ARL-TR-8619 ● JAN 2019

ARL
US Army Research Laboratory

A Facial Recognition Algorithm Comparison: Using a Hybrid EigenFace ARTMAP Neural Network vs. the Tracking-Learning-Detection (TLD) Algorithm

by Troy D Kelley, Sean McGhee, and Eric Avery

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer’s or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-8619 ● JAN 2019

US Army Research Laboratory

A Facial Recognition Algorithm Comparison: Using a Hybrid EigenFace ARTMAP Neural Network vs. the Tracking-Learning-Detection (TLD) Algorithm

by Troy D Kelley and Eric Avery
Human Research and Engineering Directorate, ARL

Sean McGhee
STG, Inc., Reston, VA

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* January 2019	**2. REPORT TYPE** Technical Report	**3. DATES COVERED (From - To)** 1 January 2014–1 February 2019

4. TITLE AND SUBTITLE A Facial Recognition Algorithm Comparison: Using a Hybrid EigenFace ARTMAP Neural Network vs. the Tracking-Learning-Detection (TLD) Algorithm	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Troy D Kelley, Sean McGhee, and Eric Avery	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory Human Research and Engineering Directorate (RDRL-HRS-E) Aberdeen Proving Ground, MD 21005	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-8619
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report describes a comparison of two facial recognition processes for continuous learning. One process used an ARTMAP neural network with features extracted using a modified EigenFace implementation. This was compared with training the Tracking-Learning-Detection (TLD) algorithm using faces from a television episode. Results indicated that the TLD algorithm was superior to the Hybrid EigenFace/ARTMAP (EA) for the entire episode but that the Hybrid EA algorithm was better for the second half of the episode. The ARTMAP was chosen because it can adaptively train to new vectors without suffering from catastrophic forgetting. However, the TLD algorithm was capable of better online learning and overall performance.

15. SUBJECT TERMS

computer vision, facial recognition, tracking, robotics, neural networks

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Troy D Kelley
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	17	**19b. TELEPHONE NUMBER (Include area code)** 410-278-5869

Contents

List of Tables

Acknowledgments

Portions of the research in this report use the FERET database of facial images collected under the FERET program, sponsored by the Department of Defense Counterdrug Technology Development Program Office. We acknowledge Zdenek Kalal for his help with the code for this experiment.

1. Introduction

The Human Research and Engineering Directorate of the US Army Research Laboratory is developing a robotics control system called the Symbolic and Sub-symbolic Robotics Intelligence Control System (SS-RICS) (Kelley 2006). The system is based on the Adaptive Character of Thought-Rational cognitive architecture, which was designed to mimic the human thought processes (Anderson and Lebiere 1998). SS-RICS incorporates cognitive modelling algorithms from traditional cognitive architectures (production utility, spreading activation) with traditional artificial intelligence techniques (Principal Component Analysis [PCA]) for robotics processing. This report describes a facial recognition algorithm comparison as part of the ongoing development of SS-RICS.

The ability to recognize a face is fundamental to the cognitive abilities of humans and many other mammals, particularly primates. Human beings appear to have a robust facial-recognition system that uses a variety of techniques for identifying faces (Young and Ellis 1989). The system appears to be localized in the brain in the fusiform gyrus area (George et al. 1999). However, for a computer, tracking and recognizing a face can be difficult. Almost as soon as the instance of a face is presented for an algorithm to learn, the instance can change in the next frame of the image due to lighting, occlusion, or changes in perspective. This continuously changing representation presents a problem for object-tracking and facial-recognition algorithms. Humans overcome these challenges by using pupil dilation to compensate for lighting changes and background knowledge of likely object shapes and sizes to estimate the features of occluded objects.

Because SS-RICS is being developed with an emphasis on human interaction, it is then only natural that it include the ability to detect people's faces, track them, and re-recognize them at a later time. This is useful if the robot is meant to be operated by a select few individuals. These people's faces can be learned by the robot and serve as a gating mechanism to prevent unauthorized use.

2. Algorithm Comparison

Face detection and recognition can be accomplished in a few different ways. One way is to use a Haar cascade feature detector (Viola and Jones 2001) for the initial face detection and EigenFace PCA for identification of a specific face (Kirby and Sirovich 1990). The Haar features are actually Haar wavelets that have been crafted to detect certain "face-like" areas in a region of interest. For example, the features of a face (the eyes being darker than the cheeks) are used to determine if a face is present in the region of interest. The collection of faces is then transformed into an

Eigenspace from which the PCA vectors are obtained for processing a candidate face. Note that this vector space needs to be calculated each time a new face is added. This is useful enough for a limited set of faces; however, for the continuous learning of new faces in real time, a different approach is needed. This led us to develop a hybrid EigenFace ARTMAP approach.

2.1 Hybrid EigenFace ARTMAP (EA) Algorithm

For the hybrid EA approach we used a three-element methodology that included the addition of an Adaptive Resonance Theory (ART) neural network (Carpenter and Grossberg 2003). This network allowed the learning of individual faces. An ART neural network allows for continuous learning and does not suffer from catastrophic forgetting (French 1999) as is the case with some other more traditional neural networks. The ARTMAP neural network is an improved version of the original ART neural network, in which a superior error correction methodology for the learning of weights is used (Carpenter et al. 1991). The hybrid EA algorithm used the following three elements:

1) Haar Cascade feature detectors trained to identify frontal faces.

2) EigenFace facial discrimination code, which extracts salient features from hundreds of random sample faces into an average face for comparison with the features extracted from an unknown face identified by the Haar classifiers.

3) ARTMAP neural network, which takes the data from the EigenFace recognition process and uses it to learn individual facial signatures. This step allows for the continuous learning of faces.

Element 2 uses a variety of sample faces across a spectrum of types and instead of including known candidates for identification (something that would require periodic updating and regeneration of the Eigenspace). The resulting Eigenspace is averaged to compute a mean value. Then the PCA values of the target face are compared using a Mahalonobis distance computed from the difference between the PCA values of the target and the mean of the Eigenspace. This was done to obtain a vector to be fed into the ARTMAP (Element 3). This approach allows any new face to be learned without recomputing the Eigenspace.

2.2 Tracking-Learning-Detection (TLD) Algorithm

The TLD algorithm developed by Kalal et al. (2010) is a general-purpose tracker that can be used to follow many different kinds of targets. The TLD algorithm is a combination of three processes, as its name implies. First, the algorithm will track an object in the current region of interest (ROI). Next, the algorithm will learn the object as the object changes from occlusion and viewpoint specific lighting. Finally, the algorithm will detect the object once it is shown again as a brand new instance.

The original TLD algorithm has since been augmented by the original authors to include face detection and tracking. In order for us to allow the original TLD algorithm to track and identify faces, we added a traditional Haar classifier to find an ROI where a candidate face might be located. This candidate was then passed to the TLD tracker to follow and identify.

The TLD uses a short-term tracker based on the Lucas-Kanade (LK) method (Lucas and Kanade 1981). The object is tracked by the LK method, and patches close to the trajectory are used to update the detector with positive exemplars. The object is detected using a randomized forest detector using a feature set developed by the authors based on local binary features called 2bitBP features. The algorithm uses errors from positive and negative examples to partially cancel each other out and allow for a convergence on the best candidate. Initially, the target is tracked using a matrix of points garnered from within the user-specified bounding box, which is then passed to the LK tracker. As the object is tracked, patches along the trajectory are used to update the detector using the positive–negative learning strategy, which reduces them to 15 × 15 gray-scaled normalized snippets. These snippets are then added to the list of positive examples. The algorithm also uses a random sampling of surrounding images (a grid of bounding boxes that covers the entire image is generated during initialization) and uses them as negative examples. These positive and negative examples have their eigenvalues extracted and are projected into a space wherein nearest neighbors can be determined. Tracking of the object using LK is continued until the tracker loses the object due to a lack of forward–backward consistency. To improve learning, the TLD algorithm uses a new sampling strategy Kalal et al. define as "quasi-random weighted sampling with trimming (QWS+)". This approach "minimizes the error estimate" leading to improved results over more-traditional sampling methodologies (i.e., bootstrapping; Kalal et al. 2009).

3. Procedure

The evaluation was conducted using images from the sitcom *The IT Crowd* (first season, first episode, as was reported in Kalal et al. 2010). From the episode we extracted the frames (at 424 × 283 resolution) containing a specific character ("Roy"), which amounted to 10,982 frames. We then split the frames into two sections (as was done in Kalal et al. 2010), part one being 5,648 frames and part two being 5,334. The 10,982 frames of the episode lasted for approximately 23 min, which we used for testing both approaches. However, the Hybrid EA algorithm used 9,252 during processing due to dropped frames.

We trained the hybrid EA algorithm with 300 faces from the FERET face database (Phillips et al. 1998). The hybrid EA data set was trained initially using carefully sculpted subject images, all sampled at the same resolution and normalized. Additional care was taken to make sure the main facial features were all lined up in the same position in each image. This created a "space" for comparing candidate faces as they were detected by the Haar classifiers. Training files had a resolution of 120 × 120 pixels.

Once a face was detected and the eigenvalues representing its difference from the mean of the training set were obtained, they were truncated to the first 10 PCA elements and fed to the ARTMAP to learn as a particular face.

Additionally, we used a method of filtering out false positive faces from the Haar classifier. The candidate faces were compared using a cross correlation algorithm. Scores above a threshold were accepted and those below were considered false positives. We used a cross correlation of 0.47 to reduce the false positives (a value we found to be consistently a good balance between rejecting bad detections while preserving good ones). This setting represented a correlation value threshold that was compared with the score that an average face image (generated by the EigenFace training process) and a candidate face must exceed. The positive candidates were then passed on to the classification function.

At the beginning of the segment, the Haar classifier detected the face of a specific character (Roy), and was trained using the ARTMAP as a classifier. For each face in each frame detected, if the patch was classified as Roy, it was written to a directory named Roy; if the detection said "Unknown", the patch was saved to a separate directory. If no face was detected, that entire scene was saved to another directory. The contents of each directory were examined visually and a count of how many hits were accurate for both Roy and Unknown was tallied. Total accuracy of Roy classifications was the number of times the character was successfully identified divided by the 9,252 frames containing Roy.

For the TLD algorithm, a bounding box was drawn around the character Roy's face using the first frame of the 10,982 available frames. As Roy was detected, the patch identified by the tracker was saved to disk and the total number of accurate classifications was determined by examining them visually and counting the number of correct classifications. Thus, total accuracy of Roy detections was how many times the character was successfully classified divided by the total number of frames.

4. Results

We found similar results to those reported by Kalal et al. (2010) for the original TLD algorithm. The character (Roy) was tracked throughout the entire episode for a total of 10,982 frames for the TLD algorithm. Of those frames, the TLD algorithm was able to correctly identify the subject 31.16% of the time, with a false positive rate of 5.1%. This is in comparison with the subject being correctly tracked 37% of the time as reported by Kalal et al. (2010). We also found, as mentioned by the authors, that the TLD algorithm functioned better in the first half of the frames used than in the second half of the show (48.99% and 12.28%, respectively). This could be because during the first half of the show there was more character development and more close-ups of the actors' faces, while in the last half of the show there were more characters together in each frame, and the camera locations were further away showing groups of actors together.

By comparison, our Hybrid EA algorithm performed somewhat worse over the entire set of frames used but better than the TLD in the second half of the frames used. As Table 1 shows, the Hybrid EA algorithm correctly found and tracked the subject 18.3% of the time for the entire frames used. In the second half of the frames used, the Hybrid found the character 14.87% of the time compared with 12.28% for the TLD.

Table 1 Hybrid EA algorithm compared with TLD algorithm for detecting one character for the first and second half of the episode, plus the overall results

Algorithm	Percent	Correct	Detection
	First half	Second half	Overall
Hybrid	20.95%	14.87%	18.30%
TLD	48.99%	12.28%	31.16%

Approved for public release; distribution is unlimited.

5. Conclusions

Our research confirmed the accuracy of the original TLD algorithm using the same data set (a sitcom episode) for comparison. We also found that the TLD algorithm performed better than our Hybrid EA algorithm for the entire episode, but the Hybrid EA algorithm did slightly better than the TLD during the second half of the episode. We are not exactly sure why there were such differences between the first and second halves of the episode. One speculation would be that there might have been differences in plot and character development in the episode that made the first half somehow different from the second. The Hybrid EA, while performing at a lower rate, was more consistent across both halves of the video. One possible problem with the Hybrid EA algorithm was that the vectors returned from the EigenFace code were not unique enough to present the ARTMAP with disparate enough data to keep the vector delineated for accurate classification. The Hybrid EA also does not include any tracking per say, as it only uses the Haar classifier to acquire a possible face prior to classification. So while it is acquiring faces it is not tracking them in the same way that the TLD algorithm is tracking an instance of a specific ROI.

6. References

Anderson JR, Lebiere C. The atomic components of thought: Mahwah (NJ): Erlbaum Associates, Inc.; 1998.

Carpenter GA, Grossberg S. Adaptive resonance theory. In: Arbib MA, editor. The handbook of brain theory and neural networks. Second edition. Cambridge (MA): MIT Press; 2003. p. 87–90.

Carpenter GA, Grossberg S, Reynolds JH. ARTMAP: supervised real-time learning and classification of non-stationary data by a self-organizing neural network. Neural Networks. 1991;4:565–588.

French RM. Catastrophic forgetting in connectionist networks. Trends in Cognitive Sciences. 1999;3(4).

George N, Dolan RJ, Fink GR, Baylis GC, Russell C, Driver J. Contrast polarity and face recognition in the human fusiform gyrus. Nature Neuroscience. 1999;2:574–580.

Kalal Z, Matas J, Mikolajczyk K. Online learning of robust object detectors during unstable tracking. Presented at the 12th IEEE International Conference on Computer Vision Workshops, ICCV Workshops; 2009 Sep 20–Oct 2.

Kalal Z, Mikolajczyk K, Matas J. Face-TLD: tracking-learning-detection applied to faces. Proceedings of the 17th IEEE International Conference on Image Processing (ICIP); 2010 Sep 12–15; Hong Kong. p. 3789–3792.

Kelley TD. Developing a psychologically inspired cognitive architecture for robotic control: the symbolic and sub-symbolic robotic intelligence control system (SS-RICS). International Journal of Advanced Robotic Systems. 2006;3(3):219–222.

Kirby M, Sirovich L. Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Transactions on Pattern analysis and Machine Intelligence. 1990;12(1):103–108.

Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision. International Joint Conferences on Artificial Intelligence. 1981;81:674–679.

Phillips PJ, Wechsler H, Huang J, Rauss P. The FERET database and evaluation procedure for face recognition algorithms. Journal of Image and Vision Computing. 1998;16(5):295–306.

Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Vision and Pattern Recognition Conference; 2001 Dec 8–14.

Young AW, Ellis HD, editors. Handbook of research on face processing. Amsterdam (Netherlands): North-Holland; 1989.

List of Symbols, Abbreviations, and Acronyms

ART	Adaptive Resonance Theory
LK	Lucas-Kanade
PCA	Principal Component Analysis
ROI	region of interest
SS-RICS	Symbolic and Sub-symbolic Robotics Intelligence Control System
TLD	Tracking-Learning-Detection

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIR ARL
(PDF) IMAL HRA
RECORDS MGMT
RDRL DCL
TECH LIB

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 AFC
(PDF) DAMR ZA
GEN J M MURRAY

1 RDECOM
(PDF) AMSRD CG
MG C T WINS

1 ARL
(PDF) RDRL HRB B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

7 ARL
(PDF) SFC PAUL RAY SMITH CTR
RDRL HRA M CLARKE
RDRL HRA I MARTINEZ
RDRL HRA C A RODRIGUEZ
RDRL HRA B J HART
RDRL HRA A C METEVIER
RDRL HRA D B PETTIT
12423 RESEARCH PARKWAY
ORLANDO FL 32826

1 USA ARMY G1
(PDF) DAPE HSI M SAMS
300 ARMY PENTAGON
RM 2C489
WASHINGTON DC 20310-0300

1 USAF 711 HPW
(PDF) 711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH
45433-7604

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

1 ARL
(PDF) RDRL HRB A
J EVERETTE
BLDG E2929
DESERT STORM DR
FORT BRAGG NC
28310-0001

ABERDEEN PROVING GROUND

14 ARL
(PDF) RDRL HR
J LANE
J CHEN
P FRANASZCZUK
K MCDOWELL
K OIE
RDRL HRB
J LOCKETT
RDRL HRB C
J GRYNOVICKI
RDRL HRB D
D HEADLEY
RDRL HRF A
A DECOSTANZA
RDRL HRF B
A EVANS
RDRL HRF C
J GASTON
RDRL HRF D
A MARATHE
RDRL HRS E
T KELLEY
RDRL WMP E
J DOE