taking an exam before he completes additional tasks, etc.

A proposed prototype consists of multiple interconnected software and hardware components. The architecture can be described by three layers - physical layer, data security layer and logical layer (see Fig. 1).

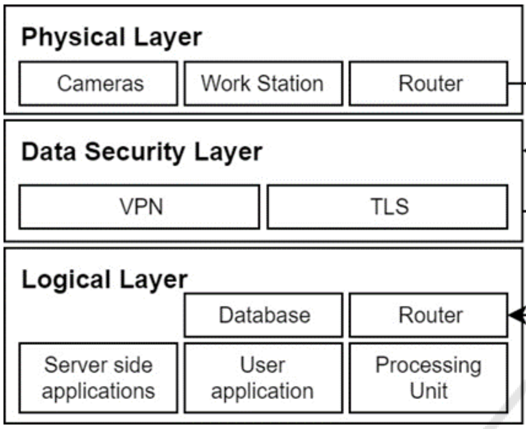

Figure 1: Stack diagram of prototype architecture.

3.1 Physical Layer

A physical layer is responsible for connectivity of hardware components by utilizing input/output interfaces of these components. The following components are included in physical layer: video cameras, Work Station and Router.

Video cameras are responsible for observation of the classroom and gathering video feed. There are multiple variations of camera placement in the classroom. The prototype uses two cameras, positioned on the front room's wall, where, in the framework of this work, we assume classroom's composition as seen in Figure 2 (see Fig. 2).

This placement provides the following benefits: the ease of camera installation because of convenient cable routing, and the encapsulation of all the classroom using two partly overlapping viewing angles for more accurate object detection and identification. However, multiple viewing angles, especially when overlapped, introduce a statistical challenge – the correct way to interpret the results for scenarios when one object is detected and identified by both cameras simultaneously.

A video feed from cameras is processed by Work Station, specifically – Processing Unit (see Section 4.2), placed at the filming location. Work Station is prepared by a service provider and has all the necessary software installed, which is described in chapter 3.3.

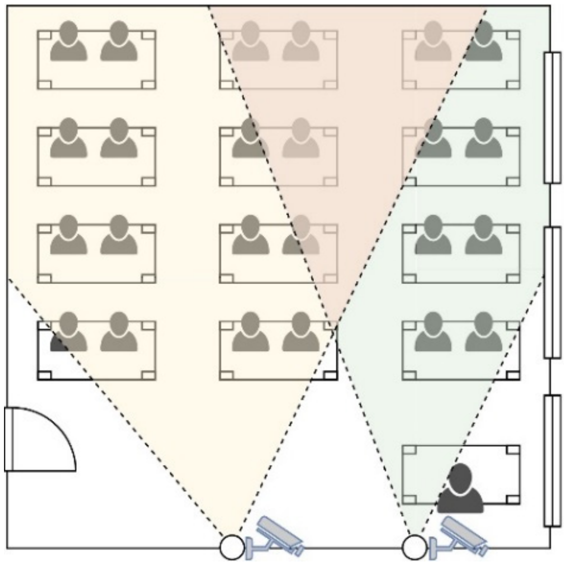

Figure 2: Proposed camera placement positions.

A physical layer includes Router used for secure connection of camera devices, Work Station and Data Centre. Security measures are described in chapter 3.2.

3.2 Data Security Layer

An intermediary level of the proposed architecture is Data security level, which describes technologies and protocols used by Router. This layer is responsible for secure connection between camera devices, Work Station and Data Centre. Connection is established using Virtual Private Network (VPN) and cryptographic protocol Transport Layer Security (TLS).

Sensitive data are placed in Data Centre, thus providing physical security of data. Access to these data is secured by VPN and TLS, and is available only to authorized and authenticated users. These measures are also effective against Main-In-The-Middle (MITM) attacks.

3.3 Logical Layer

Logical layer describes the software used by Work Station and Server, and in general is responsible for object detection and object identification processes.

The software of the proposed prototype can be divided into server side applications and a user application. Server side applications provide support for biometric data processing functionality to manage persons' biographical data, event and class attendance registers, and a persons' identification process.

Server side applications include knowledge base, created using persons' biometric data; data base; video archive, which includes video feed captured by cameras, and video cuts of a particular detected object; authentication and authorization module.

The following Server side applications are implemented as tools with corresponding methods:

- person management tools – methods for managing person's biographical and biometrical data, and this person's designation to a particular group or class;
- image quality tools – methods for image quality assessment;
- observation tools – methods for starting and stopping video recording, creation of video archive;
- classifier management tools – methods for accessing and managing classifiers;
- event management tools –methods for managing events and class attendance results;
- messaging tools – methods for implementation of messaging mechanism;
- configuration tools – methods for managing Prototype modules' settings;
- audit tools – methods for managing audit records.

The user application, installed in Work Station, provides the functionality of creation and modification of a person's biographical and biometrical profile by stating person's credentials and uploading person's face images. The user application performs identity management activities utilizing the connection between Work Station and Server and accessing Server side applications.

Biometric data processing is performed by logical Processing Unit embedded into the user application. Processing Unit is responsible for video feed processing - object detection and detected object's crop image creation; object identification using created crop image and biometric data from knowledge base. Processing Unit is also responsible for executing Server side applications' methods like creation of video archive.

Logical layer includes Data Base, which stores the following data: person's biographical and biometric data, results of object detection in the form of an event register, results of automatic person identification and expert assessment of these results in the form of a class attendance register, classifiers for teaching classes, rooms, cameras, audit records and configuration parameters.

4 PROTOTYPE FUNCIONALITY

The range of available functions depends on the authorization level of an authenticated user. Basic, teacher level authorization gives access to person management tools, observation tools with some limitations to start and end recording procedures, class attendance result register management tools and a read-only access to some classifiers; whereas advanced, administrator level authorization gives access to all existing tools. The administrator level authorization is also required to perform management of user accounts and classifiers. In addition, the prototype gives possibility to modify access rights resulting in creation of in-between level accounts.

Starting Prototype's User application, the user is offered an authorization window to input credentials. In case of a successful validation, the user is forwarded to Prototype's main page. The main page contains links to all sections available to authenticated user's authorization level. The sections correspond to User application's internal functionality or Side server application.

4.1 User Interface Functions

The sections can be divided into three main branches – managing person's identity, managing the observation process and managing the results of the observation, and one secondary branch – data base management.

The first branch is intended for development of the knowledge base, used by a person identification algorithm. The person's identity is registered by inputting biographical data, including name, surname, person identification number and the birth date. As an option, person's gender can be defined. After the definition of biographical data, person's one to four photos are uploaded; photos are taken according to image quality standards (ISO, 2011) and to ensure sufficient image quality the following parameters are validated: face recognition rate, contrast, sharpness, saturation. The final step includes selecting person's affiliated groups and classes.

The second branch is intended for the management of the filming process. The filming location and affiliated class or group are selected from pre-registered classifies, and the filming process is started by a user. The prototype supports a single filming session for up to 10 minutes, and stops it automatically after that. As an option, the filming process can be stopped manually by a user. During the filming process, the video feed is being continuously processed by Processing Unit.

The third branch is intended for managing the results of filming that includes both object detection and person identification results. Every identified person must be first detected, but not every detected object gets identified. The prototype divides these results into two sections – registered event section for object detection, and class attendance management section for person's identification and attendance of the affiliated class. In the framework of the prototype, the event is a set of frames, starting with the frame when an object was first detected, up to a frame in which the object was re-detected in the timeframe of last 3 seconds (see 4.3). By accessing the first section, the results of object detection can be observed.

These results include the following data: the event date and time, the event location, 10 seconds long video fragment with a detected object in the frame; when playing, the video is displayed in the embedded video player. In the case of a successful identification of a detected object, this detected object's name and surname, including the corresponding degree of coincidence, can also be observed.

The second section contains the results for person identification with regards to his attendance of an affiliated class. The results include metadata about the event and a list of intended participant attendance with each participant automatical identification attendance assessment. An expert assessment can be performed for each intended participant, and is commonly done by a teacher level user. Assessment values are used for statistical analysis which determines Prototype's performance in regards to the object detection and the person identification. In case of unidentified detected objects, respectively - unexpected participants, frames with these persons are included. The results can be exported in a form of a report.

A secondary branch is intended for data management processes with regards to user accounts and classifiers. Entities of school rooms, teaching classes and camera devices are registered as classifiers, and can be later modified and managed by an administrator level user. User accounts are created by the administrator. In the current stage of development the prototype does not allow creation of a user account from a work station because of security and legal rights reasons. Prototype's user authorization is realized with compliance to WS-Trust specification (OASIS, 2012).

4.2 Processing Unit

Processing Unit is a set of software algorithms which are responsible for processing video feed with the aim to detect object and identify the detected object's identity. Processing Unit realizes the previously described (Arhipova, Vitols, & Meirane, 2019) model that incorporates principles of multiple correct biometric pattern versions, providing possibility of long-term identification.

The prototype uses Single Shot Multi Box (SSD) (Liu et al., 2016) algorithm for object detection and Siamese neural network (Bielski, 2019; Koch & Koch, 2015) for person identification.

The proposed model was updated to accommodate the necessity of using two cameras. Update affects the first model's step – the processing of video feed and the creation of the object's crop image. A crop image is created during a continuous video filming process, using a captured frame with the best image quality, and is used by Siamese neural network for person identification. In the case of two cameras, only one crop image is created, decided by threshold and image quality assessment values.

4.3 Object Tracking

Object tracking, which is commonly performed by either a separate object tracking algorithm or an object detection algorithm's sub-function is implemented into the Prototype synthetically. Prototype's object tracking is based on observing the target within identical or very close proximity coordinates with the aim to determine if the object in the frame is the "same".

These calculations are performed by SSD algorithm. When an object is detected, the Prototype generates an event. During video feed processing SSD makes continuous attempts at detecting this object. SSD creates crop images of detected objects and forwards them to a person identification with Siamese neural network.

A registration process continues until the object is successfully detected within the timeframe of 3 seconds. When it happens, the registration process is stopped, and the instance of the event is recorded in the database. If object's identity is identified at any point during the registration process, a person's name and surname are also recorded.

5 CONCLUSIONS

Children face identification raises two main challenges: faster changes in face and stricter privacy control policies.

The majority of research to capture face uses few people and a single camera. Our proposal introduces two camera implementation in a classroom with synchronisation between cameras, using a router and a workstation, thus providing a possibility of further scalability and expansion.

The proposed prototype is meant to handle basic school attendance management operations with regards to object detection and person identification. However, advanced scenarios, like entry of an unintended person, introduction of additional usage and processing challenges – filming process must be stopped beforehand, and the results may be insufficient for final analysis.

The prototype uses Single Shot MultiBox Detector and Siamese neural network for the main re-identification process, where recent researches show an improved face and face expression identification and result verification with the application of Siamese Networks based on CNN.

Technically the prototype assumes correct and accurate working regime – up to 10 minutes of non-issue incurring filming and continuous processing. Furthermore, the prototype uses a remote connection to Data Centre. The potential issues which may occur during the production were not included in this paper. This requires in-depth approbation and adaptation.

Further steps include this prototype's approbation in Latvia high school. The legal permissions to execute first experiments have already been acquired.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/18/A/003 signed between IT Competence Centre and Central Finance and Contracting Agency, Research No. 2.1 "Person long-period re-identification (Re-ID) solution to improve the quality of education".

REFERENCES

Anda, F., Lillis, D., Kanta, A., Becker, B. A., Bou-Harb, E., Le-Khac, N.-A., & Scanlon, M. (2019). Improving Borderline Adulthood Facial Age Estimation through Ensemble Learning. In *Proceedings of the 14th International Conference on Availability, Reliability and Security - ARES '19* (pp. 1–8). New York, New York, USA: ACM Press. https://doi.org/10.1145/3339252.3341491

Arhipova, I., Vitols, G., & Meirane, I. (2019). Long Period Re-Identification Approach to Improving the Quality of Education: A Preliminary Study, FICC 2020. (p. In Press).

Bhattacharya, S., Nainala, G. S., Das, P., & Routray, A. (2018). Smart attendance monitoring system (SAMS): A face recognition based attendance system for classroom environment. In *Proceedings - IEEE 18th International Conference on Advanced Learning Technologies, ICALT 2018* (pp. 358–360). Institute of Electrical and Electronics Engineers Inc.

Bielski, A. (2019). Siamese and triplet networks with online pair/triplet mining in PyTorch. Retrieved August 23, 2019, from https://github.com/adambielski/siamese-triplet

Chandran, P. S., Byju, N. B., Deepak, R. U., Nishakumari, K. N., Devanand, P., & Sasi, P. M. (2018). Missing Child Identification System Using Deep Learning and Multiclass SVM. In *2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (pp. 113–116). IEEE. Retrieved from https://ieeexplore.ieee.org/document/8635054/

Dalal, A., Dalal, P., & Dalal, S. (2019). *Automatic Attendance System Using Extreme Learning Machine. International Journal of Engineering and Advanced Technology (IJEAT)*.

Dwi Putranto, R. A., & Wahyono. (2019). Alignment based siamese network model for face verification. *International Journal of Scientific and Technology Research*, *8*(10), 2577–2581.

Hayale, W., Negi, P., & Mahoor, M. (2019). Facial expression recognition using deep siamese neural networks with a supervised loss function. In *Proceedings - 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019*. Institute of Electrical and Electronics Engineers Inc.

Huang, J., Li, B., Zhu, J., & Chen, J. (2017). Age classification with deep learning face representation. *Multimedia Tools and Applications*, *76*(19), 20231–20247. Retrieved from http://link.springer.com/10.1007/s11042-017-4646-5

ISO. (2011). ISO/IEC 19794-1:2011. Information technology — Biometric data interchange formats — Part 1: Framework. Retrieved from https://www.iso.org/standard/50862.html

Jaiswal, S., & Nandi, G. C. (2019). Robust real-time emotion detection system using CNN architecture. *Neural Computing and Applications*.

Jules Verne Riga French School council. (2014). Internal Rules of Jules Verne Riga French School. Retrieved November 17, 2019, from http://www.ecolejulesverne.lv/wp-content/uploads/2015/03/Internal-rules-2014-2015.pdf

Koch, G., & Koch, G. (2015). Siamese Thesis. *Cs.Toronto.Edu*. Retrieved from http://www.cs.toronto.edu/~gkoch/files/msc-thesis.pdf

Lin, Z. H., & Li, Y. Z. (2019). Design and Implementation of Classroom Attendance System Based on Video Face Recognition. In *Proceedings - 2019 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2019* (pp. 385–388). Institute of Electrical and Electronics Engineers Inc.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

OASIS. (2012). WS-Trust 1.4. Retrieved from http://docs.oasis-open.org/ws-sx/ws-trust/v1.4/ws-trust.html

Parliament of the United Kingdom. (2011). Protection of Freedoms Act 2012. Retrieved January 6, 2020, from http://www.legislation.gov.uk/ukpga/2012/9/contents/enacted

Prangchumpol, D. (2019). Face Recognition for Attendance Management System Using Multiple Sensors. In *Journal of Physics: Conference Series* (Vol. 1335). Institute of Physics Publishing.

Ramakrishnan, A., Ottmar, E., LoCasale-Crouch, J., & Whitehill, J. (2019). Toward Automated Classroom Observation: Predicting Positive and Negative Climate. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (pp. 1–8). IEEE. Retrieved from https://ieeexplore.ieee.org/document/8756529/

Rothoft, V., Si, J., Jiang, F., & Shen, R. (2017). Monitor Pupils' Attention by Image Super-Resolution and Anomaly Detection. *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, 843–847. Retrieved from https://ieeexplore.ieee.org/document/8446759/

Sanchez-Reillo, R., Ortega-Fernandez, I., Ponce-Hernandez, W., & C. Quiros-Sandoval, H. (2018). How to Implement EU Data Protection Regulation for R&D in Biometrics. *Computer Standards & Interfaces*, *61*, 89–96.

Siddiqui, S., Vatsa, M., & Singh, R. (2018). Face Recognition for Newborns, Toddlers, and Pre-School Children: A Deep Learning Approach. *2018 24th International Conference on Pattern Recognition (ICPR)*, 3156–3161. https://doi.org/10.1109/ICPR.2018.8545742

U.K. Department for Education. (2018). Protection of children's biometric information in schools. Retrieved December 19, 2019, from https://www.gov.uk/government/publications/protection-of-biometric-information-of-children-in-schools

Wu, H., Xu, H., & Li, P. (2020). Design and Implementation of Cloud Service System Based on Face Recognition. In *Advances in Intelligent Systems and Computing* (Vol. 993, pp. 629–636). Springer Verlag.