On Constrained Local Model Feature Normalization for Facial Expression Recognition

Zhenglin Pan, Mihai Polceanu, and Christine Lisetti

School of Computing & Info. Sciences, Florida International University,
11200 SW 8th St, Miami, FL 33199, USA
zpan004@fiu.edu,{mpolcean,lisetti}@cs.fiu.edu
http://ascl.cis.fiu.edu/

Abstract. Real time user independent facial expression recognition is important for virtual agents but challenging. However, since in real time recognition users are not necessarily presenting all the emotions, some proposed methods are not applicable. In this paper, we present a new approach that instead of using the traditional base face normalization on whole face shapes, performs normalization on the point cloud of each landmark. The result shows that our method outperforms the other two when the user input does not contain all six universal emotions.

Keywords: Constrained Local Model, feature normalization, preprocessing, facial expression recognition

1 Introduction

It is important for intelligent virtual agents to have the capability that correctly detects and interprets the emotions of human users during the interaction in many areas. To improve this capability, we need to train our virtual agents to analyze facial expression, which is one of the most significant factors that we, human beings, take into consideration when we attempt to tell other people's sentiments.

While tasks such as face recognition require differentiating between individuals based on facial features, facial expression recognition relies on variations of these facial features and on their dynamics. Consequently, one major obstacle in accurately classifying users' facial expressions is the large amount of variations in expressed emotions. This makes classification difficult due to high overlap when merging data from multiple individuals.

To address these issues, many state of the art facial expression recognition systems rely on techniques to extract user-specific information, which enables multi-user data to be normalized in a more efficient manner, giving way to superior classification performance. However, some of the existing approaches have trouble with real time facial expression recognition when not all the universal emotions are provided. In this paper, we investigate these common techniques and describe a novel application of an existing point registration algorithm that performs better in this situation.

2 Related Work

Traditionally, action unit zero (AU0) has been used to normalize facial expression data of multiple subjects.

Jeni *et al.*[5] proposed Personal Mean Shape (PMS), which is the mean of neutral and extreme facial expressions of a subject. PMS turned out to work quite well on the Cohn-Kanade Extended Facial Expression (CK+) [6] Database. However, a restriction of their solution is that the base face has to be built based on explicitly labeled neutral and extreme facial expressions of the subjects. Hence, this method may have difficulties with real time video streams.

In comparison, Baltrusaitis *et al.*[1] calculated the median of all frames of a subject as their base face. This approach was based on the assumption that neutral face is the most frequently shown facial expression and the base face is very close to the neutral face, yet they found this assumption did not hold for all the situations.

In this paper, we propose a novel face landmark preprocessing approach, which works better than the mentioned approaches in some cases and could lead to better insight into improving the performance of existing facial expression recognition methods.

3 CLM Feature Normalization

Constrained Local Model (CLM) is a robust facial feature tracking algorithm proposed by Cristinacce *et al.*[3]. In our experiment, we leverage one of the open source implementations of CLM, CLM-Z, released by Baltrusaitis *et al.*[2].

We test the consistency of our method by applying the CLM-Z algorithm on 3 databases: Cohn-Kanade Extended Facial Expression Database (CK+)[6], Binghamton University 3D Facial Expression Database (BU-3DFE)[9] and Binghamton University 3D+time Facial Expression Database (BU-4DFE)[8].

To preprocess the data, we used the Generalized Procrustes Analysis (GPA) algorithm[4]. Then we apply Jeni's approach (mean-based normalization), Baltrusaitis' approach (median-based normalization) and our approach (point-wise normalization) on the normalized landmarks, respectively. Finally, we perform leave one subject out cross validation on the processed data using SVM. The results and discussions are shown in the following sections.

To generalize Jeni's mean-based normalization to datasets that do not include neutral faces, instead of using their approach that takes the mean of both neutral and extreme expressions, we take the mean of the latter only.

We also tested Baltrusaitis' approach, which takes the median of all the images as the base face. They assume that the neutral face is the most frequently shown emotion throughout a video.

In real time emotion recognition, the users may not provide all seven typical emotions. Consequently, the mean or median based normalization may have difficulties calculating the base face. However, we can deal with this situation using Coherent Point Drift (CPD) [7] point set registration algorithm.

Table 1. Comparison between mean, median and CPD, on the CK+ (subjects with ≥ 3 facial expressions), BU-3DFE and BU-4DFE databases (complete data).

Database	Technique	Anger	Disgust	Fear	Happiness	Sadness	Surprise	**Average**
CK+	Mean	64.70	89.47	60.00	100.0	75.00	92.30	**80.25**
	Median	64.70	89.47	35.71	96.00	75.0	92.30	**75.53**
	CPD	70.58	84.21	60.00	100.0	81.25	96.15	**82.03**
BU-3DFE	Mean	70.00	78.25	69.50	84.00	67.75	90.25	**76.63**
	Median	69.00	77.00	69.00	85.00	69.50	90.75	**76.71**
	CPD	64.75	77.25	63.75	85.75	70.75	89.75	**75.33**
BU-4DFE	Mean	80.59	76.73	66.04	91.39	76.04	84.75	**79.26**
	Median	83.37	76.63	66.34	90.30	83.17	85.54	**80.89**
	CPD	78.12	73.17	62.87	93.56	73.86	86.63	**78.04**

Although mean-based and median-based normalization are better than the point-wise normalization on BU-3DFE and BU-4DFE datasets, the difference among them is less than 3%, which means that after users have expressed their 6 typical emotions, all 3 approaches work more or less the same. However, in the case where users only show half of the universal emotions, CPD outperforms median, which wins both BU-3DFE and BU-4DFE, with almost 7% difference as shown in Table 1. This is consistent with the intuition that mean and median have poor performance on incomplete data. Since in real time emotion recognition users are not very likely to show all their facial expressions, detecting their base face through the incomplete input becomes inaccurate. However, the point-wise normalization can better overcome the troubles of incompleteness based on the result.

Now that the point-wise approach shows its advantage on partial input data, we expect that the advantage stays with full data. Although in our experiment it does not reach better performance than the traditional base face extraction methods, it still has the potential to outperform them with better parameter and algorithm choices. If this will the case, then the neutral face can no longer be considered the key to user-independent emotion recognition.

4 Conclusions and Future Work

Our study focused on investigating a new face landmark preprocessing approach, which shifts the landmarks to their corresponding clusters geometrically. We compared our algorithm with two existing methods that normalize on users' base face. Based on the experiment data, we claim that extracting base face is not necessarily the best approach to perform real time user independent emotion recognition. One of the alternatives is that for each CLM extracted landmark, we can construct a topology that includes the position of this very point in all the frames from real time video and map the landmarks to the corresponding clusters in our model using the CPD point registration algorithm. This approach has better classification accuracy compared to mean-based or median-based normalization methods when the input data does not include all 6 universal emotions.

In the future, we will apply the same methodology to other datasets to verify our theory and explore some means for improvement.

Additional information about the subjects may lead to better landmark clustering and therefore better recognition accuracy, due to potential similarities in expressed emotions within demographic groups. Meanwhile, instead of classifying emotions based on the geometry positions, we will work with the Action Units. Analyzing Action Units will not only help increase the classification accuracy, but also give us information about subtle micro facial expressions, which are difficult to identify if classified by geometry position only.

In this paper we did not include the importance of the face features for each emotion, but in recognizing different emotions, the same group of landmarks weigh differently. In the future, we will investigate the dominating facial features for each emotion and perform classification only upon these features. We expect to reduce the dimension, increase classification accuracy and shorten the time cost by this approach.

References

1. Baltrusaitis, T., Mahmoud, M., Robinson, P.: Cross-dataset learning and person-specific normalisation for automatic action unit detection. In: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on. vol. 6, pp. 1–6. IEEE (2015)
2. Baltrusaitis, T., Robinson, P., Morency, L.P.: Constrained local neural fields for robust facial landmark detection in the wild. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 354–361 (2013)
3. Cristinacce, D., Cootes, T.F.: Feature detection and tracking with constrained local models. In: BMVC. vol. 2, p. 6. Citeseer (2006)
4. Gower, J.C.: Generalized procrustes analysis. Psychometrika 40(1), 33–51 (1975)
5. Jeni, L.A., Lőrincz, A., Nagy, T., Palotai, Z., Sebők, J., Szabó, Z., Takács, D.: 3d shape estimation in video sequences provides high precision evaluation of facial expressions. Image and Vision Computing 30(10), 785–795 (2012)
6. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. pp. 94–101. IEEE (2010)
7. Myronenko, A., Song, X.: Point set registration: Coherent point drift. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32(12), 2262–2275 (2010)
8. Yin, L., Chen, X., Sun, Y., Worm, T., Reale, M.: A high-resolution 3d dynamic facial expression database. In: Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference On. pp. 1–6. IEEE (2008)
9. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3d facial expression database for facial behavior research. In: Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on. pp. 211–216. IEEE (2006)