



Estimating the Smile by Evaluating the Spread of Lips

Mohan Goud Kathi*, Jakeer Hussain Shaik

Vignan's Foundation for Science, Technology and Research, Vadlamudi 522213, Guntur, India

Corresponding Author Email: kathi.mohangoud@gmail.com

<https://doi.org/10.18280/ria.350207>

Received: 15 January 2021

Accepted: 10 April 2021

Keywords:

landmarks representation of lips with respect to eyes, smile diagnosis, face adjustment

ABSTRACT

Smile is one of the important emotions that is essential in computer vision tasks. The greatly influenced part due to it is the lips. By encountering the changes in lips of smile images with respect to no smile, a smile detecting model can design for the computer vision tasks. In this paper, the approach is to evaluate the spread of lips. The lips movement distance is evaluated with respect to the eyes. 68 landmark points of dlib are used for this purpose. The left and right corners of lips are evaluated with the left and right eyes respectively using the count of landmark points (l and r). The secondary parameters - average, Maximum, and maxavgsum of l and r are used for evaluating the lip expansion variation. For each value of these parameters that can attain from l and r, the count of no smile images below it and count of smile images above it is considered and calculated the attainable efficiency. The value of secondary parameter having the maximum efficiency is defined as the threshold. The maximum efficiency that is attained due to average, Maximum and maxavgsum are 80.06, 67.3 and 78.54 respectively at the thresholds 2, 3 and 4.5 respectively.

1. INTRODUCTION

Smile is one of the important emotions that is essential in computer vision tasks. It is a good gesture presenting happiness, approval friendliness etc., Ochs et al. [1] propose a method to generate virtual smiles that conveys politeness, embarrassment, amusement and a combination of these. Machines understand any image with pixels and can estimate the smile by evaluating these pixel differences [2]. In real-time scenarios, few constraints affect smile detection. Those can resolve with proper image registration, feature representation and trained using a huge number of images [3]. An efficient deep Convolution neural network can combine feature selection and classification and can efficiently involve in smile prediction [4]. Since it involves two phases – feature extraction and classification, the accuracy of smile prediction depends on the accuracy of both the phases [5]. It is important to investigate the CNN for different kinds of transformations and distortions to understand its effectiveness in the realworld [6]. By extracting useful features [7], before applying CNN can improve the efficiency of detection. Instead of implementing CNN from scratch, using robust deep learning models that are used for face recognition, can attain the best smile detection results. Those systems not only improve accuracy but also reduces training time [8].

Not only deep neural networks but also few other approaches that can estimate the smile. Smile influences the mouth shape, by extracting these features using HOG and LBP Histograms, the smile can detect [9]. Edge oriented histogram [10] is one more feature descriptor that can detect the smile from lips. Yang and Zhen [11] propose a method of smile detection for mobile platforms using Gabor transforms. Compared to Gabor features, the pyramid histogram of oriented gradients is having a shorter vector length and a high

smile recognition rate [12]. It is proved that the Fractional Fourier transform can also recognize smile better than Gabor features [13]. Instead of using one feature descriptor, using multiple descriptors, hierarchically can improve the smile detection accuracy [14]. By training the mouth and eye pair images using the KNN classifier, the smile on the face can detect [15]. The smile intensity can estimate by training the multiclass or regression models [16]. Akkoca and Gökmen [17] collected the images of smile and neutral and applied various approaches for smile detection. But this can't be used in real-time because it can able to separate the smiling image from neutral but not from other emotions. Tang and Huang [18] proposes a method to characterize the smiles, by drawing the motion vectors of feature points of facial parts like eyes, mouth and cheeks that are derived from neutral and smiling faces. While smiling, the corners of the lip slightly curve up. Using Expansion of lips, the smile can detect and the contraction of eyebrows can estimate the frown. Tsai et al., and Royce et al. [19, 20] proposes Harris corner detection and FAST corner detection methods for identifying the smile by evaluating the curve up of lip corners.

From Figure 1, one can observe the expansion of lips due to smile. Hence by extracting the lips and estimating the distance it moves, one can find the smile. The colour of lips is always different from philtrum and chin colours. It is easier to segment the lips if a moustache and beard are present [21]. But this kind of segmentation doesn't provide any information related to the spreading of lips. Haar cascade provides XML files for extraction of face parts like eyes, mouth, nose etc., But this kind of extraction also failed in providing expansion of lips, Facial landmarks can aid in providing information about all the face parts. This makes it useful in drowsiness detection [22], age estimation [23]. The expansion of lips can estimate using landmarks of the face.



Figure 1. Expansion of lips because of smile

2. DEFINING THE PARAMETERS FOR EVALUATING THE EXPANSION OF LIPS

This paper uses dlib's 68 landmark points which can provide the information about jaw (1to17), right eyebrow(18to22), left eyebrow (23to27), nose(28to36), right eye(37to42), left eye (43to48), and lips(49to68). The representation of landmarks for eyes and lips are illustrated in Figures 2 and 3. Since the left and right eyes never move sideways. The spread of lips can estimate with respect to eyes. In order to analyze this, two parameters (l and r) are defined. The total landmark points of the left eye are 6. The leftmost landmark point of the lip is compared with all the left eye landmarks and l is the number of landmark points of the left eye that crossed the left most landmark point of mouth and similarly, r is defined for the right side of the lip. Figure 4 illustrates the l and r values, the dots represent the landmark points of left and right eye for Figure 4, l=5 and r=4.



Figure 2. Left and Right eye presentation with landmarks



Figure 3. Lips presentation with landmarks

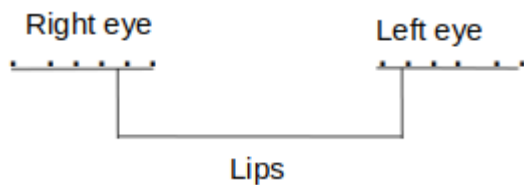


Figure 4. Representation of eye landmarks with respect to lips

3. CALCULATION OF ANGLE TO BE ROTATED FOR FACE ALIGNMENT

The face pose affects the l and r values. Hence, face alignment is necessary for obtaining reliable values. The blue

square (ABCD) in Figure 7 is the extracted face from the image, g is the right eye, l is the left eye and n is the nose. gln forms a triangle. In order to align the face, g and l must in a straight line and n is at the center. A median is drawn from n to gl and pointed as m and another median from n to AB is drawn and pointed as f. These two medians must be the same for the face alignment. Hence, $\angle P$ must be zero. In other words, the image needs to rotate $\angle P$ for face alignment. The points nmf form a triangle. With the n,m, and f points, the length of each side is calculated using the formula -. With sides, the angle of vertices can calculate using the cosine rule - $p^2 = q^2 + r^2 - 2qr\cos P$ and $\angle P = \cos^{-1}(\frac{q^2+r^2-p^2}{2qr})$.

Figures 5 and 6 illustrate the change in pose due to face alignment. It is found that the l and r values of Figure 6 are changed from 0 and 6 to 1 and 4 due to alignment.

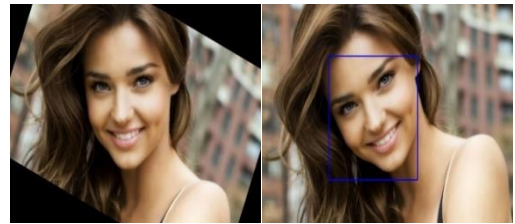


Figure 5. The change in image after rotating with $\angle P$



Figure 6. The change in face pose after rotating with $\angle P$

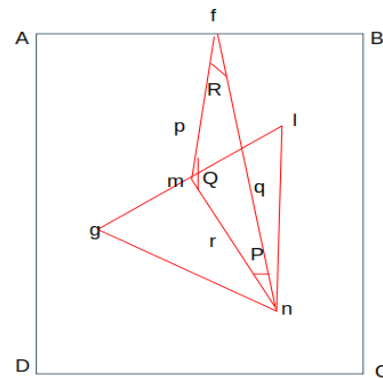


Figure 7. Face alignment angle measurement

4. INVESTIGATION OF CHANGES IN l AND r VALUES DUE TO FACE ALIGNMENT

In this paper, total 988 images among 494 smile and 494 no smile images are utilized and the threshold is defined. The amount of deviation of l and r values due to face adjustment is illustrated using Table 1 and Table 3 for no smile and smile respectively. The ll (l value after face adjustment - l value before adjustment) and rr (r value after adjustment - r value before adjustment) represents the corrected amount of

landmarks for misaligned images. As the images are more deviated, this difference is more. It is found that the most of the deviation is between -2 and 2 from Table 1 and Table 3. Table 5 illustrates the changes in terms of count. Table 2 and Table 4 illustrates the total number of images observed for all the combinations of l and r.

Table 1. Investigation of changes in l and r values for the no smile images

	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6
ll	1	1	3	9	66	89	28	19	2	1	0	0
rr	0	0	1	4	30	19	69	95	10	2	0	0

Table 2. Adjustments in l and r values due to alignment for the no smile images

	r=0		r=1		r=2		r=3		r=4		r=5		r=6	
	BA	AA	BA	AA	BA	AA	BA	AA	BA	AA	BA	AA	BA	AA
l=0	0	0	10	2	3	0	32	25	1	0	4	2	3	0
l=1	16	18	131	120	11	12	73	33	1	0	0	0	0	0
l=2	10	12	14	13	1	0	1	0	0	0	0	0	0	0
l=3	43	94	103	103	3	0	3	7	0	0	0	0	0	0
l=4	8	17	4	7	1	1	0	0	0	0	0	0	0	0
l=5	11	19	3	7	0	0	1	0	0	0	0	0	0	0
l=6	3	1	0	1	0	0	0	0	0	0	0	0	0	0

BA – Before Alignment; AA – After Alignment

Table 3. Investigation of changes in l and r values for the smile images

	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6
ll	0	3	6	14	96	48	20	26	5	1	0	0
rr	0	0	0	10	33	20	46	94	7	13	3	0

Table 4. Adjustments in l and r values due to alignment for the no smile images

	r=0		r=1		r=2		r=3		r=4		r=5		r=6	
	BA	AA	BA	AA	BA	AA	BA	AA	BA	AA	BA	AA	BA	AA
l=0	0	0	0	0	0	0	2	0	0	0	3	4	5	1
l=1	0	0	6	4	1	2	46	34	14	2	25	6	2	0
l=2	0	0	2	3	1	0	7	8	2	0	0	0	0	0
l=3	5	5	84	102	9	9	120	101	3	3	15	11	2	0
l=4	2	2	18	28	0	0	7	12	0	0	0	0	0	0
l=5	13	30	52	78	1	3	21	26	1	1	2	1	0	0
l=6	15	14	8	3	0	0	0	1	0	0	0	0	0	0

Table 5. Total changes due to face alignment

	Total changes in l (lc)	Total changes in r (rc)	Total changes in both l and r (lc∩rc)	Changes only in l = (lc - lc∩rc)	Changes only in r = (rc - lc∩rc)	Total changes = lc∪rc = lc + rc - lc∩rc	Percentage of total changes = $\frac{lc \cup rc}{494}$
No smile	219	230	153	66	77	296	59.91
Smile	219	226	150	69	76	295	59.71

5. DEFINING THE THRESHOLD

In order to predict the smile face by evaluating the spread of lips, a threshold is defined. Threshold is the point between smile and no smile. From Figure 1 Threshold can define as $(l = r) < x$, can estimate as no smile and $> x$ can predict as smile because spread of lips is equal at both the ends. But it is not true for all the cases. Figure 8 shows the one sided smiles having unequal l and r values. In some cases, even the two sided smiles failed to provide equal l and r values because of improper pose in the pictures. It is found that expansion of lip at any one side, can estimate as a smile. Hence, a parameter - Maximum of l and r is taken and derived a threshold in section 5.2. For improper posed images, the lips spread in one side more and the other side a little. In order to find the effect of combination of l and r, average of l and r is taken as a parameter and the derived the threshold in section 5.1. One

more parameter Maxavgsum which is the combination of average and maximum is taken and estimated threshold is elucidated in section 5.3.

5.1 Average

The average is denoted with AANS for no smile case and AAS for smile case respectively. AANS or AAS = $(l+r)/2$ whose range is between 0 and 6. AANS and AAS can provide the information at a specific value i.e., at 0 or 0.5 etc., as illustrated in Table 6. But in order to find the spreading of lips, it is required to get sum of the values up to that specific value. Hence two parameters $AAANS[k] = \sum_{i=0}^k AANS [i]$ for no smile case. The possible values of k are 0,0.5,1,-----,5.5,6. and $AAAS[k] = \sum_{i=0}^k AAS [i]$ for smile case are defined and their respective values are illustrated in Table 6.



Figure 8. One sided smile images

Table 6. Threshold estimation for average

Average	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
AANS	0	20	132	144	153	28	16	1	0	0	0	0	0
AAS	0	0	4	10	138	81	200	21	37	2	1	0	0
AAANS	0	20	152	296	449	477	493	494	494	494	494	494	494
AAAS	0	0	4	14	152	233	433	454	491	493	494	494	494
494-AAAS	494	494	490	480	342	261	61	40	4	1	0	0	0
AE%	50	52.02	64.98	78.54	80.06	74.69	56.07	54.05	50.3	50.1	50	50	50

AAANS – Aggregate of average aligned no smile
 AAAS – Aggregate of average aligned smile

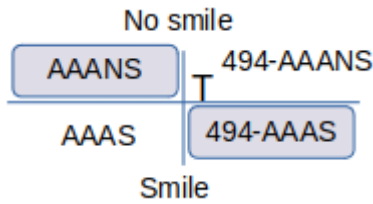


Figure 9. Threshold for average

Since for no smile, expansion of lips is less. The threshold is defined such that the value less than and equal to it is no smile and greater than that is the smiling one. Figure 9 illustrates threshold definition, horizontal line separates the smile from no smile case. As per the threshold definition, before the vertical line is no smile i.e., AAANS and after the smile is 494-AAAS.

Now the Accuracy $AE = (AAANS+494-AAAS)/988$, where 988 is the total number of images (both smile and no smile). The best threshold is the value having the maximum efficiency. From Table 6, 80.06% is the maximum efficiency which is achieved at 1.5 which is the threshold for the case average.

5.2 Maximum

Due to pose variations, there is a chance to reduce in one of the values of l and r. By considering the Maximum = Max(l,r), this kind of pose variations can resolve. Similar to AANS and AAS, MANS and MAS are the Maximum values for no smile and smile. MAANS, MAAS, and ME in Maximum are analogous to AAANS, AAAS, and AE in average, and all the values are illustrated in Table 7. The efficiency is calculated by defining the threshold value similar to the average, and the maximum efficiency achieved is 67.3 at threshold 3.

MANS – Maximum of aligned no smile,
 MAS – Maximum of aligned smile,
 MAANS – Aggregate of Maximum aligned no smile = $\sum_{i=0}^k MANS [i]$,

MAAS – Aggregate of Maximum aligned smile = $\sum_{i=0}^k MAS [i]$; $MA\% = \frac{MAANS+494-MAAS}{988} \times 100$.

Table 7. Efficiency calculation using maximum values

	0	1	2	3	4	5	6
MANS	0	140	37	262	25	28	2
MAS	0	4	5	259	47	160	19
MAANS	0	140	177	439	464	492	494
MAAS	0	4	9	268	315	475	494
494 - MAAS	494	490	485	226	179	19	0
MA%	50	63.77	67	67.3	65.08	51.9	50

5.3 Maxavgsum

In this case, efficiency is calculated by considering both Maximum and the average. The Maxavgsum is defined as Max + avg and the XANS, XAS, XAANS, XAAS, and XE, illustrated in Table 8 are analogous to AANS, AAS, AAANS, AAAS, and AE in average. The Maximum efficiency attained is 78.54, which is calculated similarly to the previous two cases.

XANS – Maxavgsum of aligned no smile,
 XAS – Maxavgsum of aligned smile,
 XAANS – Aggregate of Maxavgsum aligned no smile = $\sum_{i=0}^k XANS [i]$,
 XAAS – Aggregate of Maxavgsum aligned smile = $\sum_{i=0}^k XAS [i]$; $XE\% = \frac{XAAS+XAANS}{988} \times 100$.

A very few methods used lips in finding the smile. Royce et al. [20] identifies the lip corners, in this during the training phase lip corner y-axis positions are calculated and in the testing phase, the lip is compared with the reference identified. For corner detection, it uses Harris and Fast corner detection methods. Our present approach performs well compared to this as illustrated in Table 9. Li [9] uses mouth features that are extracted using HOG and LBP and fused together. The

maximum efficiency achieved with this method is 72.7% which is less than our maximum efficiency achieved 80.06%.

Table 8. Calculation of efficiency using Maxavgsum values

	XANS	XAS	XAANS	XAAS	494- XAAS	XE%
0	0	0	0	0	494	50
0.5	0	0	0	0	494	50
1	0	0	0	0	494	50
1.5	20	0	20	0	494	52.2
2	120	4	140	4	490	63.77
2.5	0	0	140	4	490	63.77
3	12	0	152	4	490	64.98
3.5	25	5	177	9	485	67
4	0	0	177	9	485	67
4.5	119	5	296	14	480	78.54
5	136	136	432	150	344	78.54
5.5	0	17	432	167	327	76.82
6	24	103	456	270	224	68.82
6.5	7	30	463	300	194	66.5
7	1	0	464	300	194	66.6
7.5	21	49	485	349	145	63.76
8	7	84	492	433	61	55.97
8.5	0	3	492	436	58	55.67
9	1	52	493	488	6	50.5
9.5	0	4	494	492	2	50.2
10	0	1	494	493	1	50.1
10.5	0	1	494	494	0	50
11	0	0	494	494	0	50
11.5	0	0	494	494	0	50
12	0	0	494	494	0	50

Table 9. Comparison with various methods

Method	Accuracy%
Harris Corner Detector [5]	77.5
Fast Corner Detector [5]	72.5
LBP [6]	65.05
HOG [6]	69.58
HOG+LBP [6]	72.7
Average	80.06
Max	67.3
Maxavgsum	78.54

6. CONCLUSION

For machines still detecting the smile is a challenging task. Usage of landmarks is one of the approaches in detecting the smile. The landmark points of eyes with respect to lips are considered for deriving two parameters – l and r. Average, Maximum and maxavg of the primary parameters are used to derive individual efficiencies of smile and no smile by defining different thresholds. Based on the best efficiency found, the best threshold is decided. It is observed that the threshold value of 1.5, 3 and 5 are the best for average, Maximum and Maxavgsum respectively and the overall maximum efficiency achieved is 80.06%.

REFERENCES

[1] Ochs, M., Diday, E., Afonso, F. (2016). From the symbolic analysis of virtual faces to a smiles machine. *IEEE Transactions on Cybernetics*, 46(2): 401-9. <https://doi.org/10.1109/TCYB.2015.2411432>

[2] Shan, C. (2012). Smile detection by boosting pixel

differences. *IEEE Transactions on Image Processing*, 21(1): 431-436. <https://doi.org/10.1109/TIP.2011.2161587>

[3] Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M., Movellan, J. (2009). Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11): 2106-2111. <https://doi.org/10.1109/TPAMI.2009.42>

[4] Chen, J., Ou, Q., Chi, Z., Fu, H. (2017). Smile detection in the wild with deep convolutional neural networks. *Machine Vision and Applications*, 28: 173-183. <https://doi.org/10.1007/s00138-016-0817-z>

[5] Ali, I., Dua, M. (2019). Smile detection: Current trends, challenges and future perspective. 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 151-156. <https://doi.org/10.1109/ICECA.2019.8822000>

[6] Bianco, S., Celona, L., Schettini, R. (2016). Robust smile detection using convolutional neural networks. *Journal of Electronic Imaging*, 25(6): 063002. <https://doi.org/10.1117/1.JEI.25.6.063002>

[7] Liang, S., Liang, X., Guo, M. (2015). Smile recognition based on deep Auto-Encoders. 2015 11th International Conference on Natural Computation (ICNC), pp. 176-181. <https://doi.org/10.1109/ICNC.2015.7377986>

[8] Guo, X., Polania, L., Barner, K. (2018). Smile detection in the wild based on transfer learning. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 679-686. <https://doi.org/10.1109/FG.2018.00107>

[9] Li, Y. (2014). Smile recognition based on face texture and mouth shape features. 2014 IEEE Workshop on Electronics, Computer and Applications, pp. 606-609. <https://doi.org/10.1109/IWECA.2014.6845692>

[10] Timotius, I.K., Setyawan, I. (2014). Evaluation of Edge Orientation Histograms in smile detection. 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), pp. 1-5. <https://doi.org/10.1109/ICITEED.2014.7007905>

[11] Yang, W., Zhen, S. (2011). Novel smile feature extraction algorithm using improved Gabor for mobile phone platform. Sixth International Conference on Image and Graphics, pp. 938-942. <https://doi.org/10.1109/ICIG.2011.28>

[12] Bai, Y., Guo, L., Jin, L., Huang, Q. (2009). A novel feature extraction method using Pyramid Histogram of Orientation Gradients for smile recognition. 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 3305-3308. <https://doi.org/10.1109/ICIP.2009.5413938>

[13] Zhang, L., Qi, L., Gao, L., Zheng, N., Chen, E. (2011). Recognizing smile emotion based on Fractional Fourier Transform. 2011 4th International Congress on Image and Signal Processing, pp. 940-944. <https://doi.org/10.1109/CISP.2011.6100363>

[14] Li, J., Chen, J., Chi, Z. (2016). Smile detection in the wild with hierarchical visual feature. 2016 IEEE International Conference on Image Processing (ICIP), pp. 639-643. <https://doi.org/10.1109/ICIP.2016.7532435>

[15] George, T., Potty, S.P., Jose, S. (2014). Smile detection from still images using KNN algorithm. International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICT), pp. 461-465.

- <https://doi.org/10.1109/ICCICCT.2014.6993006>
- [16] Girard, J.M., Cohn, J.F., De la Torre, F. (2015). Estimating smile intensity: A better way. *Pattern Recognition Letters*, 66: 13-21. <https://doi.org/10.1016/j.patrec.2014.10.004>
- [17] Akkoca, B.S., Gökmen, M. (2015). Automatic smile recognition from face images. *23rd Signal Processing and Communications Applications Conference (SIU)*, pp. 1985-1988. <https://doi.org/10.1109/SIU.2015.7130253>
- [18] Tang, L., Huang, T.S. (1996). Characterizing smiles in the context of video phone data compression. *Proceedings of 13th International Conference on Pattern Recognition*, pp. 659-663. <https://doi.org/10.1109/ICPR.1996.547028>
- [19] Tsai, A., Lin, T., Kuan, T., Bharanitharan, K., Chang, J., Wang, J. (2015). An efficient smile and frown detection algorithm. *International Conference on Orange Technologies (ICOT)*, pp. 139-143. <https://doi.org/10.1109/ICOT.2015.7498496>
- [20] Royce, E., Setyawan, I., Timotius, I.K. (2014). Smile recognition system based on lip corners identification. *The 1st International Conference on Information Technology, Computer, and Electrical Engineering*, pp. 222-225. <https://doi.org/10.1109/ICITACEE.2014.7065746>
- [21] Wang, S.L., Lau, W.H., S.H., Leung, A.W.C., Liew, (2004). Lip segmentation with the presence of beards. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. iii-529. <https://doi.org/10.1109/ICASSP.2004.1326598>
- [22] Jeong, M., Ko, B.C., Kwak, S., Nam, J. (2018). Driver facial landmark detection in real driving situations. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10): 2753-2767. <https://doi.org/10.1109/TCSVT.2017.2769096>
- [23] Wu, T., Turaga, P., Chellappa, R. (2012). Age estimation and face verification across aging using landmarks. *IEEE Transactions on Information Forensics and Security*, 7(6): 1780-1788. <https://doi.org/10.1109/TIFS.2012.2213812>