Evaluation of different features for face recognition in video

Prepared by:
Dmitry O. Gorodnichy
Canada Border Services Agency
Ottawa ON Canada K1A 0L8

Erico Neves, Stan Matwin
University of Ottawa,
School of Electrical Engineering and Computer Science
161 Louis Pasteur
Ottawa ON Canada K1N 6N5

Eric Granger
Université Québec
1100, rue Notre-Dame Ouest
Montréal (Québec) H3C 1K3

Scientific Authority: Pierre Meunier
DRDC Centre for Security Science
613-992-0753

The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of the Department of National Defence of Canada.

Contract Report
DRDC-RDDC-2014-C250
September 2014

IMPORTANT INFORMATIVE STATEMENTS

PROVE-IT (FRiV) Pilot and Research on Operational Video-based Evaluation of Infrastructure and Technology: Face Recognition in Video project, PSTP 03-401BIOM, was supported by the Canadian Safety and Security Program (CSSP) which is led by Defence Research and Development Canada's Centre for Security Science, in partnership with Public Safety Canada. Led by Canada Border Services Agency partners included: Royal Canadian Mounted Police, Defence Research Development Canada, Canadian Air Transport Security Authority, Transport Canada, Privy Council Office; US Federal Bureau of Investigation, National Institute of Standards and Technology, UK Home Office; University of Ottawa, Université Québec (ÉTS).

The CSSP is a federally-funded program to strengthen Canada's ability to anticipate, prevent/mitigate, prepare for, respond to, and recover from natural disasters, serious accidents, crime and terrorism through the convergence of science and technology with policy, operations and intelligence.

© Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2014

© Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2014

Science and Engineering Directorate

Border Technology Division

Division Report 2014-31 (TR)
July 2014

Evaluation of Different Features for Face Recognition in Video

Erico Neves,
Dmitry Gorodnichy,
Stan Matwin,
Eric Granger

This page left intentionally blank

Abstract

With man One of the most critical tasks in automated face recognition technology is the extraction of facial features from a facial images.

The most critical task in each face recognition (FR) technology, which contributes the most to the success of particular FR products in particular applications and which is highly protected by industries developing those products, is the extraction of facial features from a facial image. This report presents the performance comparison of several publicly reported feature extraction algorithms for face recognition in video. The evaluated features are Harris corner detection features, FAST (Features from Accelerated Segment Test), GFTT (Good Features To Track), MSER (Maximally Stable Extremal Regions), and HOG (Histograms of Oriented Gradients).

Keywords: video-surveillance, face recognition in video, instant face recognition, watch-list screening, biometrics, reliability, performance evaluation

Community of Practice: Biometrics and Identity Management

Canada Safety and Security (CSSP) investment priorities:

1. Capability area: P1.6 – Border and critical infrastructure perimeter screening technologies/ protocols for rapidly detecting and identifying threats.
2. Specific Objectives: O1 – Enhance efficient and comprehensive screening of people and cargo (identify threats as early as possible) so as to improve the free flow of legitimate goods and travellers across borders, and to align/coordinate security systems for goods, cargo and baggage;
3. Cross-Cutting Objectives CO1 – Engage in rapid assessment, transition and deployment of innovative technologies for public safety and security practitioners to achieve specific objectives;
4. Threats/Hazards F – Major trans-border criminal activity – e.g. smuggling people/ material

Acknowledgements

This work is done within the project PSTP-03-401BIOM "PROVE-IT(FRiV)" funded by the Defence Research and Development Canada (DRDC) Centre for Security Science (CSS) Public Security Technical Program (PSTP) by the following contributors:

1. **D. Gorodnichy**, Science & Engineering Directorate, Canada Border Services Agency.
2. **E. Granger**, École de technologie supérieure, Université du Québec.
3. **S. Matwin, E. Neves,** School of Electrical Engineering and Computer Science of the University of Ottawa.

Disclaimer

In no way do the results presented in this paper imply recommendation or endorsement by the Canada Border Services Agency, nor do they imply that the products and equipment identified are necessarily the best available for the purpose. The information presented in this report contains only the information available in public domain.

Release Notes

Context: This document is part of the set of reports produced for the PROVE-IT(FRiV) project. All PROVE-IT(FRiV) project reports are listed below.

1. E. Granger, P. Radtke, and D. Gorodnichy, "Survey of academic research and prototypes for face recognition in video", Border Technology Division, Division Report 2014-25 (TR).
2. D. Gorodnichy, E.Granger, and P. Radtke, "Survey of commercial technologies for face recognition in video", Border Technology Division, Division Report 2014-22 (TR).
3. E. Granger and D. Gorodnichy, "Evaluation methodology for face recognition technology in video surveillance applications", Border Technology Division, Division Report 2014-27 (TR).
4. E. Granger, D. Gorodnichy, E. Choy, W. Khreich, P. Radtke, J. Bergeron, and D. Bissessar, "Results from evaluation of three commercial off-the-shelf face recognition systems on Chokepoint dataset", Border Technology Division, Division Report 2014-29 (TR).
5. S. Matwin, D. Gorodnichy, and E. Granger, "Using smooth ROC method for evaluation and decision making in biometric systems", Border Technology Division, Division Report 2014-10 (TR).
6. D. Gorodnichy, E. Granger, S. Matwin, and E. Neves, "3D face generation tool Candide for better face matching in surveillance video", Border Technology Division, Division Report 2014-11 (TR).
7. E. Neves, S. Matwin, D. Gorodnichy, and E. Granger, "Evaluation of different features for face recognition in video", Border Technology Division, Division Report 2014-31 (TR).
8. D. Gorodnichy and E. Granger, "Evaluation of Face Recognition for Video Surveillance". Also published in Proceedings of NIST International Biometric Performance Conference (IBPC 2012), Gaithersburg, March 5-9, 2012. Online at http://www.nist.gov/itl/iad/ig/ibpc2012.cfm.
9. D. Bissessar, E. Choy, D. Gorodnichy, T. Mungham, "Face Recognition and Event Detection in Video: An Overview of PROVE-IT Projects (BIOM401 and BTS402)", Border Technology Division, Division Report 2013-04 (TR).
10. D. Gorodnichy and E. Granger "PROVE-IT(FRiV): framework and results". Also published in Proceedings of NIST International Biometrics Performance Conference (IBPC 2014), Gaithersburg, MD, April 1-4, 2014. Online at http://www.nist.gov/itl/iad/ig/ibpc2014.cfm.

The PROVE-IT(FRiV) project took place from August 2011 till March 2013. This document was drafted and discussed with project partners in March 2013 at the Video Technology for National Security (VT4NS) forum. The final version of it was produced in July 2014.

Typesetting: All tabulated content in this report was produced automatically using LATEX content for improved source control, flexibility and maintainability. The report contains automatically generated hyperlink references and table of contents for easier navigation and reading on-line.

Contact: Correspondence regarding this report should be directed to DMITRY dot GORODNICHY at CBSA dot GC dot CA.

Contents

Abstract **3**

Release notes **5**

Table of Content **6**

1 Introduction **7**

2 Test-bed and procedure for a small scale evaluation of FRiV **8**
2.1 Facial video data-set . . . 8
2.2 Classification algorithm and metrics . . . 8

3 Comparative overview of facial features for face recognition **11**
3.1 Harris . . . 11
3.2 FAST Features . . . 11
3.3 GFTT Features . . . 12
3.4 MSER Features . . . 12
3.5 HOG Features . . . 13
3.6 Performance comparison results . . . 13

4 Feature evaluation on a large-scale dataset with other ML algorithms **14**

5 Conclusions **16**

References **17**

List of Figures

1 A generic biometric system for video-based face recognition (from [10]). . . . 7
2 Video clips in the NRC-FRiV data set (Figure reproduced from [7]). The numbers underneath the images indicate the number of frames in a clip (the first number) and the number of those of them where a face detected (the second number). . . . 9
3 Facial images from the ORL database. . . . 15
4 Graph presents the performance comparison among different algorithms implemented in OpenCV (Fisherfaces, Eigenfaces and LBPH)- all use distance metrics to recognize a face. HOG features are the only algorithm implemented that uses a ML algorithm (Bayesian Network) to perform face recognition. . . . 16

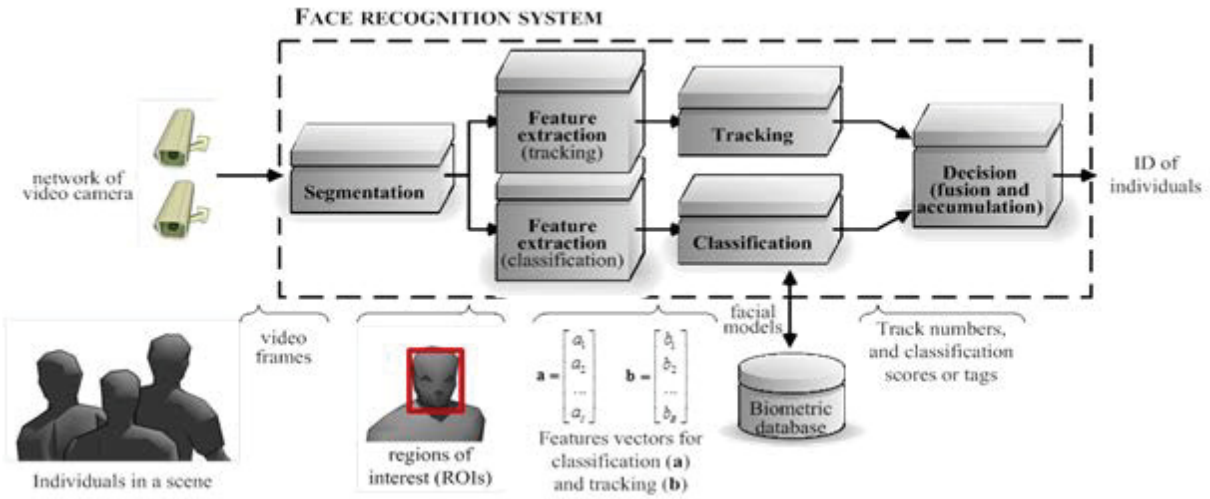

Figure 1: A generic biometric system for video-based face recognition (from [10]).

1 Introduction

As highlighted in the first report of the PROVE-IT(FRiV) project [10], one of the most critical tasks in face recognition technology is the extraction of facial features from a facial image (see Figure 1). As further presented in the second report of the PROVE-IT(FRiV) project [9], there exist several open source libraries that provide many of face recognition functions, including those required for facial feature extraction from images. These libraries are intensively used by industry and academia for in-house development of face recognition solutions.

This report presents a survey of several publicly reported feature extraction algorithms for face recognition in video, in particular those available in the OpenCV library [13]. Comparative performance analysis of these algorithms is performed for the purpose of identifying the best performing one among them.

The evaluated facial feature extraction algorithms (hereafter called simply "facial features") are are Harris corner detection features, FAST (Features from Accelerated Segment Test), GFTT (Good Features To Track), MSER (Maximally Stable Extremal Regions), and HOG (Histograms of Oriented Gradients), of which the last one is shown to perform the best.

The value of the report is seen not only in identifying the best performing publicly available facial feature extraction algorithms but also in showing a simple and efficient way of conducting a preliminary performance assessment or comparison of systems for face recognition in video (FRiV), using the NRC-FRiV data-set and such Machine Learning (ML) techniques as "Random Forests" and Synthetic Minority Oversampling Technique (SMOTE).

The report is organized as follows. First, we describe the evaluation test-bed and procedure (Section 2). Then we present the overview of the facial features, along with their performance according to the specified evaluation metrics (Section 3). Once the best performing facial feature is identified on a simpler data-set and ML algorithm, it is evaluated on a larger size still-image facial data-set using a higher complexity ML technique (Section 4). Discussions on the insights learnt and future work conclude the report.

2 Test-bed and procedure for a small scale evaluation of FRiV

2.1 Facial video data-set

Prior to conducting large-scale evaluations that take a lot of time and memory resources it is useful to pre-test the solutions to be evaluated at a small scale. Small scale evaluation is particularly helpful when it is required to select a component or a parameter for the system to be later used in a large scale evalution, instead of testing all components/parameters at a large scale.

The NRC-FRiV video database, described in [7] and which can be publicly downloaded from `http://www.videorecognition.com/FRiV`, offers convenient means to conduct such a small-scale pre-assessment evaluation for face recognition in video. This database was specifically developed for fast comparative small-scale testing of face recognition in video [**?**]. It contains eleven pairs of short low-resolution mpeg1-encoded video clips, each showing a face of a computer user sitting in front of the monitor exhibiting a wide range of facial expressions and orientations as captured by a USB webcam mounted on a computer monitor.

The video capture size is 160 x 120 pixels. With a face occupying 1/4 to 1/8 of the image (in width), this translates into a commonly observed situation on a TV screen when a face of an actor in a TV show occupies 1/8 to 1/16 of the screen.

Figure 2.1 shows 22 video clips created for this dataset, two video sequences for each of eleven registered subjects. Each video clip is about 15 seconds long, has capture rate of 20 fps and is compressed with the AVI Intel codec with bit-rate of 481 Kbps. Because of small resolution and high compression, thus created video files of person faces are very small (less than 1Mb), which makes them comparable in size to high-resolution face images such as those used e-Passports, and makes the entire video data-set easy to download and process on a limited power computer.

2.2 Classification algorithm and metrics

All tests used 10x10-fold cross-validation and used Weka [11] to execute all evaluations. Two extra programs were created to extract faces features from those videos. The first one used a generic class that exists in OpenCV (version 2.4.1), called `FeatureDetector`, which allowed the automatic extraction of Harris, FAST, GFTT and MSER features. The second program adapted a class that the program `traincascades` from OpenCV uses to extract HOG features.

All tests compared each face against all other faces in the data set. Since the number of detected faces is not same in each video clip, the training data is unbalanced. This is rectified by applying the Synthetic

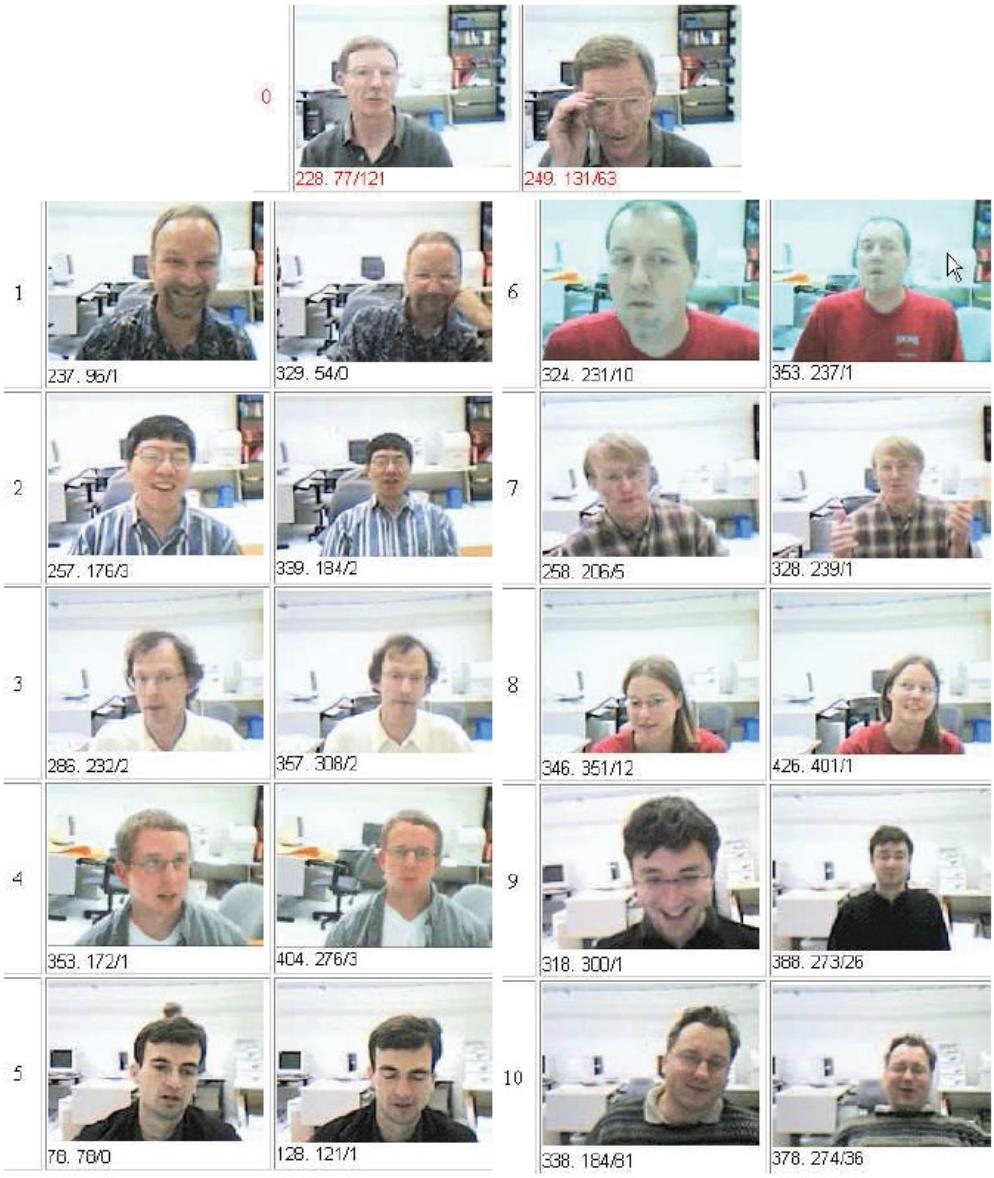

Figure 2: Video clips in the NRC-FRiV data set (Figure reproduced from [7]). The numbers underneath the images indicate the number of frames in a clip (the first number) and the number of those of them where a face detected (the second number).

Minority Oversampling Technique (SMOTE) algorithm, which generates new instances for the smallest class in the data set. Particularly, SMOTE is used to execute over-sampling of the minority class by creating

artificial data with similar distance in the feature space [4]. The algorithm uses the K-nearest neighbors for each example, and the distance is calculated according to the smallest distance along the n-dimensional feature space.

To build face models from features, the Random Forest classification algorithm is used, implemented using Weka. The Random Forest algorithm operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees [3].

The performance of features is measured using *Accuracy*, which also commonly referred to as *Recall*. The accuracy of a classifier on a given test set is defined as the percentage of test set tuples that are correctly classified by the classifier.

The *Variance* in relation to the accuracy of the Random Forest was computed by using the 10-fold cross-validation. In 10-fold cross-validation, the data set is broken in 10 exclusive sets or "folds' [12]. Training and testing are performed 10 times, which is why it is called 10x10-fold cross-validation. In each iteration, one of the data partitions is used for testing and the rest for training. For classification, the accuracy estimate is the overall number of correct classifications from the 10 iterations, divided by the total number of tuples in the initial data [12].

In the next section, the performance of each facial feature is reported in terms of the Accuracy (Recall) and Variation metrics computed for each of eleven target individuals in the NRC-FRiV dataset using the ML techniques described above.

Dataset	Recall	Variance
HARRIS_01_SMOTE	83.55	2.34
HARRIS_02_SMOTE	91.52	1.31
HARRIS_03_SMOTE	81.47	2.08
HARRIS_04_SMOTE	88.03	1.79
HARRIS_05_SMOTE	85.52	2.15
HARRIS_06_SMOTE	85.84	1.94
HARRIS_07_SMOTE	82.82	1.82
HARRIS_08_SMOTE	86.39	1.61
HARRIS_09_SMOTE	88.07	1.98
HARRIS_10_SMOTE	81.47	1.95
HARRIS_11_SMOTE	85.32	1.91
Average	85.45	

Table 1: Face recognition results for Harris features for each of eleven identities in the NRC-FRiV dataset.

3 Comparative overview of facial features for face recognition

3.1 Harris

As explained in [14], "Harris features look at the average directional intensity change in a small window around a putative interest point. This average intensity change can then be computed in all possible directions which leads to the definition of a corner as a point for which the average change is high in more than one direction. From this definition, the Harris test is performed as follows. We first obtain the direction of maximal average intensity change. Next, check if the average intensity change in the orthogonal direction is also high. If it is the case, then we have a corner". Results of simulation with Random Forest with Harris features are presented in Table 1.

Dataset	Recall	Variance
FAST_10_SMOTE	84.92	2.14
FAST_11_SMOTE	89.80	2.00
FAST_1_SMOTE	88.63	1.89
FAST_2_SMOTE	93.23	1.25
FAST_3_SMOTE	88.54	1.84
FAST_4_SMOTE	90.07	1.34
FAST_5_SMOTE	89.93	1.65
FAST_6_SMOTE	89.49	1.87
FAST_7_SMOTE	87.53	1.67
FAST_8_SMOTE	89.82	1.78
FAST_9_SMOTE	89.89	1.80
Average	89.26	

Table 2: Face recognition results for FAST features

3.2 FAST Features

Paper [14] describes the FAST (Features from Accelerated Segment Test) descriptor as follows: "(The) definition is based on the image intensity around a putative feature point. The decision to accept a keypoint is done by examining a circle of pixels centered at a candidate point. If an arc of contiguous points of length greater than 3/4 of the circle perimeter is found in which all pixels significantly differ from the intensity of the center point, then a keypoint is declared". Table 2 presents the simulation results of Random Forest with FAST features.

Dataset	Recall	Variance
GFTT_01_SMOTE	84.51	2.11
GFTT_02_SMOTE	92.05	1.14
GFTT_03_SMOTE	82.05	2.03
GFTT_04_SMOTE	86.63	2.06
GFTT_05_SMOTE	85.98	1.65
GFTT_06_SMOTE	86.74	1.69
GFTT_07_SMOTE	82.58	1.92
GFTT_08_SMOTE	86.01	1.59
GFTT_09_SMOTE	87.88	1.54
GFTT_10_SMOTE	81.01	1.74
GFTT_11_SMOTE	85.09	2.08
Average	85.50	

Table 3: Face recognition results for GFTT features

3.3 GFTT Features

As presented in [6], "Shi's and Tomasis Good Features To Track (GFTT) is a feature detector that is based on the Harris corner detector. The main improvement is that it finds corners that are good to track under affine image transformations". Table 3 presents the simulation results of Random Forest with GFTT features.

Dataset	Recall	Variance
MSER_01_SMOTE	89.76	6.71
MSER_02_SMOTE	98.51	2.67
MSER_03_SMOTE	88.28	7.05
MSER_05_SMOTE	82.07	6.85
MSER_06_SMOTE	92.04	6.27
MSER_07_SMOTE	84.58	6.65
MSER_08_SMOTE	93.70	5.68
MSER_10_SMOTE	87.90	6.46
Average	89.60	

Table 4: Face recognition results for MSER features

3.4 MSER Features

Paper [16] gives an informal explanation of MSER (Maximally Stable Extremal Regions) as follows: "Imagine all possible thresholdings of a gray-level image I. We will refer to the pixels below a thresh-

old as 'black' and to those above or equal as 'white'. If we were shown a movie of thresholded images I_t, with frame t corresponding to threshold t, we would see first a white image. Subsequently black spots corresponding to local intensity minima will appear and grow. At some point regions corresponding to two local minima will merge. Finally, the last image will be black. The set of all connected components of all frames of the movie is the set of all maximal regions; minimal regions could be obtained by inverting the intensity of I and running the same process". Table 4 presents the performance of Random Forest with MSER features.

3.5 HOG Features

[22] presents a brief explanation about the HOG features, as follows: "(...) Each detection window is divided into cells of size 8 x 8 pixels and each group of 2 x 2 cells is integrated into a block in a sliding fashion, so blocks overlap with each other. Each cell consists of a 9-bin Histogram of Oriented Gradients (HoG) and each block contains a concatenated vector of all its cells. Each block is thus represented by a 36-D feature vector that is normalized to an L2 unit length. Each 64x128 detection window is represented by 7x15 blocks, giving a total of 3780 features per detection window". Table 5 presents the recognition accuracy results for HOG features.

Dataset	Recall	Variance
HOG_60_full_10_SMOTE	91.27	1.61
HOG_60_full_1_SMOTE	92.06	1.63
HOG_60_full_2_SMOTE	96.05	1.18
HOG_60_full_3_SMOTE	92.99	1.38
HOG_60_full_4_SMOTE	89.91	1.85
HOG_60_full_5_SMOTE	93.74	1.40
HOG_60_full_6_SMOTE	94.25	1.35
HOG_60_full_7_SMOTE	91.56	1.92
HOG_60_full_8_SMOTE	93.30	1.63
HOG_60_full_9_SMOTE	92.04	1.73
Average	92.72	

Table 5: Face recognition results for HOG features.

3.6 Performance comparison results

The presented results are related to the selection of local facial features for face recognition in video-based applications. Evaluated features are Harris, FAST, GFTT, MSER and HOG.

Harris, FAST, GFTT and MSER features have shown similar performance, but in the case of MSER features there is an extra problem: this feature demands the size of the images to be bigger than the original

size. In addition, MSER features generated the smallest amount of instances than other features, which makes this feature not suitable for working with low quality images. GFTT and Harris features have very close performance, because GFTT is derived from Harris. HOG features have shown the best recognition results among tested features in terms of its recognition accuracy on a simple video data-set such as NRC-FRiV with a simple ML algorithm such as Random Forest.

This research indicates that HOG features appear to offer a reasonably good solution combined with a simple ML algorithm such as Random Forest. Other ML algorithms that could be tested to possibly further improve the recognition performance include new types of Decision Trees such as Very Fast Decision Trees (VFDT) [5]. These algorithms are design to build models from data-streams. VFDT has also the capability to learn models very fast, which can be useful to learn new faces dynamically.

The main issue with the presented evaluation is related to the assumption that the recognition problem is reduced to a binary classification problem. Real life scenarios demand the use of a database with many faces. With the current assumption, there will be a requirement to train one classifier for each target subject, which can be time consuming and demand more memory resources.

In the next section, the HOG features, which have been found to be the best performing using a simple NRC-FRiV data-set with Random Forest ML algorithm, are applied to a data set of a large scale with other ML algorithms.

4 Feature evaluation on a large-scale dataset with other ML algorithms

An implementation of the HOG features was done to evaluate their performance on a larger scale problem with real life scenario. The ORL face database was used [18, 8] It consists of 400 still images, 10 images per person for each of 40 enrolled persons, each captured with from different points of view and/or with different face expressions. The size of each image is 92 x 112 at 8-bit grey levels.

The implementation was done using OpenCV version 2.4.3, which has a class that encapsulates all functionalities of a face recognition process, called `FaceRecognition`. All new face recognition algorithms need to inherit their functionalies from this class.

The first step was to implement a C++ class, which was called `HOG`, and plug it in an application that could read the images and pass the data to this class. Some of these faces from the dataset are shown in Figure 3.

The `HOG` class depends on ML algorithms that are used to train and predict the data. In the previous section the ML algorithms were implemented in Weka. One of the objectives of the experiments was to test if the ML algorithms implemented by OpenCV would have any influence on the recognition results.

The main problem is that OpenCV's algorithms have some important limitations. For example, Boosting and SVM algorithms only deal with binary classification problems, which make it diffucult to use these algorithms with multiple subjects in the database. Despite the fact that initial simulations presented in previous section were done with binary configuration, the large-scale evaluation was conducted considering that a single database with all subjects images was created and the ML algorithms should create a final model for the whole data set. This is a important change in configuration, because in real life situations,

Figure 3: Facial images from the ORL database.

there will be a database with all faces that the algorithm must decide on. Due to these limitations, it was decided to use the algorithms implemented in Weka.

The Java Native Interface (JNI) had to be used to allow C++ classes to access Java classes, since Weka is implemented in Java. The use of Weka library requries more time to train the ML algorithms, because it reuires that all algorithms are retrained every time the program is restarted. In contrast, OpenCV's algorithms can save the model, and reuse them when the program is restarted. Another difference is related to the number of instances that can be used to train. Because Weka requires the Java Virtual Machine (JVM) to be started, it requires more memory to process all the information.

Tests were performed by selecting one face of the ten faces for each subject to be used as test data. This process emulates the 10-fold cross-validation used in previous session. This test was repeated twice for each subject in the database: the first time the *last* image on the list was used for testing, and the second time the *first* image was selected to test. This test procedure was also used on the algorithms originally implemented in OpenCV, such as: Fisherfaces [2], Eigenfaces [19] and Local Binary Patterns Histograms (LBPH) [1]. Images were not pre-processed and faces were left as is: neither localized, nor aligned. This was necessary, because the algorithms, which were used to detect faces (such as those implementing Haar cascades [20]) missed various faces of ORL database, making it impossible to evaluate certain faces.

Figure 4 presents the accuracy comparison of the algorithms. Accuracy is the measure of how many times a ML algorithm correctly classifies each image. As mentioned, the tests were done by removing the first and last face for each subject and making the algorithms learn from the remaining pictures. In total, for each training phase the algorithms were presented with nine pictures and one was used for testing.

Fisherfaces, Eigenfaces and LBPH are based on distance metric, where they calculate a distance between the faces in the database and the new face presented for testing. HOG features use the Bayesian

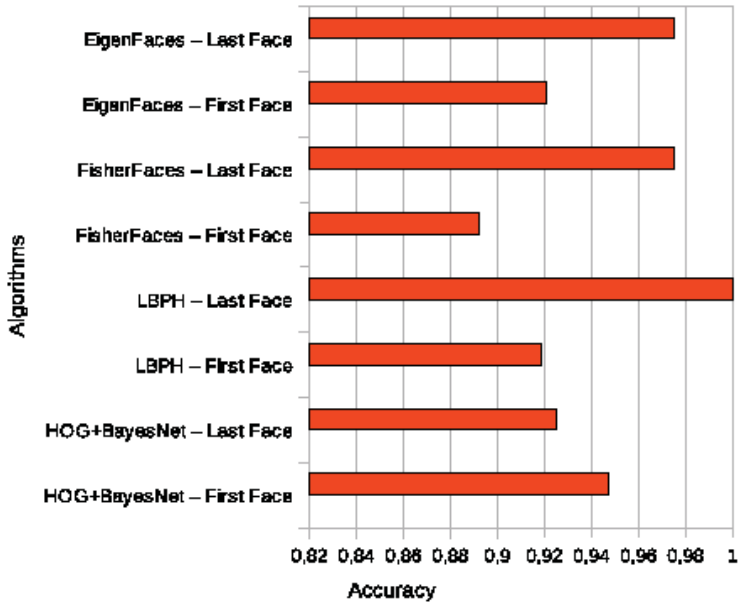

Figure 4: Graph presents the performance comparison among different algorithms implemented in OpenCV (Fisherfaces, Eigenfaces and LBPH)- all use distance metrics to recognize a face. HOG features are the only algorithm implemented that uses a ML algorithm (Bayesian Network) to perform face recognition.

Network ML algorithm to learn and predict. Distance metrics algorithms showed a variation during testing, especially when the tests removed the first faces of the subjects. These faces were usually frontal pictures. HOG features with Bayesian Network did not suffer from this problem and kept almost the same performance in both cases. The best algorithm was LBPH, which recognized all faces in the case of the last face in the list, but reduced its performance when it was requested to recognize the first face (92% accuracy). HOG features with Bayesian Network had better results than all algorithms when classifying the first face (94%), but it had worse performance than all algorithms to classify the last image (92%).

Another important observation is related to the use of Bayesian Network as an ML algorithm, instead of the Random Forest used in previous topic. The main reason is that the database is composed with data from *all* faces, and Random Forest was not able to perform well with this configuration.

5 Conclusions

This work presented an evaluation of facial feature extraction algorithms for face recognition on video using several a traditional machine learning algorithms implemented in OpenCV. The evaluated feature extraction algorithms included Harris, FAST, GFTT, MSER and HOG. Among those, HOG showed the best perfor-

mance on a small scale dataset and was chosen for further testing on a larger scale dataset using different ML algorithms. The evaluation was executed with cross-validation, because its theoretical background ensures that the results are representative of what independent test sets would yield [21].

The obtained results showed that open source face recognition codes, such as those available in OpenCV library, can be sufficient for building FRiV systems that work in the Type 1 video surveillance scenarios (i.e. person at the kiosk), provided that a good quality face picture is captured.

As future work, the techniques for automated face alignment, e.g. such as presented in report [17], should be investigated for further improvement of the face recognition performance in video. Additionally, techniques for pre-processing of images captured in poor lighting should be examined, for example such as those described in [15], which appear promising for video surveillance applications

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *ECCV (1)*, ser. Lecture Notes in Computer Science, T. Pajdla and J. Matas, Eds., vol. 3021. Springer, 2004, pp. 469–481.

[2] P. N. Belhumeur, J. a. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997. [Online]. Available: http://dx.doi.org/10.1109/34.598228

[3] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [Online]. Available: http://dx.doi.org/10.1023/A:1010933404324

[4] E. N. de Souza, "Extending adaboost:varying the base learners and modifying the weight calculation," Ph.D. dissertation, University of Ottawa, May 2014.

[5] P. Domingos and G. Hulten, "Mining high-speed data streams," in *KDD*, 2000, pp. 71–80.

[6] M. Eckmann and T. E. Boult, "Spatio-temporal consistency and distributivity as qualities of features," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 1, no. 1, pp. 1 –8, 2008.

[7] D. Gorodnichy, "Face databases and evaluation," *Chapter in Encyclopedia of Biometrics (Editor Stan Z. Li), Springer*, 2009.

[8] D. O. Gorodnichy, "Video-based framework for face recognition in video," *Proc. of Second Canadian Conference on Computer and Robot Vision (CRV'05), Workshop on Face Processing in Video*, vol. 1, no. 1, pp. 330 – 338, May 2005. [Online]. Available: http://www.videorecognition.com/FRiV

[9] D. Gorodnichy, E. Granger, and P.Radtke, "Survey of commercial technologies for face recognition in video," CBSA, Border Technology Division, Tech. Rep. 2014-22 (TR), 2014.

[10] E. Granger, P. Radtke, and D. Gorodnichy, "Survey of academic research and prototypes for face recognition in video," CBSA, Border Technology Division, Tech. Rep. 2014-25 (TR), 2014.

[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[12] J. Han, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.

[13] Intel, "Intel open source computer vision library," October 2008.

[14] R. Laganière, *OpenCV 2 Computer Vision Application Programming Cookbook*, 1st ed. Birmingham - Mumbai: Packt Publishing, 2011.

[15] H.-S. Le and H. Li, "Fused logarithmic transform for contrast enhancement," *Electronics Letters*, vol. 44, no. 1, pp. 60 – 61, January 2008.

[16] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *Proc. of British Machine Vision Conference*, vol. 1, no. 1, pp. 384–396, 2002.

[17] E. Neves, S. Matwin, D. Gorodnichy, and E. Granger, "3d face generation tool candide for better face matching in surveillance video," CBSA, Tech. Rep. 2014-11 (TR), 2014.

[18] F. S. Samaria and A. C. Hartert, "Parameterisation of a stochastic model for human face identification," in *Workshop on Applications of Computer Vision*, 1994.

[19] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991. [Online]. Available: http://dx.doi.org/10.1162/jocn.1991.3.1.71

[20] P. A. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR (1)*. IEEE Computer Society, 2001, pp. 511–518.

[21] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., ser. The Morgan Kaufmann Series in Data Management Systems, J. Gray, Ed. San Francisco, CA: Morgan Kaufmann Publishers, 2005.

[22] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, no. 1, pp. 1491–1498, 2006.