

Semi-supervised haartraining of a fast&frugal open source zygomatic smile detector

A gift to OpenCV community

Daniel Devatman Hromada
Lutin Userlab
Ecole Pratique des Hautes Etudes

prof. Charles Tijus
Cognition Humaine et Artificielle (ChART)
Université Paris 8

Abstract—Five different versions OpenCV-positive XML haarcascades of zygomatic smile-detectors as well as five SMILEsamples from which these detectors were derived had been trained and are presented hereby as a new open source package. Samples have been extended in an incremental learning fashion, exploiting previously trained detector in order to add and label new elements of positive example set. After coupling with already known face detector, overall AUC performance ranges between 77%-90.5% when tested on JAFFE dataset and <1ms per frame speed is achieved when tested on webcam videos.

Keywords—zygomatic smile detector; cascade of haar feature classifiers; computer vision; semi-supervised machine learning

I. INTRODUCTION

Great amount of work is being done in the domain of facial expression (FE) recognition. Of particular interest is a FE being at the very base of mother-baby interaction [1], a FE interpreted unequivocally in all human cultures [2] - smile. Maybe because of these reasons, maybe because of some others, smile detection is already of certain interest for computer vision (CV) community – be it for camera's smile shutter [3] or in order to study robot2children interaction [4].

Nonetheless, a publicly available, i.e. open source, smile detector is missing. This is somewhat stunning, especially given the fact that “smile” can be conceived as a “blocky” object [5] upon which a machine learning technique based on training of cascades of boosted haar-feature classifiers [6] can be applied, and that the tools for performing such a training are already publicly available as part of an OpenCV[5] project. Verily, with exceptions of detectors described in [7][8] which have not been publicly released, we did not find any reference to haarcascade-based smile detector in the literature. We aim to address this issue by making publicly available the initial results of our attempts to construct sufficiently descriptive SMILing Multisource Incremental-Learning Extensible Sample (SMILEs) and five smile detectors (smileD) generated from this sample.

From more general perspective, our aim was to study whether one can use already generated classifiers in order to facilitate such a semi-supervised extension of initial sample that a more accurate classifier can be subsequently trained.

A. SMILE sample (SMILEs)

The aim of SMILEs project is to facilitate and

accelerate the construction of smile detectors to anyone willing to do so. Since it is the OpenCV library which dominates the computer vision community, SMILEs package is adapted upon the needs of OpenCV in a sense that it contains 1) negative examples directory 2) positive examples directory 3) negatives.idx - list of files in negative examples directory 4) positives.idx - list of files in positives with associated information containing the coordinates of region of interest (ROI), i.e. the coordinates of the region within which smile can be located.

SMILEs is considered “Multisource” because it originates as an amalgam of already existing datasets like LFW and Genki both of which are, themselves, collections of images downloaded from the Internet. Images from POFA [9] of Cohn-Kanade [10] datasets were not included into SMILEs since restricted access to these datasets is in contradiction with an open source approach¹ of SMILEs project.

B. Smile Detector (smileD)

SMILEs are “Incremental-Learning Extensible” in a sense that they allow us to train new versions of smile detectors which are subsequently applied upon new image datasets in order to facilitate (or even fully automatize) the labeling of new images, and hence extending an original SMILEs with new images. Simply stated, SMILEs allow us to train smileD which helps us to extend SMILEs etc.

Since training of haar cascades is an exhaustive threshold-finding process demanding not negligible amount of time and computational resources, 5 pregenerated OpenCV-compatible XML smileD haarcascades were trained by opencv-haartraining application and are included with SMILEs in our OpenSource SMILEsmileD package, so that anybody interested could implement our smile detector in copy&use fashion.

¹ Both SMILEs & SMILEd cascades are publicly available from <http://github.com/hromi/SMILEsmileD> as a GPL-licensed package. C++ source codes of select&crop application for easy manual sample creation and of a face-coupled video stream smile detector are included as well.

II. METHOD

C. Initial Training Datasets

SMILEs project in its current state unites 3 image sets :

- cropped Labeled Faces in the Wild (LFW) dataset - LFW dataset [11] contains more than 13000 images of faces collected from the web; its cropped version contains only 25x25pixel regions detected by OpenCV's frontal face detector. No information about the presence/absence of a smile within the image is given
- Genki4K dataset – Genki4K is a publicly available part of UCSD's Genki project [12] containing 4000 images downloaded from Internet. A text file indicating the presence/absence of the smile in a given image is included.
- Ad hoc Flickr dataset – We have used the search keyword “smile” in order to download more than 4200 additional pictures from image-sharing website flickr.com. More than 2600 of them contained at least one smiling face.

D. Construction of SMILEs datasets

We have created five different version of SMILEs. All these versions exploit the same negative sample set of LFW's nonsmiling images. All manual labeling focalised solely on zygomatic smile (ZS) region²:

- **Version 0.1** is based solely upon an LFW dataset. All pictures were manually labeled by our ad hoc region selection & cropping application and divided into samples of positive (3606 images) and negative (9474 images) examples.
- **Version 0.2** added 2666 manually labeled images downloaded from flickr.com to positive examples contained already in 0.1. Labeling & region selection was realised by same application as in case of 0.1.
- **Version 0.3** also extended the positive&negative example samples of version 0.1 with images from flickr. This time, however, the flickr-originated images weren't labeled manually, but the smile-containing regions of interest were determined automatically, by applying smileD of version 0.1 upon the set of downloaded images. 1372 ROIs (1 ROI for 1image) were identified&labeled in this way.

TABLE I. BASIC COMPONENTS OF INITIAL VERSIONS OF SMILEs&SMILED PROJECT

Version	Positive examples					Neg. ex.
	LFW manual	Flickr manual	Flickr auto	Genki auto	Total	
0.1	3606	0	0	0	3606	9474
0.2	3606	2666	0	0	6262	9474
0.3	3606	0	1372	0	4978	9474
0.4	3606	0	0	624	4230	9474
0.5	3606	0	1372	624	6572	9474

- **Version 0.4** is analogous to version 0.3 in that sense that

² ZS region was defined only vaguely as a rectangular ROI in whose center are smiling lips – in preference with uncovered teeth. Whole ROI is bordered by smile&nasolabial wrinkles.

it is essentially a version 0.1 sample to which automatically labeled positive examples were added. Differently from version 0.3, Genki4K and not flickr was exploited as a source of additional data. Simply stated, positive examples, 624 of them in total, from Genki4K labeled as smile-containing by its authors were added to initial LFW-based sample.

- **Version 0.5** unites the versions 0.3 and 0.4, i.e. both Genki4K & flickr-originated images which were automatically labeled by smileD v0.1 were added to LFW samples.

E. SMILEs -> smileD Training

Identical haarcascade training parameters [width=43, height=19, number of stages=16, stage hit rate=0.995, stage false alarm rate=0.5, weak classifier decision tree depth=1(i.e. Stump), weight trimming rate=0.95] were applied for training of all five smileD versions, one smileD corresponding to one SMILEs, both referenced by same version number.

F. smileD evaluation

Training phase of every new version of smileD was followed by measuring its performance upon a Japanese Female Facial Expression (JAFFE) dataset in order to evaluate the performance of different versions of smileD classifiers when applied upon a sample having different luminosity conditions than that any imageset included in train sample

Detectors were face-detector-coupled during testing, i.e. smile detection was performed iff a face was detected in a tested image, and only in the ROI defined by well-known geometric ratios [13]

Receiver operating characteristic (ROC) curves were plotted and AUC (“area under ROC curve”) were calculated as performance measures by means of ROC library [14]. “Smile intensity” [7], i.e. the number of overlapping neighboring hit regions³, was used as a cutoff parameter.

III. RESULTS

FIGURE I. SMILED ROC CURVES

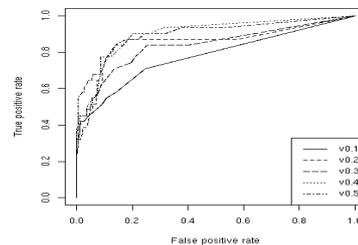


TABLE II. Roc's "AREA UNDER CURVE" PERFORMANCE OF DIFFERENT VERSIONS OF SMILED DETECTOR

Version	AUC
0.1	77.94%
0.2	85.49%
0.3	83.93%
0.4	90.21%
0.5	90.51%

DISCUSSION

Detectors we present hereby exploit the top-bottom approach, i.e. they are face-coupled. Knowing that *there can*

³ Can be obtained from undocumented *neighbors* attribute of *cvAvgComp* sequence referenced by *cvHaarDetectObjects*

be no smile without the face within which it is nested, we firstly detect the face by an OpenCV face detection solution and then smileD is applied only in very limited ROI of face's bottom third. Consequences of our decision to create face-coupled smile detector are twofold: 1) since by definition we search for smile only within the face, we have used only nonsmiling faces as negative examples (i.e. background images) 2) smile detection itself is very fast, once the position of face is specified. When applied upon the webcam-originated (320x240 resolution) video streams, the time needed for in-face smile detection in never exceeded 1ms per frame on a Mobile Intel(R) Pentium(R) 4 CPU (1.8GHz), suggesting that SmileD could be potentially embedded even into mobile devices disposing of less computational resources.

SmileD's speed can somehow neutralize its smaller accuracy handicap which it has in comparison with results reported in [8]. In its current state, our approach suffer from somewhat high false alarm rates, but our research indicates that in real life condition, these can be in great measure reduced by taking into account the dynamic sequence of subsequent frames since the probability of the same false alarm occuring within all the frames of the sequence is proportional to the product of probabilities of occurrence of that false alarm for every frame of the sequence taken individually. High speed is therefore of utmost importance and analysis of sequences of frames can substantially reduce the number of false positives.

Tuning of training parameters and the extension of negative example do remain as other possibilities how to augment the accuracy of our project. Tab.2 indicates that accuracy of such semi-supervised classifiers like smileD gets saturated at certain limit which can possibly be surmounted only by extension of negative sample set. In case of smile detection, we suggest that extension of negative example sample with more images containing "upper lip raiser" action unit (AU 10) – teeth-uncovering⁴ but associated with disgust rather than smile – could yield some significant increases in accuracy, as reported by [9]. Since such an extension is relatively easy and not much time-consuming, given that such AU10-containing images are given and marked as negative examples, it may be the subject of future research.

In this study, however, we left the negative example unchanged in order to study the effectivity of "Incremental Learning" approach during which an old detector is used to facilitate the extension of a positive example sample thanks to which a new detector is obtained. Since semi-supervised smileD versions v0.4 and v0.5 have outperformed v0.2 for which manual labeling was implemented, while the latter one performed only slightly better than v0.3 which exploited an identic flickr-originated imagebase than v0.2, it is not unreasonable to think that such semi-supervised incremental training approach can be a feasible solution for training haarcascade detectors. If that would be the case, it could possibly be stated that the machine started, in certain sense, to

ground [15] its own notion of smile.

ACKNOWLEDGMENT

We would like to thank the third section of EPHE, University Paris 8 and CROUS de Paris for their kind support.

REFERENCES

- [1] L. Strathearn, J. Li, P. Fonagy, et P.R. Montague, "What's in a smile? Maternal brain responses to infant facial cues," *Pediatrics*, vol. 122, 2008, p. 40.
- [2] C. Darwin, P. Ekman, et P. Prodger, *The expression of the emotions in man and animals*, Oxford University Press, USA, 2002.
- [3] M. Akita, K. Marukawa, et S. Tanaka, "Imaging apparatus and display control method," 2010.
- [4] J.R. Movellan, F. Tanaka, I.R. Fasel, C. Taylor, P. Ruvolo, et M. Eckhardt, "The RUBI project: a progress report," *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 2007, p. 339.
- [5] G. Bradski et A. Kaehler, *Learning OpenCV*, O'Reilly Media, Inc., 2008.
- [6] P. Viola et M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple," *Proc. IEEE CVPR 2001*.
- [7] O. Deniz, M. Castrillon, J. Lorenzo, L. Anton, et G. Bueno, "Smile Detection for User Interfaces," *Advances in Visual Computing*, p. 602–611.
- [8] J. Whitehill, M. Bartlett, G. Littlewort, I. Fasel, et J. Movellan, "Developing a practical smile detector," *Submitted to PAMI*, vol. 3, 2007, p. 5.
- [9] P. Ekman et W.V. Friesen, *Pictures of facial affect*, Palo Alto, CA: Consulting Psychologists Press, 1976.
- [10] T. Kanade, Y. Tian, et J.F. Cohn, "Comprehensive database for facial expression analysis," *fg*, 2000, p. 46.
- [11] G.B. Huang, M. Ramesh, T. Berg, et E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition" *University of Massachusetts, Amherst, Technical Report*, vol. 57, 2007, p. 07–49.
- [12] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, et J. Movellan, "Toward Practical Smile Detection," *IEEE transactions on pattern analysis and machine intelligence*, 2009, p. 2106–2111.
- [13] L. Da Vinci et J.P. Richter, *The notebooks of Leonardo da Vinci*, Dover Publications, 1970.
- [14] T. Sing, O. Sander, N. Beerenwinkel, et T. Lengauer, "ROCR: visualizing classifier performance in R," *Bioinformatics*, 2005.
- [15] S. Harnad, "The symbol grounding problem," *Physica d*, vol. 42, 1990, p. 335–346.

⁴ From anatomic point of view, disgust-expressing AU10 is associated with *Levator Labii Superioris* muscle while smile associates with *Zygomaticus Major* muscle (AU12).